


© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12850
RESEARCH ARTICLE

Application of the RAND-UCLA Appropriateness Methodology to a Large Multidisciplinary Stakeholder Group Evaluating the Validity and Feasibility of Patient-Centered Standards in Geriatric Surgery

Julia R. Berian , *Tracey L. Baker, Ronnie A. Rosenthal, JoAnn Coleman, Emily Finlayson, Mark R. Katlic, Sandhya A. Lagoo-Deenadayalan, Victoria L. Tang, Thomas N. Robinson, Clifford Y. Ko, and Marcia M. Russell*

Objectives. To explore (1) differences in validity and feasibility ratings for geriatric surgical standards across a diverse stakeholder group (surgeons vs. nonsurgeons, health care providers vs. nonproviders, including patient-family, advocacy, and regulatory agencies); (2) whether three multidisciplinary discussion subgroups would reach similar conclusions.

Data Source/Study Setting. Primary data (ratings) were reported from 58 stakeholder organizations.

Study Design. An adaptation of the RAND-UCLA Appropriateness Methodology (RAM) process was conducted in May 2016.

Data Collection/Extraction Methods. Stakeholders self-administered ratings on paper, returned via mail (Round 1) and in-person (Round 2).

Principal Findings. In Round 1, surgeons rated standards more critically (91.2 percent valid; 64.9 percent feasible) than nonsurgeons (100 percent valid; 87.0 percent feasible) but increased ratings in Round 2 (98.7 percent valid; 90.6 percent feasible), aligning with nonsurgeons (99.7 percent valid; 96.1 percent feasible). Three parallel subgroups rated validity at 96.8 percent (group 1), 100 percent (group 2), and 97.4 percent (group 3). Feasibility ratings were 76.9 percent (group 1), 96.1 percent (group 2), and 92.2 percent (group 3).

Conclusions. There are differences in validity and feasibility ratings by health professions, with surgeons rating standards more critically than nonsurgeons. However, three separate discussion subgroups rated a high proportion (96–100 percent) of standards as valid, indicating the RAM can be successfully applied to a large stakeholder group.

Key Words. RAND-UCLA Appropriateness Method, Delphi methods, geriatric surgery

Although randomized controlled trials are considered the gold standard for evidence, many medical decisions must be made without level 1 evidence. Almost 30 years ago, the RAND-UCLA Appropriateness Methodology (RAM) was developed as a way to guide medical and surgical decision making in the face of limited evidence (Brook 1994). The methodology involves two rounds of evaluation, the first by each evaluator individually, and second after a face-to-face discussion with all the evaluators. This method has frequently been applied to surgical procedures, traditionally for the purpose of designating indications as appropriate, uncertain, or inappropriate. The original studies applied RAM to individual procedures such as carotid endarterectomy, coronary artery bypass graft, and hysterectomy (Park et al. 1986; Leape et al. 1993; Shekelle et al. 1998; Broder et al. 2000). Since that time, the methodology has been used frequently and expanded beyond procedural indications to evaluate underuse and overuse and to develop quality indicators (McGory et al. 2005, 2009; Maggard, McGory, and Ko 2006; McGory, Shekelle, and Ko 2006; Lawson et al. 2011, 2012).

There is growing interest in defining quality and standardizing care for older adults, given the aging population and their increased risk for poor post-operative outcomes across procedure types (Finlayson, Fan, and Birkmeyer 2007; Bentrem et al. 2009; Makary et al. 2010; Robinson et al. 2011; Chow

Address correspondence to Marcia M. Russell, M.D., F.A.C.S., Veterans Affairs Greater Los Angeles Healthcare System and Department of Surgery, University of California, Los Angeles, Los Angeles, CA 90073; e-mail: mrrussell@mednet.ucla.edu. Julia R. Berian, M.D., M.S., is with the Department of Surgery, University of Chicago Medical Center, Chicago, IL. Tracey L. Baker, B.S., is with the Livongo Health Inc., Chicago, IL. Ronnie A. Rosenthal, M.S., M.D., F.A.C.S., is with the Department of Surgery, Yale University, New Haven, CT. JoAnn Coleman, D.N.P., A.N.P., A.D.N.P., A.O.C.N., and Mark R. Katlic, M.D., F.A.C.S., are with the Department of Surgery, LifeBridge Health, Baltimore, MD. Emily Finlayson, M.D., M.S., F.A.C.S., is with the Department of Surgery, University of California, San Francisco, San Francisco, CA. Sandhya A. Lagoo-Deenadayalan, M.D., Ph.D., F.A.C.S., is with the Department of Surgery, Duke University, Durham, NC. Victoria L. Tang, M.D., M.A.S., is with the Department of Medicine, Division of Geriatrics, University of California, San Francisco, San Francisco, CA. Thomas N. Robinson, M.D., M.S., F.A.C.S., is with the Department of Surgery, University of Colorado, Denver, Aurora, CO. Clifford Y. Ko, M.D., M.S., M.S.H.S., F.A.C.S., F.A.S.C.R.S., is with the Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL, and Department of Surgery, University of California, Los Angeles, Los Angeles, CA.

et al. 2012a). Older adults undergo a disproportionate share of surgical procedures, setting the stage for an increasing burden of complications and health care costs in this vulnerable population (Etzioni et al. 2003a, b). Increased demand in this high-risk population has resulted in several efforts to define high-quality surgical care for older adults (McGory et al. 2005, 2009; Chow et al. 2012b; Mohanty et al. 2016).

Older patients with complex medical and psychosocial issues frequently require coordination across disciplines. High-quality care requires not only medical and surgical specialists, but also coordination from nursing, pharmacy, social work, case management, and physical medicine and rehabilitation. Many of the original RAM studies included only medical and surgical specialists. As the complexity of health care delivery increases, driving a need for standardization, it is critical to adapt the RAM and similar methods of defining high-quality care to include additional health care professionals beyond physicians. Furthermore, there is a growing call to include patients, families, and caregivers in health care decision making, not just on an individual basis but for overall hospital processes and even research agendas (National Priorities Partnership 2008; Frampton et al. 2017).

