Title
Disparity-Sensitive Measures in Surgical Care: A Delphi Panel Consensus.

Journal
Journal of the American College of Surgeons, 236(1)

ISSN
1072-7515

Authors
de Jager, Elzerie
Osman, Samia Y
Levine, Adele A
et al.

Publication Date
2023

DOI
10.1097/xcs.0000000000000420

HHS Public Access

Author manuscript

J Am Coll Surg. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:

J Am Coll Surg. 2023 January 01; 236(1): 135–143. doi:10.1097/XCS.0000000000000420.

Disparity-Sensitive Measures in Surgical Care: A Delphi Panel Consensus

Elzerie de Jager, MBBS(Hons), PhD[1,2], **Samia Y Osman, MD, MPP**[1], **Adele A Levine, MPH**[1], **Charles Liu, MD, MS**[3], **Melinda A Maggard Gibbons, MD, MHSH, FACS**[4], **Clifford Y Ko, MD, MS, MSHS, FACS**[4,5,6], **Helen R Burstin, MD, MPH**[7], **Adil H Haider, MD, MPH, FACS**[1], **David B Hoyt, MD, FACS**[5], **Andrew J Schoenfeld, MD, MSc**[1], **LD Britt, MD, MPH, FACS**[8], **Joel S Weissman, PhD**[1]

[1]Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School and Harvard T.H. Chan School of Public Health, Boston, MA

[2]College of Medicine and Dentistry, James Cook University, Townsville, QLD, AUS

[3]Department of Surgery, Stanford University School of Medicine, Stanford, CA

[4]Department of Surgery, David Geffen School of Medicine at University of California, Los Angeles, CA

[5]American College of Surgeons, Chicago, IL

[6]Department of Surgery, VA Greater Los Angeles Healthcare System, Los Angeles, CA

[7]Council of Medical Specialty Societies, Chicago, IL

[8]Department of Surgery, Eastern Virginia Medical School, Norfolk, VA

Abstract

Background—In the United States, disparities in surgical care impede the delivery of uniformly high-quality care to all patients. There is a lack of disparity sensitive measures related to surgical care. The American College of Surgeons Metrics for Equitable Access and care in SURgery group, through research and expert consensus, aimed to identify disparity-sensitive measures in surgical care.

Study Design—An environmental scan, systematic literature review and subspecialty society surveys were conducted to identify potential disparity-sensitive surgical measures. A modified Delphi process was conducted where panelists rated measures on both importance and validity. In addition to this a novel literature-based disparity-sensitive scoring (DSS) process was used.

Correspondence address: Elzerie de Jager, MBBS(Hons), PhD, Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School and Harvard T.H. Chan School of Public Health, Boston, MA, One Brigham Circle, 1620 Tremont Street, Suite 2-016, Boston, MA 02120, (elzerie.dejager@my.jcu.edu.au), 617-525-7300.

Disclosure Information: Nothing to disclose.

Disclosures outside the scope of this work: Dr Schoenfeld receives royalty payments from Wolters Kluwer and Springer is an editorial board member at *The Journal of Bone and Joint Surgery* and is the Editor-in-Chief of *Spine*. Other authors have nothing to disclose.

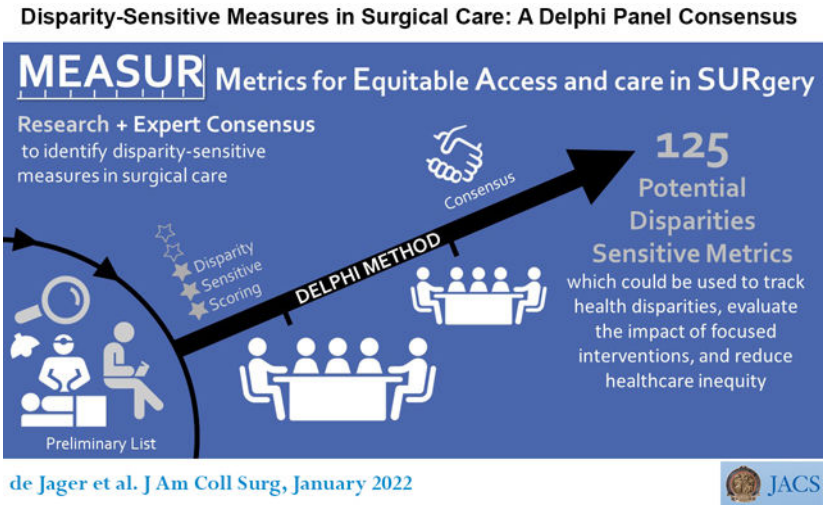
Results—We identified 841 potential disparity-sensitive surgical measures. From these, our Delphi and literature-based approaches yielded a consensus list of 125 candidate disparity-sensitive measure. These measures were rated as both valid and important and were supported by the existing literature.

Conclusion—There are profound disparities in surgical care within the United States healthcare system. A multidisciplinary Delphi panel identified 125 potential disparity-sensitive surgical measures which could be used to track health disparities, evaluate the impact of focused interventions, and reduce healthcare inequity.

Precis:

The American College of Surgeons Metrics for Equitable Access and Care in Surgery group aimed to identify disparity-sensitive measures in surgical care. A Delphi panel identified 125 potential disparity-sensitive measures which could be used to track health disparity, evaluate the impact of focused interventions, and reduce healthcare inequity.

