Randomized Pilot Test of a Decision Support Tool for Acute Appendicitis

Decisional Conflict and Acceptability in a Healthy Population

Joshua E. Rosen, MD, MHS,*†‡ David R. Flum, MD, MPH,*†‡ Giana H. Davidson, MD, MPH,*† and Joshua M. Liao, MD, MSc‡§

Objective: To test the effect of a new decision support tool for acute appendicitis and assess its efficacy and acceptability.

Background: Mounting evidence from randomized controlled trials have shown that antibiotics can be a safe and effective treatment for appendicitis. Patients and surgeons must work together to choose the optimal treatment approach for each patient based on their own preferences and values. We developed a decision support tool to facilitate shared decision-making for appendicitis and its effect on decisional outcomes remains unknown.

Methods: We conducted an online randomized field test in at-risk individuals comparing the decision support tool to a standard infographic. Individuals were randomized 3:1 to view the decision support tool or infographic. The primary outcome was the total decisional conflict scale (DCS) score measured before and after exposure to the decision support tool. Secondary outcomes included between-group DCS scores, and between-group comparisons of the acceptability.

Results: One hundred eighty individuals were included in the study. Total DCS scores decreased significantly after viewing the decision support tool (59 [95% confidence interval (CI): 55–63] to 15 [95% CI: 12–17], $P < 0.001$) representing movement from a state of high to low decisional conflict. Individuals exposed to the decision support tool reported higher acceptability ratings (3.7 [95% CI: 3.6–3.8] vs 3.3 [95% CI: 3.2–3.5] out of 4) and demonstrated increased willingness to consider both treatment options.

Conclusions: These data support the further use and testing of this novel decision support tool in patients with acute appendicitis.

Keywords: appendicitis, decision aid, decisional conflict, decision support

INTRODUCTION

Appendicitis has traditionally been treated with appendectomy, but randomized controlled trials have demonstrated the safety and efficacy of antibiotics.[1–5] In particular, the Comparison of Outcomes of Antibiotic Drugs and Appendectomy (CODA) trial found that surgery and antibiotics had similar impacts on patient-reported quality of life and time to symptom resolution.[1] However, the 2 treatments can differ with respect to other outcomes that patients and clinicians may prioritize differently.[6] Given that multiple competing outcomes must be considered—and the reality that appendicitis treatment decisions occur in acute settings between patients and clinicians without long-standing relationships—materials and interventions that support informed shared decision-making are needed.[7]