The Coalition for Quality in Geriatric Surgery (CQGS) is a multidisciplinary effort to establish high-quality standards for perioperative care of older adults. Fifty-eight organizations are stakeholders in the CQGS, representing a range of expertise from nursing, social work, and case management to medicine, geriatrics, and surgical specialists, to payers and regulatory health care agencies, to organizations representing the interests of patients and families. With a wide-ranging set of players shaping the current health care environment, the CQGS was designed to include all the relevant voices. To our knowledge, this is the first time the RAM has been applied to such a large and diverse group; therefore, the RAM has been adapted by splitting the stakeholders into three smaller, interdisciplinary panels for simultaneous discussion sessions. Prior research indicates that panel composition does have some effect on appropriateness ratings, with those who perform an intervention often rating appropriateness more favorably than nonproceduralists (Coulter, Adams, and Shekelle 1995; Kahan et al. 1996). We therefore sought to (1) explore differences in validity and feasibility ratings across types of health professionals (surgeons vs. nonsurgeons), and compared to representatives who do not provide clinical patient care, such as patient-family representation, payers, advocacy, and regulatory groups (health care providers vs. nonproviders); and (2) evaluate whether three parallel multidisciplinary panels would reach

similar conclusions regarding the validity and feasibility of geriatric surgical standards.

METHODS

Panel Composition

The CQGS is a collection of 58 stakeholder organizations engaged in setting standards for high-quality surgical care of older adults. The CQGS stakeholder organizations were identified and convened by the American College of Surgeons and The John A. Hartford Foundation, intended to represent a diverse, interdisciplinary, and wide-ranging set of voices critical for high-quality surgical care of the older adult. Stakeholder organizations submitted Letters of Support for the initiative. Individual representatives were nominated by their respective organizations. There were 48 individuals scheduled to attend the in-person RAM panel. Of these, four were unable to attend due to unanticipated circumstances. Three were absent due to illness or emergency, one due to unexpected scheduling conflict. Regardless of attendance by individual representatives, no organizations have dropped support of the project.

Traditional RAM approaches refer to expert “panelists,” which will be termed “stakeholders” for the purpose of the CQGS project. CQGS stakeholders represented anesthesia, care transitions (including physical medicine and rehabilitation, and postacute and long-term care facilities), ethics and palliative care, internal medicine (including primary care, medicine and hospitalists), geriatrics, measurement science, nursing, patient and family advocates, payers, pharmacy, regulatory agencies, social work and case management, and multiple surgical specialties. For the purpose of categorizing stakeholders into surgeon versus nonsurgeon and health care provider versus nonprovider, the clinical practice of the individual representative was taken into account. For example, the representative for the ACS Committee on Surgical Palliative Care provides input and expertise on the perioperative role of palliative care; however, the representative’s daily clinical practice is surgical. As such, this representative is categorized in the “surgeon” and “health care provider” groups. The group of health care providers included any stakeholder representative who participates in patient care. The nonprovider group included representatives from patient-family organizations, payers, advocacy, and regulatory agencies.

Preliminary Standards Development

The CQGS standards describe care elements intended to identify or contribute to higher quality care; they are not designed to distinguish between perfect and excellent care, rather to set the bar for reasonable expectations of “high-quality” geriatric surgical care according to the CQGS stakeholders. In September 2015, the CQGS was convened in an initial Stakeholder Kickoff Meeting to define current gaps in surgical care for older adults and ideal future solutions. The results from this meeting set the framework for the development for the CQGS Standards (see Appendix S1), which is described in detail elsewhere (Berian et al. 2017). In brief, development of candidate standards was conducted from September 2015 through March of 2016 drawing upon a structured literature review, with duplicate title/abstract review to identify the highest level of available evidence. Supporting evidence was evaluated and summarized for distribution to CQGS stakeholders in the preliminary standards rating packet.

Evaluation Criteria

Coalition for Quality in Geriatric Surgery stakeholders rated standards separately with regard to validity and feasibility. Validity was defined by (1) adequate scientific evidence or professional consensus exists to support a link between the performance of care specified by the standard and (a) the accrual of health benefits to the patient or (b) improved delivery and experience of care; (2) a hospital that consistently meets the standard would be considered higher quality (and inconsistent performance or failure to meet the standard would be considered poor quality). Feasibility was defined by whether hospitals within a range of practice settings (private or public, academic or nonacademic, urban or rural) could meet the standard given a reasonable effort.

Potential validity and feasibility ratings ranged from 1 to 9, with 1 = definitely not valid/feasible, 5 = uncertain or equivocal validity/feasibility, and 9 = definitely valid/feasible. A median validity and feasibility score was calculated for each standard, as well as a measure of statistical agreement based upon the spread of ratings. Traditionally, disagreement in the RAM process was defined as the number of panel members that voted outside the 3-point range that contained the median. However, for panels consisting of more than nine members, another method based on the interpercentile range adjusted for symmetry (IPRAS) to determine statistical agreement or disagreement is recommended (Fitch et al. 2001). The interpercentile range (IPR) that

identifies disagreement is known to be smaller when ratings are symmetric with respect to the median rating on a 9-point scale. In cases of asymmetry, if the IPR is greater than the IPRAS, the rated item displays disagreement.

A standard was considered valid or feasible if the median score was ≥ 7 without disagreement. A standard was considered uncertain if the median score was 4–6, or the distribution demonstrated disagreement. A standard was invalid or infeasible if the median score was ≤ 3 without disagreement.

RAM Process and Data Collection

The RAM, a modified Delphi technique (Brook 1994; Fitch et al. 2001; Lawson et al. 2011, 2012), requires that expert stakeholders evaluate candidate standards twice: once independently and again during an in-person meeting. The preliminary standards, including rationale and literature summary, were distributed for individual stakeholder rating in March 2016 (Round 1). After returning individual stakeholder rating forms by mail, the ratings were entered in duplicate into a database for analysis. Random error checking was performed for six complete stakeholder rating sets of 308 standards for 2 scores each (validity and feasibility). The error rate was identified at 2 errors of 1,848 data points, or 0.11 percent. The data were analyzed for median score and degree of agreement to categorize the standards as valid/uncertain/invalid and feasible/uncertain/infeasible.

The in-person meeting was conducted May 12–13, 2016, in Chicago, IL, during which stakeholders discussed standards with uncertain validity or feasibility and then, individually, rerated all items on validity and feasibility (Round 2). The meeting began with a gathering of the entire CQGS group, which included an update on the preliminary standards development and a review of RAM process. Stakeholders were then divided into three subgroups during which discussions were conducted in parallel, guided by experienced RAM moderators. The three subgroups were designed to achieve a balance across the multiple stakeholder professions. Each stakeholder received a unique rating sheet that recapped his or her own rating from Round 1 and provided a de-identified distribution of ratings across all CQGS stakeholders. Stakeholders discussed validity and feasibility, focusing on areas of uncertainty as identified in Round 1 ratings. The three subgroups were reconvened into the larger CQGS group to summarize and report-out major points made during the subgroup discussions. To conclude the meeting, CQGS stakeholders individually rated the standards and submitted their Round 2 ratings. Stakeholder data were again entered into an electronic dataset, with the error

rate for Round 2 calculated at 4 of 1,848 data points or 0.21 percent. Stakeholder data were included for analysis only if there were ratings available for both Round 1 and Round 2.