Graphical Abstract



Keywords

Surgery; Disparities; Quality measures; Equity measurement; Disparity sensitive measures

Introduction

Social risk factors, such as socioeconomic status, rurality, race, and ethnicity play a major role in healthcare delivery.[1,2] In the United States, socially at-risk patients receive lower quality healthcare and experience worse health outcomes when compared to other patients.[1,2] These disparities in healthcare, including surgical care, represent one of the greatest challenges in achieving uniformly high-quality care.[3–5]

An underused tool for advancing health equity is performance measurement.[6] In 2020, the Office for the Assistant Secretary for Planning and Evaluation recommended that

the Centers for Medicare and Medicaid Services include measures of health equity in public reporting and value-based purchasing programs.[1,2] Measures of health equity, or disparity-sensitive measures, assess were defined as measures "…that accurately reflect disparities in surgical access or quality for a healthcare organization, and where the health care organization can reasonably be expected to influence important factors related to its performance." The National Quality Forum (NQF) have developed criteria for identifying such measures.[6,7] However, these criteria have not identified any disparity-sensitive measures related to surgical care.

To identify a suite of disparity-sensitive surgical measures, the American College of Surgeons Metrics for Equitable Access and care in SURgery (ACS MEASUR) group was formed in collaboration with Eastern Virginia Medical School, Brigham and Women's Hospital, the NQF, and the University of California – Los Angeles.[8] A modified Delphi process was used to ascertain measures that represented disparities in high quality surgical care.[9–16] The purpose of this study was to identify a suite of measures that would be tested empirically for evidence of disparities using real world evidence. This suite of metrics could then be used to guide quality improvement efforts and facilitate the development of targeted interventions aimed at eradicating surgical disparities.

Methods

Study design and participants

The UCLA-RAND modified Delphi method is an established technique to elicit quasi-anonymous, structured opinions from a multidisciplinary group of experts with the goal of achieving consensus.[9–16] In this context quasi-anonymous means that the panelists know one another, but only study staff will be able to see their opinions and judgment to enable analysis. Delphi panelists were invited from the ACS Health Disparities Standing Committee, the NQF Disparities Standing Committee, the National Institute of Health (NIH)-ACS Surgical Disparities Symposium attendees, and various non-profit, academic, and medical institutions. Seventeen panelists were selected and 13 agreed to participate in the study; this included representatives from national professional and quasi-public organizations including researchers, surgeons, other physicians, economists, historians with expertise in race in the United States, medical ethicists, and health policy experts. A $500 honorarium was offered to each panelist, as well as covering all travel costs for in-person meetings. The Delphi process was conducted between May 2018 and October 2020. The project received ethics approval from the Partners Healthcare Institutional Review Board (2017P002831) and was considered IRB exempt at Eastern Virginia Medical School(18–05-NH-0145) (LD Britt).

A modified Delphi process was conducted where panelists rated measures on both importance and validity. In addition, a novel literature-based disparity-sensitive scoring (DSS) process was used. Responses were collected from our expert panel with quasi-anonymity, fed back their individual responses so they knew how their opinion related to the groups, and used statistical analysis techniques to interpret the data.

Identifying potential measures

We identified a suite of potential disparity-sensitive measures by performing: 1) an environmental scan of existing surgical quality measures 2) a systematic literature review on surgical access[8]; and a survey of subspecialty surgical societies. The environmental scan included measures from peer-reviewed and grey literature, the NQF's portfolio of endorsed measures, the Agency for Healthcare Research and Quality's National Quality Measures Clearinghouse and National Guidelines Clearinghouse, and the Centers for Medicare and Medicaid Services Measures Inventory. The systematic literature review employed a broad PubMed search strategy, with the following terms: *healthcare disparities, health status disparities* and *surgery* (2008–2018). [8] A total of 1,375 original articles were screened and individual measures of surgical access disparities were extracted. A measure was defined as a quantitative primary or secondary study end point that demonstrated a statistically significant disparity.[8]

Thirteen surgical societies selected by our Expert Advisory Panel were surveyed: American Surgical Association, American Society of Anesthesiologists, American Society of Colon and Rectal Surgeons, American Association of Endocrine Surgeons, General Surgery, Society of Gynecologic Oncology, American College of Obstetrics and Gynecologists, American Academy of Ophthalmology, American Academy of Orthopedic Surgeons, American Thoracic Society, American Association for the Society of Trauma, Society for Vascular Surgery, and the American Urological Association. Measures suggested by the surgical societies were included in the suite of potential disparity-sensitive measures.

Recognizing the importance of addressing disparities across the continuum of surgical care, measures were categorized into the ACS five phases of surgical care: pre-operative, intra-operative, peri-operative, post-operative, post-discharge. A relative lack of access to surgical services may be a contributing factor to surgical disparities and as such surgical access was added as the sixth phase of surgical care.[5]

Importance rating

Before the first Delphi round, panelists were provided with an abbreviated summary of each measure's specifications and asked to rate Importance on a 5-point Likert scale after reading the following prompt: "In your expert opinion, how important is this metric in identifying or addressing disparities in surgical access or quality?" Since we included surgeons and non-surgeons, the question required a minor clarification to capture the appropriate perspective.