We used the results from CODA to develop a novel appendicitis decision support tool (DST, Fig. 1, www.appyornot.org), which has been described previously.[8] DSTs or decision aids are a class of tools that can support decision-making between patients and healthcare providers by providing evidence-based information, exploring patient preferences and values, and facilitating value-concordant decisions.[9] They may be particularly helpful for decisions where no clear "best choice" exists, such as for the many patients with acute appendicitis. Our appendicitis DST consists of an informational video and interactive decision aid to help patients understand differences in outcomes between surgery and antibiotics and select the treatment that best aligns with their preferences and values. In this article, we present the results of a pilot study that assesses (1) the impact of the DST on decisional conflict, a measure of uncertainty and (2) whether the DST is acceptable to patients and promotes informed consideration of multiple treatment options. The goal of this study was to ensure that the DST decreases decisional conflict after viewing it, and to assess participants opinions of its use in decision-making to justify further clinical testing and deployment.
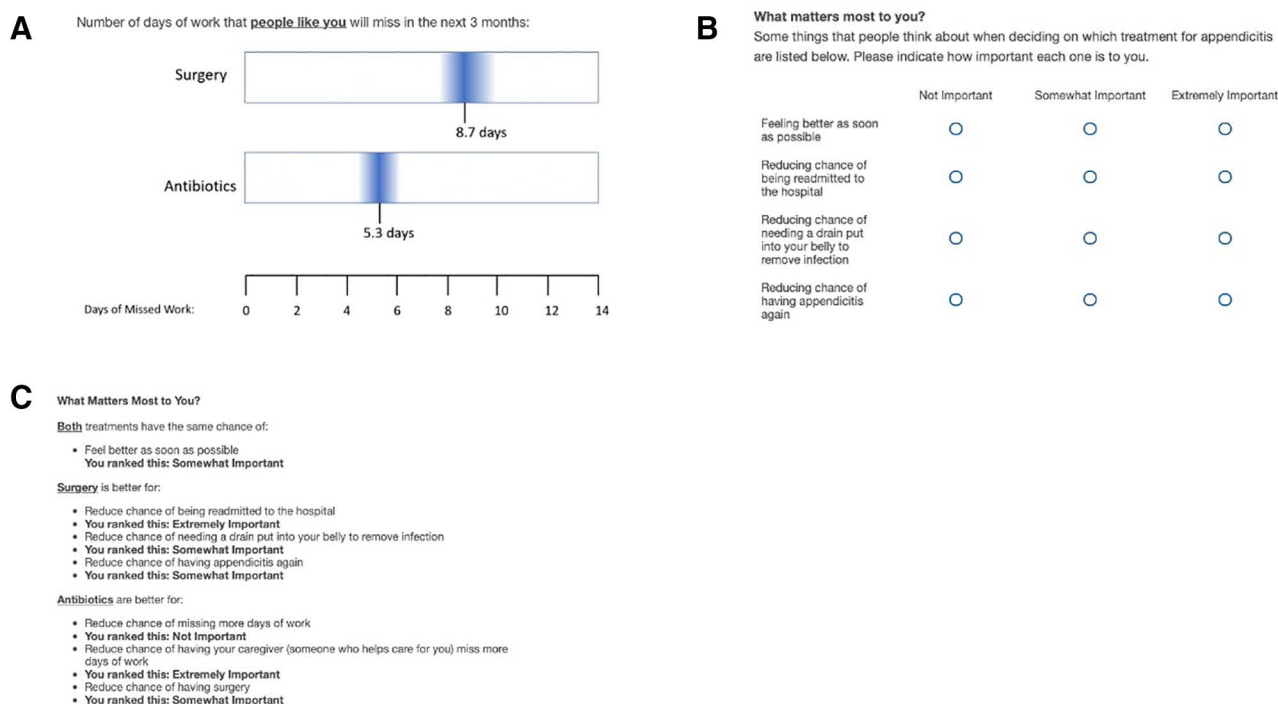
FIGURE 1. Screenshots from DST (A) graphical comparison of an outcome using point estimates and confidence intervals. B, Preference elicitation screen where users rank which outcomes are most important to them. C, Screen showing which outcomes are most likely to be favored by surgical or antibiotic treatment.

METHODS

Study Population

We used Amazon's Mechanical Turk (MTurk) to test the DST in US adults without a history of appendicitis. MTurk workers with an approval rating of at least 95% and at least 50 prior completed tasks were invited to participate until the maximum specified number of participants was reached. Participants are an at-risk population since appendicitis is a common acute surgical condition in US adults in the typical age range of MTurk users.[6,10] The University of Washington institutional review board exempted this study. Individuals were paid $5 for participation.

Survey Quality Control

We implemented a Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) test, an attention-check question at the end of the survey, and manual review of survey responses for duplicate MTurk IDs. To ensure that participants viewed the entire DST, we required those in the DST arm to enter a passcode displayed at the end of the DST.