Definitions

The CQGS preliminary standards book included a glossary and key terms. Older adult was defined as 65 years of age and older. Additional key terms included elective versus nonelective surgery, inpatient versus outpatient surgery, and risk factors that would classify an older adult as “high risk.” These key terms and a full list of the 308 standards are provided in Appendix S1.

Software

Data entry was performed in *Microsoft Excel*, with dataset import for aggregation and analysis utilizing *SAS9.4* (SAS Institute, Inc., Cary, NC, USA).

RESULTS

The Coalition for Quality in Geriatric Surgery includes 58 Stakeholders, all of whom received the 308 preliminary standards for rating. Forty-six stakeholders (79.3 percent) returned the Round 1 ratings. Although 48 stakeholders were scheduled to attend Round 2, there were four stakeholders unable to complete Round 2 ratings due to unanticipated circumstances. Therefore, Round 2 ratings were completed by 44 (95.7 percent) of the 46 stakeholders that had completed Round 1 ratings. All three subgroups were designed to include representation from care transitions, ethics/palliative care, geriatrics, nursing, patient/family perspective, regulatory agencies, and surgery. The remaining stakeholder groups were limited to fewer than three representatives, and therefore stakeholders were present in one or two of the parallel groups. The final subgroup composition was relatively balanced, with each including representatives from care transitions, ethics/palliative care, nursing, and surgery (Table 1).

Standards Ratings According to Stakeholder Type

Stakeholders were grouped into surgeons ($n = 17$) versus nonsurgeons ($n = 27$; Table 2A). Round 1 validity ratings included more “Uncertain”

Table 1: Composition of Stakeholder Subgroup Panels

<i>Panelist</i>	<i>Group 1 N = 16</i>	<i>Group 2 N = 16</i>	<i>Group 3 N = 16</i>
Anesthesia	0	1	1
Care transitions	1	1	2
Ethics/palliative care	1	1	1
Geriatrics	1	1*	2
Measurement science	1	1	0
Medicine	1	0	1
Nursing	2	1	2
Patients and families	1*	2	1
Payers	2	1	0
Pharmacy	1*	0	1
Regulatory	1	1*	1
Social work and case management	0	1	1
Surgery and surgical specialists	4	5	3

Note. Medicine includes primary care, internal medicine, and hospitalist medicine. Care transitions include physical medicine and rehabilitation, postacute and long-term care.

*Stakeholders unable to submit Round 2 ratings.

standards as rated by surgeons ($n = 27$, 8.8 percent) compared with nonsurgeons ($n = 0$; Table 2A). There were no standards rated invalid by either surgeons or nonsurgeons. Round 1 feasibility ratings demonstrated more “Uncertain” and “Infeasible” standards as rated by surgeons ($n = 105$, 34.1 percent Uncertain and $n = 3$, 1 percent Infeasible) compared with nonsurgeons ($n = 40$, 13.0 percent Uncertain, $n = 0$, Infeasible). In Round 2, ratings for validity and feasibility demonstrated a decrease in Uncertain and Infeasible standards. Surgeons rated 4 (1.3 percent) standards uncertain validity, compared with 1 (0.3 percent) standard rated uncertain validity by nonsurgeons. Surgeons rated 28 (9.1 percent) standards uncertain feasibility, compared with 12 (3.9 percent) standards uncertain by nonsurgeons. There was 1 (0.3 percent) standard rated infeasible by surgeons and none by nonsurgeons.

Stakeholders were classified as health care providers ($n = 35$) or nonproviders ($n = 9$; Table 2B). Round 1 ratings indicated more “Uncertain” validity for health care providers compared with nonproviders (6, 2 percent vs. 3, 1 percent) and more “Uncertain” feasibility for health care providers compared with nonproviders (69, 22.4 percent vs. 43, 14 percent). In contrast, Round 2 ratings were similar between health care providers compared with nonproviders, with both high validity (307, 99.7 percent vs. 307, 99.7 percent) and feasibility (290, 94.2 percent vs. 291, 94.5 percent).

Table 2: Number and Percentage of Ratings (A) Surgeons versus Nonsurgeons. (B) Health Care Providers versus Nonproviders

	Validity Ratings N = 308 Standards				Feasibility Ratings N = 308 Standards		
	Valid	Uncertain	Invalid		Feasible	Uncertain (4-6)	Infeasible
Round 1	Surgeons (17)	27 (8.8%)	0	200 (64.9%)	105 (34.1%)	3 (1.0%)	
	Nonsurgeons (27)	0	0	268 (87%)	40 (13%)	0	
Round 2	Surgeons (17)	4 (1.3%)	0	279 (90.6%)	28 (9.1%)	1 (0.3%)	
	Nonsurgeons (27)	1 (0.3%)	0	296 (96.1%)	12 (3.9%)	0	
	Validity Ratings N = 308 Standards				Feasibility Ratings N = 308 Standards		
	Valid	Uncertain	Invalid		Feasible	Uncertain	Infeasible
Round 1	Health care providers (35)	6 (2%)	0	239 (77.6%)	69 (22.4%)	0	
	Nonproviders (9)	3 (1%)	0	265 (86%)	43 (14%)	0	
Round 2	Health care providers (35)	1 (0.3%)	0	290 (94.2%)	18 (5.8%)	0	
	Nonproviders (9)	1 (0.3%)	0	291 (94.5%)	17 (5.5%)	0	

Standards Ratings across Parallel Subgroups

Round 2 validity ratings were compared across the three parallel RAM discussion groups (Table 3A). Most standards were rated “Valid” in all three groups—298 (96.8 percent) standards in group 1, 308 (100 percent) standards in group 2, and 300 (97.4 percent) standards in group 3. Among the standards rated “Uncertain,” there were more ratings with “disagreement” in group 3 (4 of 8), compared with group 1 (1 of 10) and group 2 (0). Examples of uncertain validity items in group 1 included mandatory consideration for minimally invasive surgical approaches and the inclusion of the patient and family into multidisciplinary conference. Examples of uncertain validity items in group 3 included restrictions on the use of neuromuscular blockade and mandatory discussion of patient preferences for hemodialysis and tube feedings preoperatively. Both groups demonstrated uncertain validity ratings for the inclusion of Hematology-Oncology into multidisciplinary conference and the use of swallow studies preoperatively.