- For surgeons: When thinking about importance, consider how meaningful this measure is to you as a surgeon and how improving performance on this measure would reduce surgical disparities.
- For non-surgeons: When thinking about importance, consider issues like impact on health and how improving performance on this measure would reduce surgical disparities.

An option for "unable to rate" was provided for each measure, and if this option was selected it was not counted in the rating for that measure. The rating process was completed remotely via an online survey platform (Qualtrics). To reduce survey burden, panelists

were split into two groups, each rating one half of the measures. Measures with a mean importance rating for addressing disparities that fell below the 50th percentile were then excluded from further consideration.

Disparity-sensitive scoring description

Measures with a mean importance rating of 50th percentile or higher were then 'scored' using literature-based Disparity-Sensitive Scoring criteria, which was adapted for surgical measures from NQF criteria.[6,7] The MEASUR DSS criteria contained five domains: Quality gap or Access gap (demonstrated disparities in quality or access based on literature), Impact (improving measure performance improves overall quality of care), Prevalence, Discretion (surgeons rely on clinical judgment versus practice guidelines), and Feasibility (improvement can realistically be achieved).

Each measure was reviewed by a member of the study group, and, for each domain, a summary of relevant peer reviewed literature was recorded. Measures were then assigned a point value for each criterion, based on the strength of the supporting literature. The sum of the five criteria was calculated for each metric; this was termed the Disparity-Sensitive Score. Each measure received a minimum score of zero (lack of disparity-sensitivity) and a maximum score of 12 (high disparity-sensitivity). Measures were excluded if they scored a total of less than six, or if they scored a zero on the Quality or Access domain (indicating that studies had been conducted and had found no disparities or gaps in quality of care).

Validity ratings rounds 1 and 2 stages

Panelists were provided with the disparity-sensitive scores and supporting literature summary for the first Delphi rating on validity. The panelists were asked to rate validity based on the materials provided and their own knowledge. Panelists were also asked to consider that a valid disparity sensitive measure must pass two validity bars: measures must be valid measures of quality and valid measures of disparity-sensitivity. To be a valid *measure of quality*, it must be strongly linked to an outcome that is important to patients, or an outcome needed to be linked to prior process. To be a valid *Disparity-Sensitive Measure*, it must accurately reflect disparities in surgical access or quality for a healthcare organization, which can reasonably be expected to influence important factors related to its performance. The following six criteria were provided to the panelists to consider when evaluating the validity of each measure.

1. There exists a large gap in access to or quality of the procedure/treatment in question.
2. The measure has a meaningful impact on patient quality of life and care.
3. The condition or procedure is prevalent in the disparity population.
4. The measure refers to a procedure that may have a high degree of discretion by the surgeon.
5. The measure is feasible to improve at a surgeon or organizational level.
6. The measure's Quality Measure Status.

Panelists were provided with each measure's Quality Measure Status. This is a designation given to measures to distinguish well-vetted measures (quality indicators that are endorsed, specified, or validated by a quality organization) and measure concepts (quality indicators that were included in our potential suite of metrics, but were not endorsed, specified, or validated by a quality organization).

Ratings were conducted using a 9-point Likert scale with an unable to rate option. The rating was completed remotely via an online survey platform (Qualtrics). To reduce the burden of rating, the panelists were split into two groups and were given four weeks to complete the survey.

Panelists' ratings were analyzed using the median rating and the dispersion of panel ratings.[17] The number of panelists rating each measure fluctuated, based on how many panelists selected the 'unable to rate' response option. The median validity ratings were placed into terciles (median rating 1–3, 4–6 and 7–9). Median ratings were rounded up to the nearest whole number (recognizing that this was biased towards making measures valid).

The dispersion of panel ratings is an indicator of the level of disagreement between panelists. Measures with a wide dispersion of ratings indicate that the panelists disagreed. Disagreement was defined as 80% of ratings not falling within the same tercile of the observed mean.[17] The number of ratings required to be in disagreement were rounded up to the nearest whole number (Figure 1).[18]

Measures that were rated not valid (median rating 1–3, not in disagreement) were removed from further consideration, and those rated as uncertain (median rating 4–6, or any median rating in disagreement) were re-rated in a second Delphi round. Measures rated valid (median rating 7–9, not in disagreement) were included in the final suite of measures.

The second validity rating round was conducted as an in-person in Dallas, Texas on April 13th, 2019. Participants unable to attend in-person were invited to join by videoconference. In-person meetings allow the 'Estimate-Talk-Estimate' process which allows experts to interact, allowing clarification on matters and justification of viewpoints. Limiting the potential for bias and avoiding the influence of dominant participants in the face-to-face group was achieved by effective group leadership and quasi-anonymous voting procedures.[19] All panelists were invited to discuss and re-rate measures.[20] Panelists were provided with individual rating sheets which depicted each measure's title, Quality Measure Status, DSS, dispersion of validity ratings, their own individual rating, and the panel median rating (Figure 2). To minimize the impact of "groupthink" on biasing the ratings towards a group consensus, panelists re-rated measures independently after discussion.

Of the 13 panelists, six were able to attend the meeting; two panelists attended the meeting in person and four panelists joined remotely by videoconference. To reduce survey burden, panelists were split into two groups (one of six, and one of seven) for the importance rating and for the first validity rating. As such the number of panelists rating the measures in each round was consistent (six-seven panelists).