Description of Survey

Participants were randomized in a 3:1 ratio to the view the DST (www.appyornot.org) or a standardized infographic (http://bec-ertain.org/coda-infographic). The infographic provided basic information about appendicitis and treatment options without the detailed discussions, comparisons, and value-clarification activities present in the DST. The infographic is an appropriate comparator in this online setting since it provides information but does not incorporate many of the hallmarks of a true decision aid such as values clarification exercises, clarification of the decision to be made, or multiple formats (eg, graphical, numerical) for representing the outcomes. Participants also answered demographic and health history questions as well as a 3-item test of objective numeracy.[11]

Outcomes

The primary outcome was the change in the decisional conflict scale (DCS) score in the DST arm before and after viewing the DST. Decisional conflict is a "state of uncertainty about a course of action"[12] and the DCS is a validated instrument used in multiple studies to measure it.[9,13] The DCS has been used as the primary outcome in many studies evaluating decision support interventions,[9] has well-characterized psychometric properties,[12] and is highly relevant to the proximate experience of conflict and confusion while making a decision in the emergency department—aspects that we hope the DST will address. The DCS has subscales, including feeling *informed*, *values clarity*, feeling *uncertain*, feeling *supported*, and making an *effective decision*, scored out of 100. Lower scores represent less decisional conflict with scores >37.5 considered to reflect very high decisional conflict.[12,14] We used a modified version of the DCS excluding the *Support* subscale and a question about "sticking with my decision" from the *Effective Decision* subscale, as these were not relevant to our scenario. Scores were re-scaled per instructions in the DCS User Manual.[12]

Secondary/exploratory analyses included between-arm comparisons of the DCS, decision aid acceptability (4-point scale), and perceptions of trust and accuracy in the information (5-point scale). Acceptability of a DST refers to "ratings regarding the comprehensibility of components of a decision aid, its length, … balance in presentation of information about options, and overall suitability for decision making."[15]

Parameters are reported with 95% confidence intervals (CIs). We calculated *P* values using t-tests for the main outcomes specified above. An alpha of 0.05 was used for statistical significance. This study followed the American Association for Public Opinion Research reporting guidelines for survey studies (see

Table S1, for checklist, http://links.lww.com/AOSO/A175). This study was retrospectively registered on ClinicalTrials.gov (NCT05219786).

RESULTS

Two hundred ten individuals opened the survey. Ten were screened out after indicating that they were an employee of the University of Washington and 6 were removed due to duplicate responses leaving 194 participants who passed quality control and completed the survey. Fourteen reported or were unsure about a prior diagnosis of appendicitis and were excluded from the analysis (one in the infographic arm, 13 in the DST arm). Among 180 participants in the final analysis, demographics were well-balanced between study arms (Table 1). At the start of the study, most individuals in both arms knew that appendicitis was commonly treated with surgery (88% [95% CI = 74%–95%] infographic, 90% [95% CI = 83%–94%] DST), while a few were aware that antibiotics were a treatment option (12% [95% CI = 5%–26%] infographic, 6% [95% CI = 3%–12%] DST).

Primary Analysis: Pre- Versus Postcomparison of Decisional Conflict

The mean total DCS score decreased from 59 (95% CI=55-63) to 15 (95% CI=12-17) (p<0.001) after viewing the DST, moving from a very high to a low state of decisional conflict.[12] DCS scores decreased across all subscales (Table 2). Participant's knowledge of appendicitis improved after viewing the DST (3.4 [95% CI = 3.3–3.5] vs 2.0 [95% CI = 1.8–2.2] out of 4).

Exploratory Analysis: Between-Arm Comparisons

Baseline DCS were similar between DST (59, SD = 22.6) and infographic (53, SD = 26.4) arms. Compared to the infographic, the DST had higher acceptability ratings (3.7 [95% CI = 3.6–3.8] vs 3.3 [95% CI = 3.2–3.5] out of 4), and was associated with greater perceived trust in (4.5 [95% CI = 4.4–4.7] vs 4.3 [95% CI = 4.1–4.5] out of 5, $P = 0.02$) and accuracy of (4.7 [95% CI=4.6–4.8] vs 4.4 [95% CI=4.2–4.6] out of 5, $P = 0.005$) information. Knowledge scores were similar between groups (Table 2). The mean total DCS score was 15 (95% CI = 12–17) for the DST and 18 (95% CI = 13–23) for the infographic ($P = 0.14$). The effect size (defined as the mean difference divided by the pooled standard deviation) for the total DCS score was 0.25 (95% CI = −0.08 to 0.58).

