



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Original Article

# Reliability and Validity of 2 Surgical Prioritization Systems for Reinstating Nonemergent Benign Gynecologic Surgery during the COVID-19 Pandemic

Cherie Q. Marfori, MD, Jordan S. Klebanoff, MD, Catherine Z. Wu, MD, Whitney A. Barnes, MD, MPH, Charelle M. Carter-Brooks, MD, MSc, and Richard L. Amdur, PhD

From the Departments of Obstetrics and Gynecology (Drs. Marfori, Klebanoff, Wu, Barnes, Carter-Brooks), and Surgery (Dr. Amdur), The George Washington University Hospital, Washington, District of Columbia

**ABSTRACT** **Study Objective:** Scientifically evaluate the validity and reproducibility of 2 novel surgical triaging systems, as well as offer modifications to the Medically-Necessary, Time-Sensitive (MeNTS) criteria for improved application in gynecologic surgeries.

**Design:** Retrospective cohort study.

**Setting:** Academic university hospital.

**Patients:** Ninety-seven patients with delayed benign gynecologic procedures owing to the coronavirus disease 2019 pandemic.

**Intervention(s):** Surgical prioritization was assessed using 2 novel scoring systems, the Gynecologic Medically-Necessary Time-Sensitive (Gyn-MeNTS) and modified Elective Surgery Acuity Scale (mESAS) systems for all 93 patients included.

**Measurements and Main Results:** The interrater reliability and validity of 2 novel surgical prioritization systems (Gyn-MeNTS and mESAS) were assessed. The Gyn-MeNTS scores were calculated by 3 raters and analyzed as continuous variables, with a lower score indicating more urgency/priority. The mESAS score was calculated by 2 raters and analyzed as a 3-level ordinal variable with a higher score indicating more urgency/priority. All 5 raters were blinded to reduce bias. The Gyn-MeNTS interrater reliability was tested using Spearman  $r$  and paired  $t$  tests were used to detect systematic differences between raters. Weighted  $\kappa$  indicated mESAS reliability. Concurrent validity with mESAS and surgeon self-prioritization (SSP) was examined with Spearman  $r$  and logistic regression. Spearman  $r$ 's for all Gyn-MeNTS rater pairs were above 0.80 (0.84 for 1 vs 2; 0.82 for 1 vs 3; and 0.82 for 2 vs 3, all  $p < .001$ ) indicating strong agreement. The weighted  $\kappa$  for the 2 mESAS raters was 0.57 (95% confidence interval, 0.40–0.73) indicating moderate agreement. When used together, both scores were significantly independently associated with SSP, with strong discrimination (area under the curve, 0.89).

**Conclusion:** Interrater reliability is acceptable for both scoring systems, and concurrent validity of each is moderate for predicting SSP, but discrimination improves to a high level when they are used together. Journal of Minimally Invasive Gynecology (2020) 00, 1–12. © 2020 AAGL. All rights reserved.

**Keywords:** Case management; Coronavirus; Delivery of healthcare; Elective surgical procedures; Triage

The respiratory illness caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), also known as coronavirus disease 2019 (COVID-19), has

upended global healthcare delivery. In March 2020, the American College of Surgeons (ACS), the US Surgeon General, and the Centers for Medicare & Medicaid services

The authors declare that they have no conflict of interest.  
Institutional review board: NCR202525, May 7, 2020.

**Data sharing:** All data collected for the study including the study protocol, a data dictionary defining each field, and deidentified patient data will be made available to qualified researchers using a data-sharing platform on reasonable request with publication.

Corresponding author: Cherie Q. Marfori, MD, Department of Obstetrics and Gynecology, The George Washington University Hospital, 2150

Pennsylvania Ave NW, Suite 6A, Washington, DC 20037.  
E-mail: [cmarfori@mfa.gwu.edu](mailto:cmarfori@mfa.gwu.edu)

Submitted June 20, 2020, Revised July 21, 2020, Accepted for publication July 23, 2020.

Available at [www.sciencedirect.com](http://www.sciencedirect.com) and [www.jmig.org](http://www.jmig.org)

all made recommendations to cancel elective surgery throughout the United States to preserve resources to treat those critically ill and to reduce transmission between doctors and patients [1–3]. However, the term elective does not mean unnecessary, and postponement of these surgeries may lead to significant morbidity or even mortality. Surgeons are faced with the potentially overwhelming and morally exhausting task of prioritizing postponed patients within their departments and hospital systems.