Panelists' ratings were analyzed on the median rating, and the dispersion of panel ratings, as was done for the first validity rating. Measures that were rated as uncertain or not valid were removed from further consideration.

Results

There were 841 potential disparity-sensitive measures: 350 from the environmental scan, 223 from the literature review, and 268 from the surgical society surveys (Figure 3). This included performance measures, measure concepts, survey instruments, study outcomes demonstrating disparities for a population, and novel surgical society suggestions. Many of the measures had not been specified or validated. Measures spanned all surgical specialties and phases of surgical care (access to care, pre-operative, peri-operative, intra-operative, post-operative, and post-discharge).

Importance ratings

Measures from the environmental scan and literature review (573 measures) were rated by panelists on importance as a disparity measure. Measures suggested by the subspecialty society surveys were all deemed to be important. There were 307 measures excluded: 261 measures with a mean rating below the 50^{th} percentile of 3.8 out of 5, and 46 duplicative measures.

Disparity-sensitive scoring

A total of 534 measures, including 266 measures rated as important, and 268 subspecialty survey recommended measures, were scored using the DSS scoring criteria. There were 83 measures excluded; 59 measures which scored a total DSS less than six, 11 measures which scored a zero on the quality or access domain, and 13 duplicate measures. In addition to these, 22 measures were removed due to being hospital level indicators, population screening measures, non-surgical measures, and survey instruments.

Validity ratings round 1 and round 2

Panelists rated 429 measures on validity; 159 were rated as valid, 270 were rated as uncertain, and no measures were rated as not valid (Figure 4). After discussion, it was determined that the focus of this project was on identifying rather than developing quality measures, and thus 228 measure concepts (quality indicators that are not endorsed, specified, or validated by a quality organization) were removed from further consideration. In addition, four measures examining hospital level characteristics and two duplicates were removed. A total of 195 measures remained; 90 were categorized as valid, and 105 were categorized as uncertain.

There were 105 measures that were re-rated on validity; 35 deemed to be valid and 70 were deemed to be not valid (Figure 5).

Suite of Candidate Disparity-Sensitive Measures

This process generated 125 candidate disparity-sensitive measures: 90 from the first validity rating and 35 from the second validity rating (Supplemental Digital Content 1). Each

of these measures had been rated as both important and valid by the panelists and had peer reviewed literature (DSS) supporting them as potential disparity-sensitive measures. Twenty-two measures spanned all surgical specialties. The remaining measures were surgical specialty specific; cardiothoracic surgery (31), vascular surgery (14), obstetrics and gynecology (13), general (12), gynecologic oncology (11), orthopedic (7), ophthalmology (6), neurosurgery (5), pediatric surgery (3), and colorectal surgery (1). The measures spanned across all phases of surgical care; access to care (13), pre-operative (12), peri-operative (13), intra-operative (5), post-operative (55), post-discharge (24), and multiple phases of care (3).

Discussion

This work identified 125 potential surgical disparity-sensitive measures. Validation of the measures as disparity-sensitive was achieved via expert determination. Prior to these measures being implemented by surgical care providers, health systems or quality organizations, more work is necessary to assess their performance in a clinical context and ascertain their capacity to address existing disparities in surgical care.

Defining validity

We focused on identifying disparity-sensitive measures from existing validated quality measures, rather than developing these measures solely for the purposes of detecting disparities in healthcare. Panelists were asked to ensure selected health equity measures achieved a two-pronged definition; a valid measure of quality and a valid disparity-sensitive measure. However, there may be merit in developing measures for the specific purpose of detecting disparities. A disparity-sensitive metric may not have to be a valid measure of healthcare quality.

Surgical access

In examining only validated measures of quality for disparity-sensitivity most surgical access measures were removed; in the final suite there were only 13 measures that captured surgical access. Numerous studies have found correlations between lack of or delayed access to surgical care, complications[21–25], and mortality.[26–31] The surgical access literature review identified 223 study outcomes demonstrating disparities for a population.[8] There is a need to develop more disparity-sensitive surgical access measures.[32] The current suite of measures is unlikely to address critical issues in the disparity landscape, such as healthcare segregation (i.e., the clustering of minority and patients of low socioeconomic status at lower quality hospitals).

Distribution of measures by surgical specialty

The final suite of measures did not equally represent all surgical specialties. Almost a quarter of the measures selected were in the field of cardiothoracic surgery, whilst there were no measures pertaining to certain surgical specialties (i.e., oral and maxillofacial, otolaryngology, plastic, and urology). Some specialties, such as cardiothoracic surgery, have many well developed and well-studied quality measures, and our Delphi panelists may have

had more familiarity with these. This likely resulted in a bias towards selecting measures from specialties with well-established quality metrics.

Discretionary surgery

One of the six criteria which were provided to panelists to consider when evaluating the validity of each measure was whether the measure refers to a procedure that may have a high degree of discretion by the surgeon. This criterion was established as there was a concern that there tends to be greater disparities in discretionary surgical procedures compared with non-discretionary. However, there may be disparities in non-discretionary surgeries. The final suite of measures relate to both discretionary and non-discretionary surgical procedures, although the suite may be skewed towards discretionary procedures due to the evaluation criteria.