Round 2 feasibility ratings were compared across the three parallel RAM discussion groups (Table 3B). More standards were rated “Uncertain” feasibility in group 1 (71, 23.1 percent) compared with group 2 (12, 3.9 percent) and group 3 (23, 7.5 percent). There was 1 standard rated “Infeasible” by group 3. Of the standards rated “Uncertain” feasibility, there was disagreement in 3 of 71 standards in group 1, 1 of 12 in group 2, and 8 of 23 in group 3. There were nine items that were rated as uncertain feasibility among groups 1, 2, and 3. Examples of these nine shared uncertain feasibility items were the use of patient navigators throughout the perioperative period for high-risk patients, the need for large rooms to accommodate family visitation, the ability to respond to home safety concerns preoperatively, and collection of 90-day data on geriatric-specific outcomes. Of note, the inclusion of Hematology-Oncology was rated infeasible by group 3 and uncertain feasibility by groups 1 and 2.

Standards Ratings across Rounds

Taking into account all stakeholder ratings, the vast majority of standards (228, 74.0 percent) had no change in median validity score between Round 1 and Round 2 (Figure 1). Among standards that did change, 63 of 80 increased in score. Despite a change in median score, only three standards were recategorized as a result of the Round 2 ratings. Two standards moved from “Uncertain” to “Valid” and one from “Valid” to “Uncertain.” The two items that

Table 3: (A) Validity Ratings by Group and Standard Section (Round 2). (B) Feasibility Ratings by Group and Standard Section (Round 2)

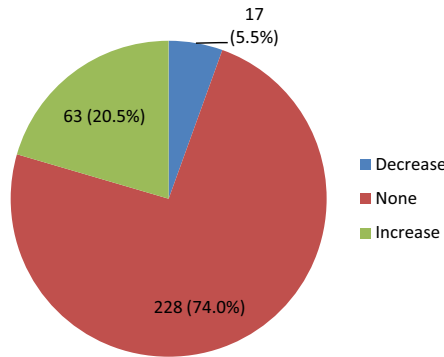
	Group 1			Group 2			Group 3		
	Valid	Uncertain	Invalid	Valid	Uncertain	Invalid	Valid	Uncertain	Invalid
1.1 Goals and decision making (n = 29)	29 (100%)	0	29 (100%)	29 (100%)	0	25 (86.2%)**	25 (86.2%)	0	4 (13.8%)**
1.2 Preoperative optimization (n = 89)	82 (92.1%)	7 (7.9%)	89 (100%)	89 (100%)	0	87 (97.8%)	87 (97.8%)	0	2 (2.2%)
1.3 Transitions of care (n = 39)	39 (100%)	0	39 (100%)	39 (100%)	0	39 (100%)	39 (100%)	0	0
2.1 Immediate preoperative care (n = 9)	8 (88.9%)	1 (11.1%)*	9 (100%)	9 (100%)	0	9 (100%)	9 (100%)	0	0
2.2 Intraoperative care (n = 9)	9 (100%)	0	9 (100%)	9 (100%)	0	8 (88.9%)	8 (88.9%)	0	1 (11%)*
2.3 Postoperative care (n = 66)	65 (98.5%)	1 (1.5%)	66 (100%)	66 (100%)	0	66 (100%)	66 (100%)	0	0
3.1 Personnel and committee structure (n = 52)	51 (98.1%)	1 (1.9%)	52 (100%)	52 (100%)	0	51 (98.1%)	51 (98.1%)	0	1 (1.9%)*
3.2 Credentialing and education (n = 7)	7 (100%)	0	7 (100%)	7 (100%)	0	7 (100%)	7 (100%)	0	0
4.0 Data and follow-up (n = 8)	8 (100%)	0	8 (100%)	8 (100%)	0	8 (100%)	8 (100%)	0	0
Total	298 (96.8%)	10 (3.3%)	308 (100%)	308 (100%)	0	300 (97.4%)	300 (97.4%)	0	8 (2.6%)

	Group 1			Group 2			Group 3		
	Feasible	Uncertain	Infeasible	Feasible	Uncertain	Infeasible	Feasible	Uncertain	Infeasible
1.1 Goals and decision making (n = 29)	21 (72.4%)	8 (27.6%)*	28 (96.6%)	28 (96.6%)	1 (3.4%)	22 (75.9%)	22 (75.9%)	7 (24.1%)***	0
1.2 Preoperative optimization (n = 89)	50 (56.2%)	39 (43.8%)**	82 (92.1%)	82 (92.1%)	7 (7.9%)	78 (87.6%)	78 (87.6%)	11 (12.4%)**	0
1.3 Transitions of care (n = 39)	34 (87.2%)	5 (12.8%)	39 (100%)	39 (100%)	0	38 (97.4%)	38 (97.4%)	1 (2.6%)*	0
2.1 Immediate preoperative care (n = 9)	9 (100%)	0	9 (100%)	9 (100%)	0	9 (100%)	9 (100%)	0	0
2.2 Intraoperative care (n = 9)	9 (100%)	0	9 (100%)	9 (100%)	0	8 (88.9%)	8 (88.9%)	1 (11.1%)*	0
2.3 Postoperative care (n = 66)	59 (89.4%)	7 (10.6%)	64 (97.0%)	64 (97.0%)	2 (3.0%)	65 (98.5%)	65 (98.5%)	1 (1.5%)*	0
3.1 Personnel and committee structure (n = 52)	43 (82.7%)	9 (17.3%)	51 (98.1%)	51 (98.1%)	1 (1.9%)*	51 (98.1%)	51 (98.1%)	0	1 (1.9%)
3.2 Credentialing and education (n = 7)	7 (100%)	0	7 (100%)	7 (100%)	0	7 (100%)	7 (100%)	0	0
4.0 Data and follow-up (n = 8)	5 (62.5%)	3 (37.5%)	7 (87.5%)	7 (87.5%)	1 (12.5%)	6 (75.0%)	6 (75.0%)	2 (25.0%)	0
Total	237 (76.9%)	71 (23.1%)	296 (96.1%)	296 (96.1%)	12 (3.9%)	284 (92.2%)	284 (92.2%)	24 (7.8%)	1

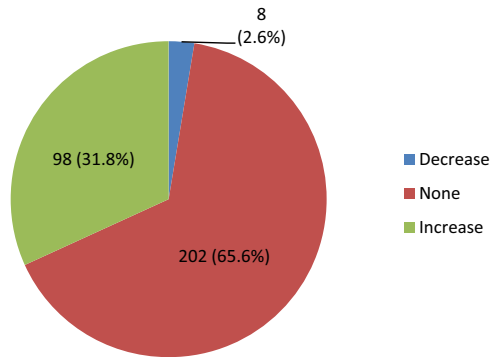
Note. Each * designates one standard for which uncertainty was associated with "disagreement." (A) There were no standards rated Invalid in Round 2. (B) There were no standards rated Infeasible in groups 1 and 2.