In response, the ACS developed a tiered ranking system for prioritization of elective surgeries based on the risk of delay and patient comorbidities, the Elective Surgery Acuity Scale (ESAS) [4]. Although this system provides a framework, many hospitals and surgeons find implementation difficult given its vagueness. Recognizing these limitations, the Society for Gynecologic Surgeons (SGS) issued a joint statement, which applied the ACS tier to numerous benign gynecologic procedures to serve as a guide of acuity [5]. In addition, a third and novel scoring system by Prachand et al [6] serves to individually rank nonemergent surgeries coined Medically-Necessary, Time-Sensitive (MeNTS) procedures. This scoring system attempts to objectively prioritize surgeries by grading 21 factors within the broad categories of procedure variables, disease state, and comorbidities. All of these scoring systems are based on expert opinion and consensus, and none have been externally validated nor tested for reliability.

At our institution, the Division of Benign Gynecologic Surgery incurred a large number of case cancellations with the orders to halt elective surgery. We recognized a need for a reproducible system beyond a 3-tiered scale to help triage and prioritize affected patients. We took great interest in the MeNTS tool as a potential solution. Although our institution elected not to adapt the MeNTS criteria system-wide, our division sought to critically evaluate it as a solution to ethically and efficiently prioritize patients beyond the ACS criteria. Our objective was to scientifically evaluate the validity and reproducibility of both of these triaging systems, as well as offer modifications to the MeNTS criteria for improved application in gynecologic surgeries.

## Materials and Methods

After obtaining institutional review board exemption (IRB#NCR202525), we performed a single-center retrospective cohort study evaluating the interrater reliability and validity of 2 novel prioritization systems, the SGS adaptation of the ESAS (modified ESAS [mESAS]) and a Gynecologic Medically-Necessary Time-Sensitive tool (Gyn-MeNTS). A total of 97 benign gynecologic procedures were affected between March 16, 2020, and April 30, 2020, in our tertiary academic institution in Washington, DC, including cases from 8 general obstetrician gynecologists, 2 minimally invasive gynecologic surgeons, a urogynecologist, and a gynecologic oncologist. We excluded 3 patients who had not completed their preoperative

evaluations, and thus the severity of their disease was unclear. One additional patient was excluded because she suffered from a subarachnoid hemorrhage and was no longer eligible for her planned procedure.

## SSP

Beginning March 16, as required by our hospital system, all patients were categorized by their respective surgeons into 1 of 3 categories on the basis of their level of morbidity if delayed. Level 1 indicated no morbidity with delay, level 2 indicated some morbidity with delay, and level 3 indicated significant morbidity and/or mortality with delay. Only those deemed level 3 were initially allowed to proceed as scheduled. Levels 1 and 2 would be postponed until further notice, likely to extend 2 to 3 months. For the purpose of the validity analysis, those patients who were allowed to proceed with surgery will be referred to as urgent SSP.

## ESAS and mESAS

The ESAS scale divides elective surgeries into 3 tiers, low acuity (tier 1) defined as “nonlife-threatening illness,” intermediate acuity (tier 2) defined as “nonlife-threatening but with potential for future morbidity and mortality,” and high acuity (tier 3). In addition, each tier is further divided into subtype A (healthy patients) or subtype B (unhealthy patients) to better discriminate patient risk and hospital resource use. The ACS recommended proceeding with tier 3 surgery and postponing tier 1 and 2 surgery (or performing at an ambulatory surgery center) [4]. The joint statement published by SGS on April 28 applied the ACS–tiered system to numerous benign gynecologic procedures (Fig. 1) [5]. This framework categorizes specific surgical procedures taking into consideration their indications and severity of disease symptoms. Two authors (W.A.B. and C.M.C.B.) assigned all patients into 1 of the 3 mESAS tiers by reviewing the patient’s electronic medical record. These authors were queried about their awareness of the alternative MeNTS scoring system and, after finding them unfamiliar, were instructed to intentionally proceed in this manner to reduce potential bias toward another scoring system.