Upstream disparities

Disparities in surgical care provision may stem from upstream disparities, differing access to surgical care and differing quality of surgical care provision. This project examined areas that healthcare organizations have traditionally focused on improving; surgical access and quality. There are potentially additional measures of upstream disparities that healthcare organizations can influence. Future efforts addressing surgical inequities should examine how to identify, measure, and impact environmental influences on disparities in surgical care.

There are many upstream factors that may result in disparities in surgical care. For example, patients with social risk factors tend to have higher rates of medical comorbidities and may have higher surgical risk profiles. Institutional, societal and structural racism may each impact disparities in surgical access and care, such as when underlying segmentation of the insurance market may adversely affect low-income racial and ethnic minority patients.[33] There are also environmental influences on the incidence of surgical disease; for example, socially at-risk patients have higher rates of surgical procedures for firearm related injuries. These upstream disparities must be considered when identifying, implementing, and interpreting the results of disparity sensitive surgical measures.

Next Steps

Going forward, we envision a process whereby the performance of the surgical measures identified here can be assessed to determine their ability to accurately identify healthcare disparities as a springboard to recommending their adoption by quality organizations, societies, healthcare providers and third-party payers. This would necessitate large scale data, that cuts across systems with appropriate levels of clinical and social variation in the substrates used to support such investigation. A final phase would require the assessment of such measures on reducing or eliminating disparities once they are implemented. In parallel, additional work is called for to develop measures that can address other known contributors to disparities in surgical care that are not accounted for in this battery of surgical measures, such as implicit bias, representation in healthcare and issues around hospital-level segregation.

Conclusions

We present the results of a multidisciplinary Delphi panel that identified 125 potential candidates for disparity-sensitive surgical measures. These measures represent a robust battery that could be deployed to track disparities in surgical care across healthcare organizations, evaluate the impact of tailored interventions, and reduce healthcare inequity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the following individuals from the collaborating institutions for their contributions: Zara Cooper, MD, MSc, FACS, N Rhea Udyavar, MD, Esther Moberg, MPH, Christina Sheu, Nizar Bhulani, MD, MPH, Margherita Lamaina, MD, Sarah O'Sullivan, and Fatima Rahim.

We thank the panelists for their enormous contribution and guidance throughout the Delphi process. Members of the panelists are Waddah B. Al-Refaie, MD, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Surgical Oncology; Marshall Chin, MD, MPH, University of Chicago, Healthcare Ethics; Edward Cornwell III, MD, Howard University, Surgery; Bruce Hall, MD, PhD, BJC Healthcare, Washington University School of Medicine, Surgery, Healthcare Management; Romana Hasnain-Wynia, PhD, MS, Denver Health and Hospital Authority; Elliot Haut, MD, PhD, Johns Hopkins Medicine, Surgery; Arden M. Morris, MD, MPH, Stanford University School of Medicine, Surgery; Ninez Ponce, PhD, MPP, UCLA Center for Health Policy Research, UCLA Fielding School of Public Health, Health Policy and Management; Gary Puckrein, PhD, National Minority Quality Forum; Ali Salim, MD, Harvard Medical School, Surgery, Brigham and Women's Hospital, Trauma, Burn, Surgical Critical Care & Emergency General Surgery; Thomas Sequist, MD, MPH, Mass General Brigham, Harvard Medical School, Medicine, Health Care Policy; Girma Tefera, MD, American College of Surgeons Operation Giving Back, University of Wisconsin School of Medicine and Public Health, Vascular Surgery; Sandra Wong, MD, MS, Dartmouth-Hitchcock Medical Center, Surgery.

Specialty entities who contributed are Society for Vascular Surgery, American Society of Anesthesiologists, Society of Gynecologic Oncology, American Academy of Ophthalmologists, Society of Thoracic Surgeons, American Surgical Association, American College of Obstetricians and Gynecologists, American Society of Colon and Rectal Surgeons, American Association of Endocrine Surgeons, American Association for the Surgery of Trauma, and Marie Crandall, MD, MPH, FACS (General Surgery).

Support:

Research reported in this publication was supported by the National Institute on Minority Health and Health Disparities of the National Institute of Health under award number: 5RO1MD011695–02. Dr de Jager was a Doctor of Philosophy candidate at an Australian university, this research degree is supported by an Australian Government Research Training Program (RTP) Scholarship. Drs Osman and Hoyt receive grant support from the American College of Surgeons. Dr Hoyt is the former Executive Director of the American College of Surgeons.

References

1. Office of the Assistant Secretary for Planning and Evaluation USD of H& HS. Second Report to Congress on Social Risk Factors and Performance in Medicare's Value-Based Purchasing Program.; 2020. Accessed July 12, 2021. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-based-purchasing-programs
2. RAND Health Care. Developing Health Equity Measures.; 2021. Accessed June 14, 2021. https://aspe.hhs.gov/social-risk-factors-and-
3. Haider AH, Scott VK, Rehman KA, et al. Racial Disparities in Surgical Care and Outcomes in the United States: A Comprehensive Review of Patient, Provider, and Systemic Factors. Journal of the American College of Surgeons. 2013;216(3):482–492.e12. [PubMed: 23318117]