Participants were asked their opinions about treating appendicitis with antibiotics and surgery (Table 3). More participants who viewed the DST thought that it is a good idea to treat appendicitis with antibiotics compared to those viewing the infographic (71% [95% CI = 63%–79%] vs 48% [95% CI = 34%–63%]). When grouping those who agreed or completely agreed, 98% [95% CI = 93%–99%] of those viewing the DST felt it was safe to treat appendicitis with antibiotics compared to 79% [95% CI = 65%–89%] viewing the infographic. Additionally, 77% [95% CI = 68%–83%] of those viewing the DST felt antibiotics would work for them if they had appendicitis compared to 65% [95% CI = 49%–77%] of those viewing the infographic. When asked which treatment they would choose if they developed appendicitis 51% [95% CI = 42%–60%] of those viewing the DST chose antibiotics compared to 38% [95% CI = 24%–54%] of those viewing the infographic.

DISCUSSION

In this pilot study, a novel DST for acute appendicitis decreased decisional conflict among those viewing it by a clinically

TABLE 1.

Participant Characteristics.

Variable	Level	Infographic n = 48	DST n = 132
Age, mean (SD)		38.5 (10.4)	40.4 (11.8)
Sex			
	Female	17 (35.4)	54 (40.9)
	Male	30 (62.5)	78 (59.1)
	Prefer not to say	1 (2.1)	0 (0.0)
Gender identity			
	Female	17 (35.4)	54 (40.9)
	Male	30 (62.5)	78 (59.1)
	Other	1 (2.1)	0 (0.0)
Racial identity			
	Black/African American	3 (6.2)	10 (7.6)
	East Asian	3 (6.2)	8 (6.1)
	Multiple identities	2 (4.2)	5 (3.8)
	Other	0 (0.0)	3 (2.3)
	South Asian	1 (2.1)	2 (1.5)
	Unknown	1 (2.1)	0 (0.0)
	White	38 (79.2)	104 (78.8)
Ethnicity			
	Hispanic/Latino/Latinx	3 (6.2)	5 (3.8)
	Non-Hispanic/Latino/Latinx	44 (91.7)	126 (95.5)
	Prefer not to say	1 (2.1)	1 (0.8)
Education level			
	2-year college degree	4 (8.3)	13 (9.8)
	4-year college degree	21 (43.8)	57 (43.2)
	Graduate degree	7 (14.6)	9 (6.8)
	High school/ GED	6 (12.5)	23 (17.4)
	Some college	9 (18.8)	28 (21.2)
	Some high school	0 (0.0)	2 (1.5)
	Unknown	1 (2.1)	0 (0.0)
Insurance			
	Employer-provided insurance	25 (52.1)	56 (42.4)
	Government	2 (4.2)	3 (2.3)
	MediCAID	5 (10.4)	17 (12.9)
	MediCARE	2 (4.2)	10 (7.6)
	Not insured	8 (16.7)	18 (13.6)
	Other	2 (4.2)	4 (3.0)
	Private insurance	4 (8.3)	24 (18.2)
Employment status			
	Employed full-time	32 (66.7)	84 (63.6)
	Employed part-time	6 (12.5)	17 (12.9)
	Prefer not to say	1 (2.1)	0 (0.0)
	Retired	0 (0.0)	5 (3.8)
	Self-employed	6 (12.5)	21 (15.9)
	Student	2 (4.2)	0 (0.0)
	Unemployed (looking for work)	1 (2.1)	4 (3.0)
	Unemployed (not looking for work)	0 (0.0)	1 (0.8)
Annual household income			
	<$25,000	8 (16.7)	20 (15.2)
	>$100,000	9 (18.8)	12 (9.1)
	$25,000–$50,000	12 (25.0)	43 (32.6)
	$50,001–$75,000	13 (27.1)	33 (25.0)
	$75,001–$100,000	5 (10.4)	23 (17.4)
	Prefer not to say	1 (2.1)	1 (0.8)
Objective numeracy, mean (SD)	Max score = 3	2.4 (0.9)	2.3 (0.8)

TABLE 2.