Figure 1: Change in Median Scores—All Stakeholders, Round 2 Compared with Round 1 [Color figure can be viewed at wileyonlinelibrary.com]

(A) Validity Ratings



(B) Feasibility Ratings



Notes. Ratings could increase, decrease, or remain unchanged between Round 1 and Round 2. (A) The majority of standards did not change in validity between Rounds 1 and 2 ($n = 228$, 74.0 percent). There were 63 (20.5 percent) standards with an increase in validity ratings and 17 standards (5.5 percent) with a decrease in validity ratings. (B) The majority of standards did not change in feasibility between Rounds 1 and 2 ($n = 202$, 65.6 percent). There were 98 (31.8 percent) standards with an increase in feasibility ratings and 8 (2.6 percent) standards with a decrease in feasibility ratings.

became “Valid” were the use of alternative, nonpharmacologic methods for anxiety control and the use of a patient navigator in high-risk patients. The item which became “Uncertain” was including hematology-oncology in pre-operative multidisciplinary conference. Similarly, the majority of standards

(202, 65.6 percent) had no change in median feasibility score between Round 1 and Round 2. Among standards that did change, 98 of 106 increased in score. There were 40 standards rated “Uncertain” feasibility in Round 1 that were recategorized as “Feasible” in Round 2.

Surgeon and nonsurgeon subgroups were compared to examine how validity ratings changed across Round 1 and Round 2 and to examine whether the subgroups changed scores for the same or different standards. Surgeons increased validity scores for 117 standards and 24 standards shifted categories from “Uncertain” to “Valid.” Surgeons decreased validity scores for 5 standards, and 1 standard shifted from “Valid” to “Uncertain.” In comparison, nonsurgeons increased validity scores for 32 standards; however, all 32 of these were already considered “Valid” in Round 1 ratings and therefore were not recategorized. Nonsurgeons decreased validity scores for 32 standards; however, only 1 of these 32 shifted from “Valid” to “Uncertain.” In sum, changes in surgeons’ ratings between Round 1 and Round 2 resulted in 24 standards being recategorized from “Uncertain” to “Valid,” while nonsurgeons’ ratings resulted in only 1 standard recategorized from “Valid” to “Uncertain.” More detailed representation of the change in validity scores for surgeon versus nonsurgeon and provider versus nonprovider groups can be found in Figures S1 and S2.

DISCUSSION

The CQGS hospital standards for geriatric surgery were developed using a modified RAND-UCLA Appropriateness Methodology, as applied across a large stakeholder group of 58 organizations, conducted by three parallel subgroups and including non-health care providers (such as patient-family organizations, advocacy groups, payers, and regulatory organizations). There were differences observed between surgeons and nonsurgeons, but little difference between ratings across all health care providers compared with nonproviders. Surgeons rated the standards more critically than nonsurgeons (i.e., more uncertainty in validity and feasibility); however, surgeons also tended to increase their ratings more from Round 1 to Round 2. Group discussions focused on areas of uncertainty, with stakeholders expressing a range of opinions and explaining their reasoning for different ratings. Presumably surgeons were willing to change their minds after hearing an alternative point of view. For example, surgeons’ ratings increased from “Uncertain” to “Valid” for the role of a preoperative multidisciplinary conference for elective surgery among

high-risk older patients. This may reflect new insight based upon input from stakeholder groups with a different perspective (e.g., nursing, geriatrics, patient-family experience). The end result was that both surgeons and nonsurgeons rated the validity of the standards highly (≥ 98 percent), although differences persisted regarding feasibility. Despite conducting the RAM panel in three parallel, separate discussion groups, they came to similar conclusions (ranging from 96.8 to 100 percent of standards rated as valid), reinforcing the common perception that mixed-panels are important to generate robust results.

The inclusion of patient-family perspectives in the CQGS is a novel adaptation and critical for advancing toward patient-centered care. Although patients are not traditionally included in a RAM process, a recent study included underserved communities and patient-family advocacy organizations to provide recommendations for appropriate use and access to ancillary health care services (Escaron et al. 2016). In an alternative approach, Stelfox et al. (2013) proposed a protocol engaging patients and families in focus groups prior to conducting the RAM process for patient-centered quality indicators among critically injured patients. Although a qualitative analysis of the RAM discussions is outside of the scope of this paper, there were multiple instances in which the “patient voice” shifted the conversation to better consider the patient’s experience of care. Standards concerning the preoperative goals and decision-making process were heavily influenced by patient-family stakeholders, emphasizing that the decision to undergo surgery must first align with the patient’s goals and values to meet any criteria for a successful outcome. The patient-family representatives expressed a need for better transparency in multidisciplinary care committees and encouraged establishment of patient-family advisory councils, both incorporated as standards. These standards are well aligned with ongoing national efforts to promote increased patient activation, such as the Choosing Wisely[®] campaign. Involvement of patient-family representatives as stakeholders within the CQGS discussion subgroups provided an invaluable contribution: the ability to refocus the conversation on the patient. The current project and aforementioned studies underscore the importance of including patient-family voices for their unique viewpoint and insight into the patient experience of care.

Prior studies have shown rating differences across medical specialties, with proceduralists rating appropriateness higher if they perform a given procedure. In an analysis of five separate nine-member panels, Kahan et al. (1996) found that specialists who perform a given procedure (e.g., angiography, percutaneous transluminal coronary angioplasty, or coronary artery

bypass graft surgery) were more likely to rate the indication as appropriate or necessary compared with a primary care provider. This tendency informs the preference for multispecialty panels rather than reliance on experts from a single discipline (Coulter, Adams, and Shekelle 1995). Many of the CQGS standards were processes of care, often dependent upon the participation of a multidisciplinary health care team. In our study, it may be that interdisciplinary nonsurgeons more routinely use multidisciplinary practices in their daily clinical activities and—as such—are likely to rate the validity and feasibility of the standards higher. Although perioperative decision making is often interdisciplinary, surgical training follows a strict hierarchy and the attending surgeon is often considered the leader in the operating room. Although surgery is increasingly viewed as a team sport, with coordinated roles for anesthesia, surgery, and nursing, it is possible that a traditional, hierarchical view of surgical care may influence surgeons to rate interdisciplinary care processes lower. If surgeons, in fact, use these interdisciplinary care processes less in their daily practice, then the current study findings could be interpreted as similar to those observed in prior studies: just as the nonproceduralists tended to rate procedure indications lower, the surgeons tend to rate processes requiring interdisciplinary coordination lower.