4. Torain MJ, Maragh-Bass AC, Dankwa-Mullen I, et al. Surgical Disparities: A Comprehensive Review and New Conceptual Framework. Journal of the American College of Surgeons. 2016;223(2):408–418. [PubMed: 27296524]
5. de Jager E, Levine AA, Udyavar NR, et al. Disparities in Surgical Access: A Systematic Literature Review, Conceptual Model, and Evidence Map. Journal of the American College of Surgeons. 2019;228(3):276–298. [PubMed: 30803548]
6. National Quality Forum. A Roadmap for Promoting Health Equity and Eliminating Disparities: The Four I's for Health Equity.; 2017.
7. National Quality Forum. Healthcare Disparities and Cultural Competency Consensus Standards: Disparities-Sensitive Measure Assessment.; 2012. 20
8. de Jager E, Levine AA, Udyavar NR, et al. Disparities in Surgical Access: A Systematic Literature Review, Conceptual Model, and Evidence Map. Journal of the American College of Surgeons. 2019;228(3):276–298. [PubMed: 30803548]
9. Boulkedid R, Abdoul H, Loustau M, et al. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PloS one. 2011;6(6):e20476.
10. Schwartz HP, Bigham MT, Schoettker PJ, et al. Quality Metrics in Neonatal and Pediatric Critical Care Transport. Pediatric Critical Care Medicine. 2015;16(8):711–717. [PubMed: 26181297]
11. Ingraham A, Nathens A, Peitzman A, et al. Assessment of emergency general surgery care based on formally developed quality indicators. Surgery. 2017;162(2):397–407. [PubMed: 28647046]
12. Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG. Metric Development for an Arthroscopic Bankart Procedure: Assessment of Face and Content Validity. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2015;31(8):1430–1440. [PubMed: 26239785]
13. Villafane J, Edwards TC, Diab KA, et al. Development of quality metrics for ambulatory care in pediatric patients with tetralogy of Fallot. Congenital Heart Disease. 2017;12(6):762–767. [PubMed: 28880457]
14. Gottlieb A, Salame N, Armstrong AW, et al. A provider global assessment quality measure for clinical practice for inflammatory skin disorders. Journal of the American Academy of Dermatology. 2019;80(3):823–828. [PubMed: 30244058]
15. Kojima K, Graves M, Taha W, et al. AO international consensus panel for metrics on a closed reduction and fixation of a 31A2 pertrochanteric fracture. Injury. 2018;49(12):2227–2233. [PubMed: 30268512]
16. MacLean CH, Kerr EA, Qaseem A. Time Out — Charting a Path for Improving Performance Measurement. New England Journal of Medicine. 2018;378(19):1757–1761. 21 [PubMed: 29668361]
17. Fitch K, Bernstein S, Aguilar MD, et al. The RAND / UCLA Appropriateness Method User's Manual.; 2001.
18. MacLean CH, Kerr EA, Qaseem A. Time Out — Charting a Path for Improving Performance Measurement. New England Journal of Medicine. 2018;378(19):1757–1761. [PubMed: 29668361]
19. Kahan J, Bernstein S, Leape L, et al. Measuring the necessity of medical procedures. Medical care. 1994;32(4):357–365. [PubMed: 8139300]
20. MacLean CH, Kerr EA, Qaseem A. Time Out — Charting a Path for Improving Performance Measurement. New England Journal of Medicine. 2018;378(19):1757–1761. [PubMed: 29668361]
21. Ibrahim AM Thumma JR Dimick JB. RSE. Emergency Surgery for Medicare Beneficiaries Admitted to Critical Access Hospitals. Ann Surg. 2018;267(3):473–477. [PubMed: 28288068]
22. Moja L, Piatti A, Pecoraro V, et al. Timing matters in hip fracture surgery: patients operated within 48 hours have better outcomes. A meta-analysis and meta-regression of over 190,000 patients. PloS one. 2012;7(10):e46175.
23. Papandria D, Goldstein SD, Rhee D, et al. Risk of perforation increases with delay in recognition and surgery for acute appendicitis. The Journal of surgical research. 2013;184(2):723–729. [PubMed: 23290595]
24. Mahdi H, Lockhart D, Moslemi-Kebria M, Rose PG. Racial disparity in the 30-day morbidity and mortality after surgery for endometrial cancer. Gynecologic oncology. 2014;134(3):510–515. [PubMed: 24905775]

25. Terplan M, Smith EJ, Temkin SM. Race in ovarian cancer treatment and survival: a systematic review with meta-analysis. Cancer causes & control: CCC. 2009;20(7):1139–1150. 22 [PubMed: 19288217]

26. Gonzalez PC, Gauvreau K, Demone JA, et al. Regional racial and ethnic differences in mortality for congenital heart surgery in children may reflect unequal access to care. Pediatric cardiology. 2003;24(2):103–108. [PubMed: 12360393]

27. McIsaac DI, Abdulla K, Yang H, et al. Association of delay of urgent or emergency surgery with mortality and use of health care resources: a propensity score-matched observational cohort study. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2017;189(27):E905–E912.

28. Kim C, Diez Roux A V, Hofer TP, et al. Area socioeconomic status and mortality after coronary artery bypass graft surgery: the role of hospital volume. American heart journal. 2007;154(2):385–390. [PubMed: 17643593]

29. Eaglehouse YL, Georg MW, Shriver CD, Zhu K. Time-to-surgery and overall survival after breast cancer diagnosis in a universal health system. Breast cancer research and treatment. 2019;178(2):441–450. doi:10.1007/s10549-019-05404-8 [PubMed: 31414244]

30. Sabik LM, Bradley CJ. Differences in mortality for surgical cancer patients by insurance and hospital safety net status. Medical care research and review: MCRR. 2013;70(1):84–97. [PubMed: 22951313]

31. Moja L, Piatti A, Pecoraro V, et al. Timing matters in hip fracture surgery: patients operated within 48 hours have better outcomes. A meta-analysis and meta-regression of over 190,000 patients. PloS one. 2012;7(10):e46175.