Pre- Versus Postexposure Comparisons in the Group Who Viewed the DST

	Preexposure to DST (Mean (95% CI))	Postexposure to DST (Mean (95% CI))	P
Knowledge	2.0 (1.8, 2.2)	3.4 (3.3, 3.5)	
DCS total	59 (55, 63)	15 (12, 17)	<0.001
DCS informed subscale	52 (49, 55)	11 (9, 13)	
DCS values clarity subscale	60 (55, 65)	13 (11, 16)	
DCS uncertainty subscale	67 (62, 72)	20 (16, 23)	
DCS effective decision subscale	57 (52, 61)	14 (11, 16)	

Between-Arm Comparisons of Postexposure Outcomes

Outcome	Levels	Infographic n = 48	DST n = 132
Proportion who would choose antibiotics (n (% [95% CI]))		18 (38% [24%, 53%])	67 (51% [42%, 60%])
Do you think the presentation of data was slanted?			
	Antibiotics	2 (4% [0.7%, 15%])	13 (10% [6%, 17%])
	Balanced (no slant)	43 (90% [77%, 96%])	115 (87% [80%, 92%])
	Surgery (appendectomy)	3 (6% [2%, 18%])	4 (3% [1%, 8%])
Knowledge, mean [95% CI]		3.3 [3.1, 3.5]	3.4 [3.3, 3.5]
DCS total, mean [95% CI]		18.1 [13, 23]	14.6 [12, 17]
DCS informed subscale, mean [95% CI]		12.5 [9, 16]	11.4 [9, 13]
DCS values clarity subscale, mean [95% CI]		16.3 [11, 21]	13.3 [11, 16]
DCS uncertainty subscale, mean [95% CI]		23.3 [17, 29]	19.8 [16, 23]
DCS effective decision subscale, mean [95% CI]		20.1 [14, 26]	13.8 [11, 16]
Do you think it is a good idea to treat appendicitis with antibiotics?			
	No	9 (19% [9%, 33%])	17 (13% [8%, 20%])
	Not sure	16 (33% [21%, 49%])	21 (16% [10%, 25%])
	Yes	23 (48% [34%, 63%])	94 (71% [63%, 79%])
It is safe to treat appendicitis with antibiotics			
	Completely disagree	1 (2% [0.11%, 12%])	0 (0% [0.00%, 3.5%])
	Moderately disagree	3 (6.2% [1.6%, 18%])	2 (1.5% [0.26%, 5.9%])
	Neither agree nor disagree	6 (12% [5.2%, 26%])	1 (0.8% [0.04%, 4.8%])
	Moderately agree	26 (54% [39%, 68%])	69 (52% [43%, 61%])
	Completely agree	12 (25% [14%, 40%])	60 (45% [37%, 54%])
It is safe to treat appendicitis with surgery			
	Completely disagree	1 (2.1% [0.11%, 12%])	0 (0% [0.00%, 3.5%])
	Moderately disagree	0 (0% [0.00%, 9.2%])	1 (0.8% [0.04%, 4.8%])
	Neither agree nor disagree	3 (6.2% [1.6%, 18%])	3 (2.3% [0.59%, 7.0%])
	Moderately agree	16 (33% [21%, 49%])	60 (45% [37%, 54%])
	Completely agree	28 (58% [43%, 72%])	68 (52% [43%, 60%])
If I had appendicitis, antibiotics would work to treat it			
	Completely disagree	1 (2.1% [0.11%, 12%])	1 (0.8% [0.04%, 4.8%])
	Moderately disagree	7 (15% [6.5%, 28%])	4 (3.0% [1.0%, 8.1%])
	Neither agree nor disagree	9 (19% [9.4%, 33%])	26 (20% [13%, 28%])
	Moderately agree	26 (54% [39%, 68%])	74 (56% [47%, 65%])
	Completely agree	5 (10% [3.9%, 23%])	27 (20% [14%, 29%])
If I had appendicitis, I would be willing to try antibiotics			
	Completely disagree	3 (6.2% [1.6%, 18%])	4 (3.0% [1.0%, 8.1%])
	Moderately disagree	9 (19% [9.4%, 33%])	16 (12% [7.3%, 19%])
	Neither agree nor disagree	1 (2.1% [0.11%, 12%])	11 (8.3% [4.4%, 15%])
	Moderately agree	25 (52% [37%, 66%])	41 (31% [23%, 40%])
	Completely agree	10 (21% [11%, 35%])	60 (45% [37%, 54%])