Next steps for the standards, particularly given some surgeons' reluctance toward interdisciplinary care, must include pilot testing their implementation. The original 308 standards were necessarily subdivided into specific elements for the purpose of the formal rating process; however, their content will be revised and repackaged to facilitate practical implementation. For example, there were 6 standards describing an advanced directives discussion, itemized as cardiopulmonary resuscitation, mechanical ventilation, feeding tube access, hemodialysis, anticoagulant use, and blood transfusions. Revision for the purpose of pilot testing combines these elements so the advanced directive discussion constitutes a single standard. The CQGS anticipates the standards will be culled to a much lower number as the project moves forward. Given the multidisciplinary nature of the CQGS project, the standards are so-called "living document," requiring much input and revision. The pilot began in 2017, starting with an alpha phase in which end-user hospitals provide feedback on likely challenges and relative value of the CQGS standards in context of the hospital's environment and patient population. During the beta phase, these hospitals will fully implement the standards in their daily workflow, and external site visits by peer reviewers will help clarify issues of feasibility and areas for improvement. The pilot will allow the CQGS an in-depth evaluation to determine whether differences in surgeons and nonsurgeons perceived

feasibility ratings affect implementation. Lessons from the pilot will inform future dissemination and adoption of the standards. Certainly, we anticipate a range of reactions, from early adopters to resisters. However, the CQGS standards align well with several nonfinancial motivators of behavior change, such as promoting goal-concordant patient-centric care (i.e., “social purpose”) and delivery of consistent high-quality care processes (i.e., “mastery”; Phillips-Taylor and Shortell 2016). Future work will analyze barriers and facilitators of real-world implementation of the CQGS standards and explore perceptions of the CQGS standards among providers, nonproviders, surgeons, and non-surgeons. These ongoing efforts may identify strategies to implement quality improvement in surgical care of older adults.

The current health care reimbursement landscape has been consistently shifting away from the fee-for-service, volume-based model toward one that is based upon quality metrics. The Medicare Access and CHIP Reauthorization Act of 2015 continues the push toward value-based reimbursement. With so-called “bundled care” on the horizon, coordination across multidisciplinary health care professionals becomes even more important. In the older surgical population, coordination of care is critical for decreasing readmission rates and improving postoperative outcomes. The standards established by the CQGS will be made publicly available to all hospitals and, for those interested in developing a formal geriatric surgery center, may be incorporated into a more comprehensive “quality program,” similar to existing programs (e.g., trauma accreditation by the American College of Surgeons’ Committee on Trauma or cancer centers through the Commission on Cancer). Such quality programs not only establish evidence-based standards for care, they also outline resources required for the implementation of those standards, create a database for tracking quality indicators and patient outcomes, and verify delivery of high-quality care through a process of in-person site visits by third-party external peer-review. The RAM is a critical tool to establish a shared vision for high-quality care of older adults given the multidisciplinary nature of geriatric surgical care. Utilizing RAM can help identify processes of care perceived as important across specialties. This study indicates that involvement of multiple stakeholders is possible and can be conducted across a large group, allowing for engagement of different voices and arriving at consistent conclusions across subgroups.

This study has several limitations. First, the RAM process has never been applied to such a large group and its generalizability to other large coalitions is unknown. Because of the diverse multidisciplinary representation,

there were insufficient numbers to represent each specialty in each group. The logistic challenges of conducting large RAM panels may limit future efforts to replicate such studies. However, we believe the importance of engaging diverse stakeholders in improving surgical care of the older adult requires the inclusions of all participants in this particular study. Second, the inclusion of patient-family advocacy groups and other nonproviders (e.g., regulatory agencies) into the RAM panel could be considered both a limitation and a strength. From a strictly traditional viewpoint, critics may question the medical and surgical knowledge of such stakeholders. However, we categorically consider inclusion of the patient voice a strength. The need to address patient experience of care and patient engagement has gained prominence in national health care priorities and cannot be ignored. We believe the contributions and engagement from these stakeholder groups have been invaluable in this process and must be considered for future RAM processes. Third, the definition of feasibility in the current study is highly subjective. Certainly, feasibility will vary by the perceived cost and effort of implementing changes, balanced against the anticipated benefit. There is currently an ongoing pilot evaluation of the standards' feasibility. Fourth, the cost of implementation is as yet undetermined. There are ongoing efforts to determine both the anticipated costs of such an effort and the potential return on investment (e.g., by decreasing rates of postoperative delirium). Finally, and importantly, the direct relationship between implementation of these standards and improvement in outcomes is an untested hypothesis implicit in the CQGS project. The development of this project is still in its early phases; however, future work will strive to examine this fundamental question with a thoughtfully designed multifaceted implementation analysis.

CONCLUSION

In this study, a diverse group of stakeholders—the Coalition for Quality in Geriatric Surgery—has engaged in a modified RAM process to define high-quality standards of surgical care for older adults. This study demonstrates that the RAM can be applied to a large, interdisciplinary group, with the inclusion of patient and family advocacy groups, payers, and regulatory agencies. Despite differences across health professionals (surgeons vs. nonsurgeons), three parallel, multidisciplinary subgroups arrived at similar conclusions on the validity and feasibility of the proposed standards.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Funding for the CQGS project was provided by a grant from the John A. Hartford Foundation. The funder did not influence the interpretation of findings or decision to publish.

The CQGS leadership committee would like to thank the CQGS stakeholder organizations and their designated representatives, which are detailed below, for their contributions to this project.

- 1 American Association of Retired Persons, Susan Reinhard, RN, PhD, FAAN
- 2 ACS Advisory Council for Rural Surgery, Phillip Caropreso, MD, FACS
- 3 ACS Committee on Surgical Palliative Care, Anne Mosenthal, MD, FACS
- 4 Aetna, Joseph Agostini, MD
- 5 The Society for Post Acute and Long Term Care Medicine (“AMDA”), Naushira Pandya, MD, CMD, FACP
- 6 American Academy of Ophthalmology, Hilary Beaver, MD
- 7 American Academy of Orthopaedic Surgeons/American Association of Orthopaedic Surgeons, Timothy Brox, MD
- 8 American Academy of Otolaryngology, Kourosh Parham, MD, PhD, FACS
- 9 American Academy of Physical Medicine and Rehabilitation, Richard Zorowitz, MD
- 10 American College of Emergency Physicians, Ula Hwang, MD
- 11 American College of Physicians, Cynthia Smith, MD, FACP
- 12 American Geriatrics Society, Nancy Lundebjerg, MPA
- 13 American Hospital Association, Health Research & Educational Trust, Ken Anderson, DO, MS, CPE
- 14 American Society of Anesthesiologists, Stacie Deiner, MD, and Mark Neuman, MD, PhD
- 15 American Society of Consultant Pharmacists, Sharon Clackum, PharmD, CGP, CDM, FASCP
- 16 American Society of Health-System Pharmacists, Deborah Pasko, PharmD and Kasey Thompson, PharmD, MS, MBA
- 17 American Society of PeriAnesthesia Nurses, Vallire Hooper, PhD, RN, CPAN, FAAN
- 18 American Urological Association, Peter Albertsen, MD