32. Levine AA, de Jager E, Britt LD. Perspective: Identifying and Addressing Disparities in Surgical Access: A Health Systems Call to Action. Annals of Surgery. 2020;271(3).

33. Yearby R, Clark B, Figueroa. Structural racism in Historical and Modern US Health Care Policy. Health Affairs. 2022; 41(2):187–194. [PubMed: 35130059]

Number of ratings	Disagreement definition
10, 11, 12 or 13	Having 3 or more ratings outside the tercile with the observed median.
6, 7, 8 or 9	Having 2 or more ratings outside the tercile with the observed median.
4 or 5	Having 1 or more ratings outside the tercile with the observed median
3 or fewer	Classed as in disagreement due to insufficient panelist ratings.

Figure 1.
Disagreement parameters for rating validity.

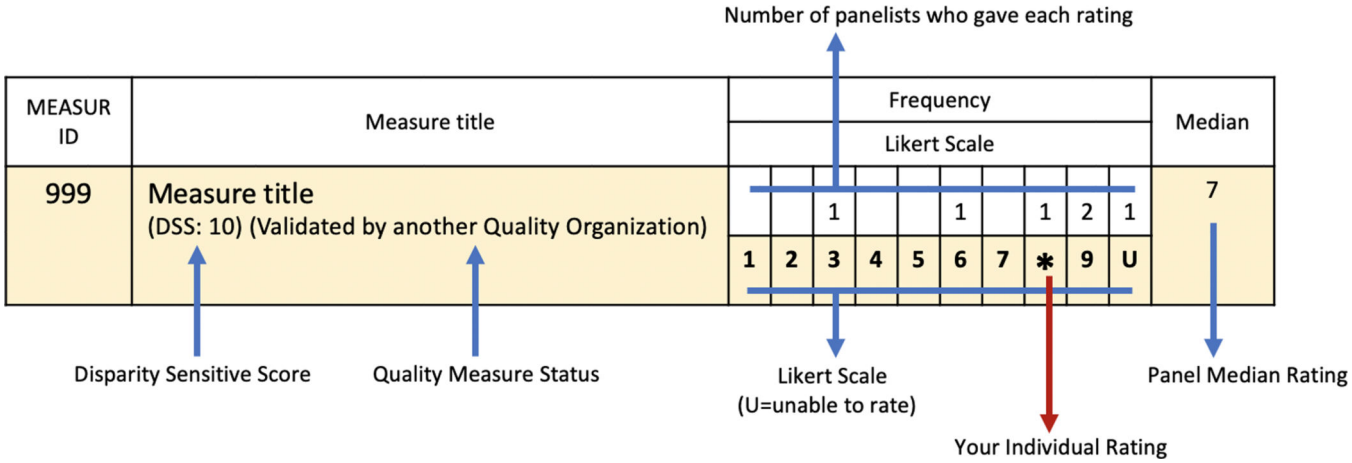


Figure 2.

An example of how the rating sheet information was presented to the panelists. DSS, disparity-sensitive score.