Values are reported as (%, [95% CI]) unless otherwise specified.

meaningful amount. Compared to a publicly available infographic summarizing recent study results, the DST achieved higher acceptability, trust, and accuracy ratings. Pilot studies such as this are important to ensure that a new DST has the intended effect on decisional outcomes, is acceptable to the intended users, and adequately informs users about novel treatments.

The DST moved DCS scores by a large magnitude from high to low decisional conflict across all subscales, indicating that the DST addressed multiple domains of decisional conflict. It is notable that this occurred in a testing environment where most participants were unaware of one of the treatment options before using the DST, and where no other information sources (eg, conversations with a surgeon) were available. When compared with the infographic, the effect size of 0.25 [95% CI –0.08 to 0.58] is like those seen in other studies of DSTs.[9,12] A formal trial to test for a difference of this magnitude would require approximately 500 participants. Irrespective of the effect size, the practical implications of a difference between already low DCS scores may not be clinically meaningful. We interpret these results to indicate that a more complex tool (DST) achieved very low DCS scores while being rated as more acceptable, useable, and trustworthy by participants. This is a promising signal that patients may find the DST usable in clinical environments and supports future clinical testing.

The DST informed patients about antibiotics and was associated with increased perceptions of the safety and efficacy of antibiotics without being perceived as biased. Since few patients knew that antibiotics are a treatment for appendicitis at the start of the survey, increased willingness to consider them indicates that the DST encouraged informed consideration of both treatment options. The DST achieved this without concomitant perceptions that the tool biased decisions toward either surgery or antibiotics. It is possible that increased trust and accuracy ratings among participants viewing the DST contributed to their willingness to consider an alternative treatment option.

These results are in line with evidence for a wide array of other DSTs which have been broadly found to increase knowledge and decrease decisional conflict in participants using them.[9,16,17] Notably, most existing decision aids or DSTs have focused on decisions made in elective settings such as perinatal testing, cancer screening, or breast reconstruction options after mastectomy and not in acute surgical conditions.[9,17] One potential limitation of applying these technologies in acute surgical conditions is that while DSTs can be somewhat tailored to pre-specified patient characteristics (eg, appendicolith vs no appendicolith in the present case) they cannot capture all of the nuance and complexity of each unique patients (eg, comorbidities, prior surgeries, disease characteristics, etc) that may affect surgical risk and treatment decision-making. Thus, it is crucial that they be used as a complement to thorough discussions with a surgeon, and a single tool will not be appropriate for use in

all patients. Additionally, the acute care setting may introduce logistical barriers to the implementation of a DST—such as time pressures, resource and workflow constraints, and patient factors such as acute pain that may interfere with the desire or ability to engage in shared decision-making with or without a DST.[7] While the present study provides pilot data for applying this intervention to an acute surgical condition, further testing will be required in patients with acute appendicitis to assess the efficacy and feasibility of delivering this intervention to individuals with an acute surgical condition in the emergency department.