- 19 Association of periOperative Registered Nurses, Susan Bakewell, MS, RN-BC
- 20 Association of Veterans Affairs Surgeons, Jason Johanning, MD, MS, FACS
- 21 Case Management Society of America, Cheri Lattimer
- 22 Center to Advance Palliative Care, Diane Meier MD, FACP, and Brynn Bowman
- 23 Centers for Medicare and Medicaid Services, Shari Ling, MD
- 24 Eastern Association for the Surgery of Trauma, James Forrest Calland, MD, FACS
- 25 Family Caregiver Alliance, Kathleen Kelly, MPA
- 26 Florida Hospital Association, Martha DeCastro, MS, RN
- 27 Geriatrics for Specialists Initiative, William Lyons, MD
- 28 Gerontological Advanced Practice Nurses Association, Julie Stanik-Hutt, PhD, CRNP
- 29 Hartford Institute for Geriatric Nursing and Nurses Improving Care for Healthsystem Elders, Tara Cortes, PhD, RN
- 30 Hospital Elder Life Program, Sharon Inouye, MD, MPH
- 31 Kaiser Permanente, Laurel Imhoff, MD, FACS
- 32 Memorial Sloan Kettering Cancer Center, Larissa Temple, MD, FACS
- 33 National Association of Social Workers, Chris Herman, MSW, LICSW
- 34 National Committee for Quality Assurance, Judy Ng and Erin Giovannetti, PhD
- 35 National Gerontological Nursing Association, Elizabeth Tanner, PhD, RN, and Joanne Alderman, MS-N, APRN-CNS, RN-BC, FNGNA
- 36 Nurses Improving Care for Healthsystem Elders, Barbara Bricoli, MPA, and Holly Brown, MSN
- 37 Patient and Family Centered Care Partners, Libby Hoy
- 38 Patient Priority Care, Mary Tinetti, MD
- 39 Pharmacy Quality Alliance, Woody Eisenberg, MD
- 40 University of Pennsylvania Center for Perioperative Outcomes Research and Transformation, Sushila Murthy, MD
- 41 Society for Academic Emergency Medicine, Teresita Hogan, MD
- 42 Society of Critical Care Medicine, Jose Diaz, MD, FCCM
- 43 Society of Hospital Medicine, Melissa Mattison, MD, FACP, SFHM

- 44 Society of General Internal Medicine, Lee Lindquist, MD, MPH, MBA
- 45 The American Association for the Surgery of Trauma, Robert Bar-raco, MD, FACS, and Zara Cooper, MD, FACS
- 46 The American Board of Surgery, Mark Malangoni, MD, FACS
- 47 The American Congress of Obstetricians and Gynecologists, Holly Richter, MD
- 48 The Beryl Institute, Jason Wolf, PhD
- 49 The Gerontological Society of America
- 50 The John A. Hartford Foundation, Marcus Escobedo, MPA
- 51 The Society of Thoracic Surgeons, Joseph Cleveland, MD
- 52 University of Alabama – The Division of Gerontology, Geriatrics and Palliative Care, Kellie Flood, MD
- 53 University of Chicago – the MacLean Center for Clinical Medical Ethics, Peter Angelos, MD, FACS
- 54 University of Colorado, Division of Health Care Policy and Research, Eric Coleman, MD, MPH
- 55 University of Pittsburgh Medical Center, Daniel Hall, MD
- 56 University of Wisconsin School of Medicine & Public Health, Gretchen Schwarze, MD, MPP, FACS
- 57 US Department of Veterans Affairs, Geriatrics and Extended Care, Thomas Edes, MD, MS
- 58 Yale New Haven Patient Experience Council, Michael Bennick, MD

Disclosures: Emily Finlayson is founding shareholder Ooney, Inc. No additional disclosures for the remaining authors.

Disclaimer: None.

REFERENCES

- Bentrem, D. J., M. E. Cohen, D. M. Hynes, C. Y. Ko, and K. Y. Bilimoria. 2009. "Identification of Specific Quality Improvement Opportunities for the Elderly Undergoing Gastrointestinal Surgery [Erratum Appears in Arch Surg. 2010 Mar; 145 (3):225]." *Archives of Surgery* 144 (11): 1013–20.
- Berian, J. R., R. A. Rosenthal, T. L. Baker, J. Coleman, E. Finlayson, M. R. Katlic, S. A. Lagoo-Deenadayalan, V. L. Tang, T. N. Robinson, C. Y. Ko, and M. M. Russell. 2017. "Hospital Standards to Promote Optimal Surgical Care of the Older Adult: A Report from the Coalition for Quality in Geriatric Surgery." *Annals of Surgery* 267: 280–90.

- Broder, M. S., D. E. Kanouse, B. S. Mittman, and S. J. Bernstein. 2000. "The Appropriateness of Recommendations for Hysterectomy." *Obstetrics and Gynecology* 95 (2): 199–205.
- Brook, R. 1994. "The RAND/UCLA Appropriateness Method." In *Clinical Practice Guideline Development: Methodology Perspectives*, edited by K. McCormick, S. Moore, and R. Siegel, pp. 59–67. Rockville, MD: US Dept Health and Human Services, Public Health Service; Agency for Health Care Policy and Research.
- Chow, W. B., R. P. Merkow, M. E. Cohen, K. Y. Bilimoria, and C. Y. Ko. 2012a. "Association Between Postoperative Complications and Reoperation for Patients Undergoing Geriatric Surgery and the Effect of Reoperation on Mortality." *American Surgeon* 78 (10): 1137–42.
- Chow, W. B., R. A. Rosenthal, R. P. Merkow, C. Y. Ko, and N. F. Esnaola. 2012b. "Optimal Preoperative Assessment of the Geriatric Surgical Patient: A Best Practices Guideline from the American College of Surgeons National Surgical Quality Improvement Program and the American Geriatrics Society." *Journal of the American College of Surgeons* 215 (4): 453–66.
- Coulter, I., A. Adams, and P. Shekelle. 1995. "Impact of Varying Panel Membership on Ratings of Appropriateness in Consensus Panels: A Comparison of a Multi- and Single Disciplinary Panel." *Health Services Research* 30 (4): 577–91.
- Escaron, A. L., R. Chang Weir, P. Stanton, S. Vangala, T. R. Grogan, and R. M. Clarke. 2016. "Testing an Adapted Modified Delphi Method: Synthesizing Multiple Stakeholder Ratings of Health Care Service Effectiveness." *Health Promotion Practice* 17 (2): 217–25.
- Etzioni, D. A., J. H. Liu, M. A. Maggard, and C. Y. Ko. 2003a. "The Aging Population and Its Impact on the Surgery Workforce." *Annals of Surgery* 238 (2): 170–7.
- Etzioni, D. A., J. H. Liu, J. B. O'Connell, M. A. Maggard, and C. Y. Ko. 2003b. "Elderly Patients in Surgical Workloads: A Population-Based Analysis." *American Surgeon* 69 (11): 961–5.
- Finlayson, E., Z. Fan, and J. D. Birkmeyer. 2007. "Outcomes in Octogenarians Undergoing High-Risk Cancer Operation: A National Study." *Journal of the American College of Surgeons* 205 (6): 729–34.
- Fitch, K., S. J. Bernstein, M. D. Aguilar, B. Burnand, and J. R. LaCalle. 2001. "The RAND/UCLA Appropriateness Method User's Manual" [accessed on January 25, 2016, 2001]. Available at https://www.rand.org/content/dam/rand/pubs/monograph_reports/2011/MR1269.pdf
- Frampton, S. B., S. Guastello, L. Hoy, M. Naylor, S. Sheridan, and M. Johnston-Fleece. 2017. *Harnessing Evidence and Experience to Change Culture: A Guiding Framework for Patient and Family Engaged Care*. Washington, DC: National Academy of Medicine Perspectives.
- "Guide to Patient and Family Engagement in Hospital Quality and Safety" [accessed on April 20, 2017]. Available at <http://www.ahrq.gov/professionals/systems/hospital/engagingfamilies/guide.html>
- Kahan, J. P., R. E. Park, L. L. Leape, S. J. Bernstein, L. H. Hilborne, L. Parker, C. J. Kamberg, D. J. Ballard, and R. H. Brook. 1996. "Variations by Specialty in