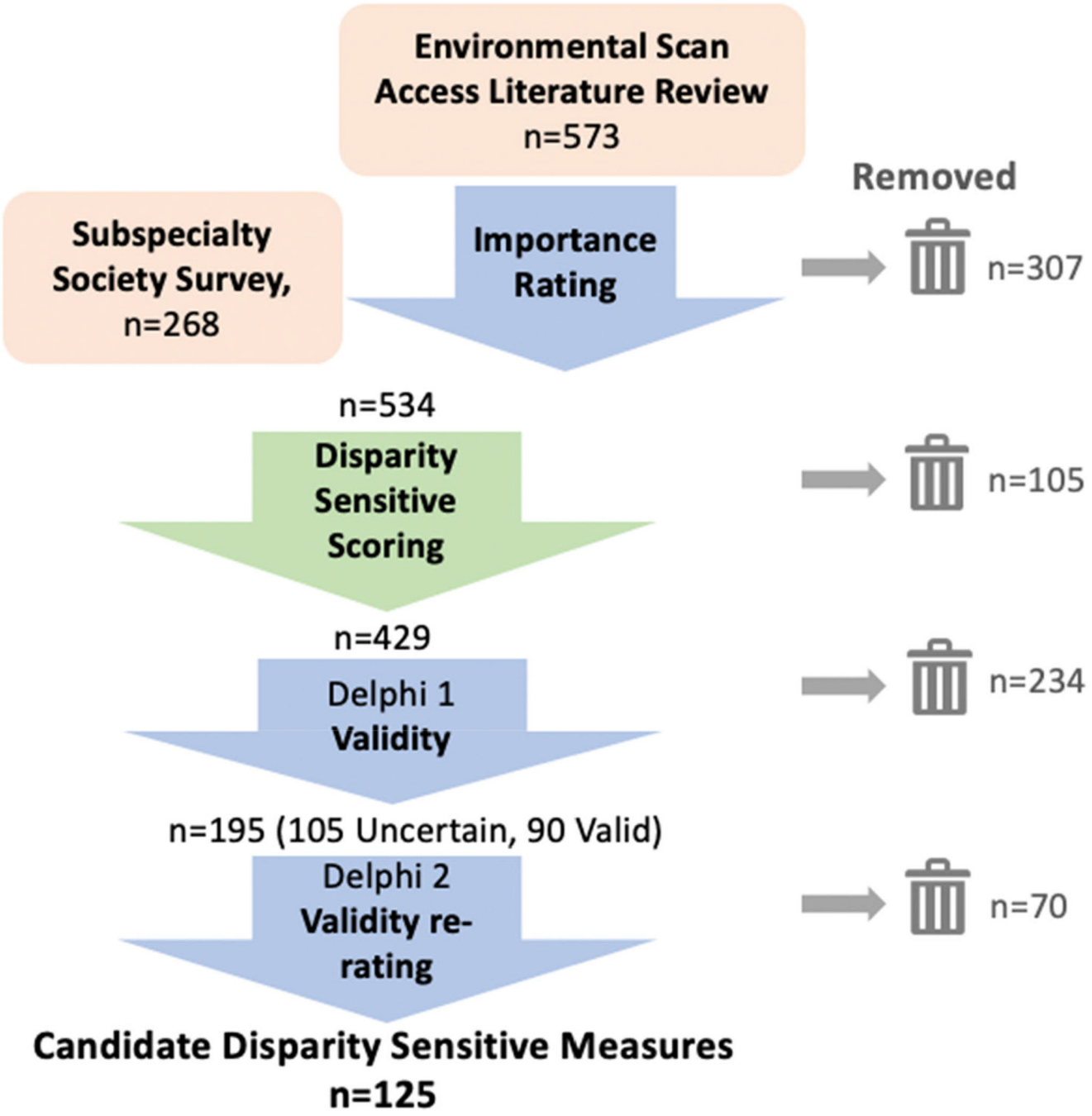


Figure 3.
Roadmap of Metrics for Equitable Access and care in SURgery project.

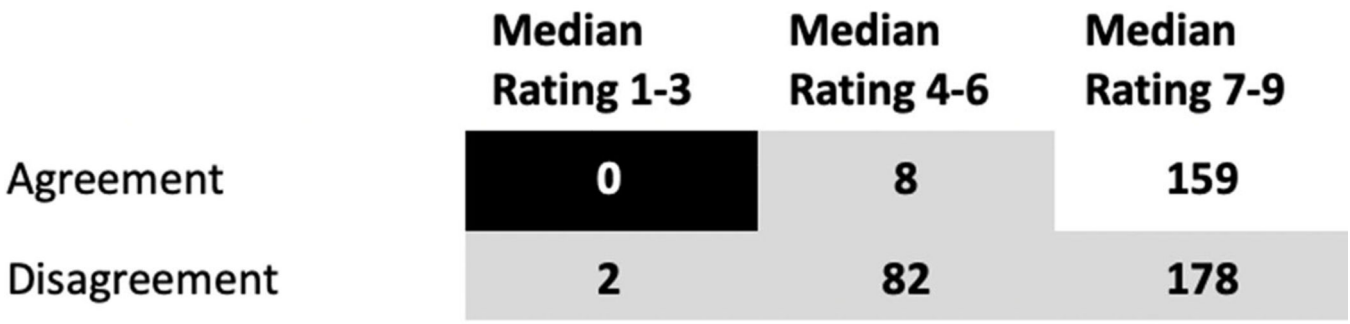


Figure 4.
Summary table of the median ratings and rating dispersion (agreement/disagreement) of the first Delphi validity rating round.

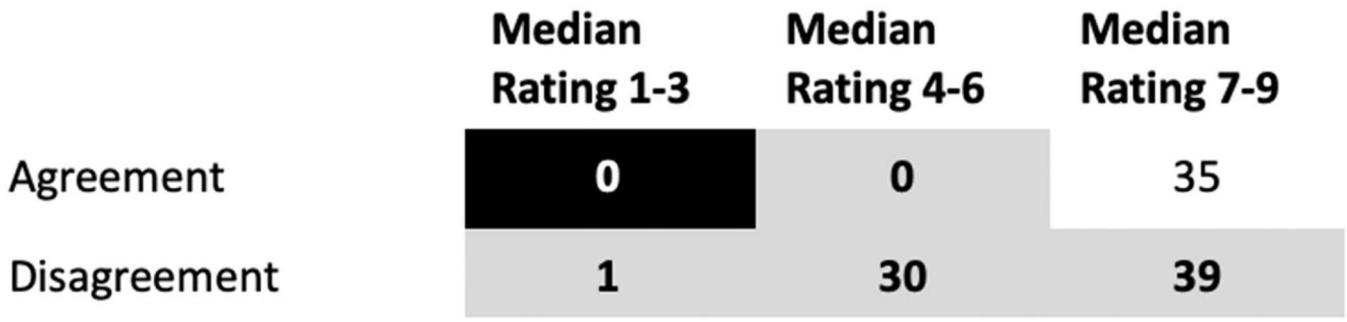


Figure 5.
Summary table of the median ratings and rating dispersion (agreement/disagreement) of the second Delphi validity rating round.