Study limitations include insufficient power to detect between-arm differences and the use of an at-risk versus actively ill population. MTurk users differ from the general population in terms of demographic and health status, which may affect the generalizability of results.[18] However, we believe that this is an appropriate venue for initial pilot testing of a DST, particularly since the tool must act on its own (ie, without additional healthcare provider interaction) demonstrating its independent marginal benefit. Furthermore, given the randomized design for assessing between-arm differences in acceptability and perceptions of the DST, we believe the marginal differences between groups still provide valid and inference for justifying the further development and use of this tool in clinical populations. Finally, we used a publicly available infographic as a comparator, which eliminates the variability inherent in real clinical discussions. Based on our typical experiences, the infographic contains more information than is typically provided by clinicians, which if anything may have biased between-arm comparisons toward the null.

In conclusion, in a randomized pilot test, a novel DST for acute appendicitis decreased decisional conflict, was highly acceptable to users, and encouraged consideration of antibiotics as a treatment approach. These results are promising and support the implementation and testing of this DST and its effect on clinical decision-making among patients with acute appendicitis.

ACKNOWLEDGMENTS

REFERENCES

1. Flum DR, Davidson GH, Monsell SE, et al; CODA Collaborative. A randomized trial comparing antibiotics with appendectomy for appendicitis. N Engl J Med. 2020;383:1907–1919.
2. O'Leary DP, Walsh SM, Bolger J, et al. A randomized clinical trial evaluating the efficacy and quality of life of antibiotic-only treatment of acute uncomplicated appendicitis: results of the COMMA trial. Ann Surg. 2021;274:240–247.
3. Salminen P, Tuominen R, Paajanen H, et al. Five-year follow-up of antibiotic therapy for uncomplicated acute appendicitis in the APPAC randomized clinical trial. JAMA. 2018;320:1259–1265.
4. Sallinen V, Akl EA, You JJ, et al. Meta-analysis of antibiotics versus appendicectomy for non-perforated acute appendicitis. Brit J Surg. 2016;103:656–667.
5. Park HC, Kim MJ, Lee BH. Randomized clinical trial of antibiotic therapy for uncomplicated appendicitis. Br J Surg. 2017;104:1785–1790.
6. Rosen JE, Agrawal N, Flum DR, et al. Willingness to undergo antibiotic treatment of acute appendicitis based on risk of treatment failure. Br J Surg. 2021;108:e361–e363.
7. Rosen JE, Flum DR, Liao JM. The need for patient decision aids in acute care settings. Healthc. 2022;100639.
8. Rosen JE, Liao JM, Flum DR, et al. Development of a decision support tool for acute appendicitis. medRxiv. 2021. doi: 10.1101/2021.11.08.21266077
9. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2017;4:CD001431.
10. Rosen JE, Agrawal N, Flum DR, et al. Verbal descriptions of the probability of treatment complications lead to high variability in risk perceptions: a survey study. Ann Surg. 2021. doi: 10.1097/sla.0000000000005255
11. Schwartz LM, Woloshin S, Black WC, et al. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127:966–972.
12. O'Connor A. User Manual - Decisional Conflict Scale (16 item statement format). 2010. Available at: http://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf. Accessed October 19, 2021.
13. O'Connor AM. Validation of a decisional conflict scale. Med Decis Making. 1995;15:25–30.
14. Kryworuchko J, Stacey D, Bennett C, et al. Appraisal of primary outcome measures used in trials of patient decision support. Patient Educ Couns. 2008;73:497–503.
15. O'Connor A, Cranney A. User Manual - Acceptability. 2002. Available at: http://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Acceptability.pdf. Accessed February 24, 2021.
16. Knops AM, Legemate DA, Goossens A, et al. Decision aids for patients facing a surgical treatment decision: a systematic review and meta-analysis. Ann Surg. 2013;257:860–866.
17. Leinweber KA, Columbo JA, Kang R, et al. A review of decision aids for patients considering more than one type of invasive treatment. J Surg Res. 2019;235:350–366.
18. Walters K, Christakis DA, Wright DR. Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? PLoS One. 2018;13:e0198835.