- Physician Ratings of the Appropriateness and Necessity of Indications for Procedures." *Medical Care* 34 (6): 512–23.
- Lawson, E. H., M. M. Gibbons, A. M. Ingraham, P. G. Shekelle, and C. Y. Ko. 2011. "Appropriateness Criteria to Assess Variations in Surgical Procedure Use in the United States." *Archives of Surgery* 146 (12): 1433–40.
- Lawson, E. H., M. M. Gibbons, C. Y. Ko, and P. G. Shekelle. 2012. "The Appropriateness Method has Acceptable Reliability and Validity for Assessing Overuse and Underuse of Surgical Procedures." *Journal of Clinical Epidemiology* 65 (11): 1133–43.
- Leape, L. L., M. D. Bernstein, C. J. Kamberg, M. Sherwood, and M. D. Brook. 1993. "The Appropriateness of use of Coronary Artery Bypass Graft Surgery in New York State." *Journal of the American Medical Association* 269 (6): 753–60.
- Maggard, M. A., M. L. McGory, and C. Y. Ko. 2006. "Development of Quality Indicators: Lessons Learned in Bariatric Surgery." *American Surgeon* 72 (10): 870–4.
- Makary, M. A., D. L. Segev, P. J. Pronovost, D. Syin, K. Bandeen-Roche, P. Patel, R. Takenaga, L. Devgan, C. G. Holzmüller, J. Tian, and L. P. Fried. 2010. "Frailty as a Predictor of Surgical Outcomes in Older Patients." *Journal of the American College of Surgeons* 210 (6): 901–8.
- McGory, M. L., P. G. Shekelle, and C. Y. Ko. 2006. "Development of Quality Indicators for Patients Undergoing Colorectal Cancer Surgery." *Journal of the National Cancer Institute* 98 (22): 1623–33.
- McGory, M. L., P. G. Shekelle, L. Z. Rubenstein, A. Fink, and C. Y. Ko. 2005. "Developing Quality Indicators for Elderly Patients Undergoing Abdominal Operations." *Journal of the American College of Surgeons* 201 (6): 870–83.
- McGory, M. L., K. K. Kao, P. G. Shekelle, L. Z. Rubenstein, M. J. Leonardi, J. A. Parikh, A. Fink, and C. Y. Ko. 2009. "Developing Quality Indicators for Elderly Surgical Patients." *Annals of Surgery* 250 (2): 338–47.
- Mohanty, S., R. A. Rosenthal, M. M. Russell, M. D. Neuman, C. Y. Ko, and N. F. Esnaola. 2016. "Optimal Perioperative Management of the Geriatric Patient: A Best Practices Guideline from the American College of Surgeons NSQIP and the American Geriatrics Society." *Journal of the American College of Surgeons* 222 (5): 930–47.
- National Priorities Partnership. 2008. *National Priorities and Goals: Aligning Our Efforts to Transform America's Healthcare*. Washington, DC: National Quality Forum.
- Park, R. E., A. Fink, R. H. Brook, M. R. Chassin, K. L. Kahn, N. J. Merrick, J. Koseoff, and D. H. Solomon. 1986. "Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures." *American Journal of Public Health* 76 (7): 766–72.
- "Patient-Centered Outcomes Research Institute" [accessed on December 28, 2016]. Available at <http://www.pcori.org/>
- Phillips-Taylor, M., and S. M. Shortell. 2016. "More Than Money: Motivating Physician Behavior Change in Accountable Care Organizations." *Milbank Quarterly* 94: 832–61.
- "RAND-UCLA Appropriateness Method Online Supplementary Methods Section: Analysis of Questionnaire Results."

- Robinson, T. N., J. I. Wallace, D. S. Wu, A. Wiktor, L. F. Pointer, S. M. Pfister, T. J. Sharp, M. J. Buckley, and M. Moss. 2011. "Accumulated Frailty Characteristics Predict Postoperative Discharge Institutionalization in the Geriatric Patient." *Journal of the American College of Surgeons* 213 (1): 37–42; discussion 42–4.
- Shekelle, P. G., J. P. Kahan, S. J. Bernstein, L. L. Leape, C. J. Kamberg, and R. E. Park. 1998. "The Reproducibility of a Method to Identify the Overuse and Underuse of Medical Procedures." *New England Journal of Medicine* 338 (26): 1888–95.
- Stelfox, H. T., J. M. Boyd, S. E. Straus, and A. R. Gagliardi. 2013. "Developing a Patient and Family-Centred Approach for Measuring the Quality of Injury Care: A Study Protocol." *BMC Health Services Research* 13: 31.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.

Appendix SA2:

Appendix S1. Key Terms and List of 308 Standards.

Figure S1. Change in Validity Ratings Between Round 1 and Round 2 for Surgeons vs. Non-Surgeons.

Figure S2. Change in Validity Ratings Between Round 1 and Round 2 for Providers vs. Non-Providers.