## MeNTS Modifications for Gyn-MeNTS

The original MeNTS scoring criteria attempts to objectively prioritize surgeries by grading 21 factors within the broad categories of procedure variables, disease state, and comorbidities [6]. The cumulative score ranges between 21 and 105 and serves as a rank in priority, with lower numbers equating to greater priority (Table 1). Higher scores equate to poorer perioperative outcomes, higher hospital resource use, an increased risk of COVID-19 transmission, and an increased ability to safely defer surgery.

When attempting to apply the MeNTS model to our gynecologic patients, we believed adaptations could be

Fig 1

SGS mESAS for benign gynecologic indications and surgeries. AMH, antimüllerian hormone; ASA, American Society of Anesthesiologists; ASC = ambulatory surgery center; AUB = abnormal uterine bleeding; CIN = cervical intraepithelial neoplasia; EIN = endometrial intraepithelial neoplasm; EMB = endometrial biopsy; GYN = gynecology; LARC = long-acting reversible contraception; mESAS = modified Elective Surgery Acuity Scale; MUS = midurethral sling; PMB = postmenopausal bleeding; SGS = Society for Gynecologic Surgeons; UTI = urinary tract infection; QoL = quality of life.

Tier	Definition	Gynecology Examples	Urogynecology Examples	Reproductive Examples	Suggested Location
Tier 1a	<b>Low acuity surgery/healthy patient</b> Not life-threatening illness	<ul style="list-style-type: none"> <li>- Benign appearing adnexal mass, asymptomatic</li> <li>- Hysterectomy, without anemia, stable symptoms</li> <li>- Myomectomy, asymptomatic</li> <li>- Hysteroscopic metroplasty</li> <li>- Endometriosis, stable symptoms</li> <li>- Chronic pelvic pain requiring diagnosis (i.e. diagnostic laparoscopy)</li> <li>- AUB, without anemia, low suspicion malignancy</li> <li>- Bleeding requiring diagnosis (i.e. diagnostic/operative hysteroscopy)</li> <li>- Persistent PMB, low suspicion (neg EMB)</li> <li>-</li> </ul>	<ul style="list-style-type: none"> <li>- Stress urinary incontinence surgery</li> <li>- Surgery for symptomatic prolapse in patients without other significant sequelae.</li> <li>- Asymptomatic mesh exposure</li> <li>- Neuromodulation for urinary urgency incontinence, retention or fecal incontinence</li> <li>- Diverticulectomy</li> </ul>	<ul style="list-style-type: none"> <li>- Benign ovarian cyst aspiration</li> </ul>	Outpatient ASC, or Hospital
Tier 1b*	<b>Low acuity surgery/unhealthy patient based on ASA or Charlson comorbidity index</b>				ASC or hospital
Tier 2a	<b>Intermediate acuity surgery/healthy patient</b> Not life threatening but potential for near future morbidity or mortality Significant impairment of QoL	<ul style="list-style-type: none"> <li>- AUB with secondary anemia, stable</li> <li>- AUB age &gt; 50y with inability to sample in office</li> <li>- Endometriosis, poorly controlled pain or fertility desired, low AMH/ovarian reserve</li> <li>- Endometrial polyps age &gt; 50y</li> <li>- Symptomatic ovarian cyst</li> <li>- Myomectomy, symptomatic fibroid</li> </ul>	<ul style="list-style-type: none"> <li>- Fistula repair stable</li> <li>- Mesh-related complication (infection or severe pain)</li> <li>- Surgery for advanced prolapse in patients with significant sequelae (urinary retention, recurrent UTI, pain, etc)</li> </ul>	<ul style="list-style-type: none"> <li>- Missed abortion &lt; 12 weeks</li> </ul>	Outpatient or overnight hospitalization ASC or Hospital
Tier 2b*	<b>Intermediate acuity surgery/unhealthy patient</b>				ASC or hospital
Tier 3a	<b>High acuity surgery/healthy patient</b> High potential for near future morbidity or mortality Severe impairment of QoL	<ul style="list-style-type: none"> <li>- Endometrial hyperplasia/EIN, highly suspicious for atypia/malignancy</li> <li>- Adnexal mass, complex, suspicious for torsion, malignancy</li> <li>- AUB, secondary anemia, unstable, worsening bleeding</li> <li>- Deeply infiltrating endometriosis with severe bowel/bladder symptoms, intractable/severe pain</li> <li>- CIN II, CIN III treatment</li> <li>- Symptomatic ovarian cyst, intractable, severe pain</li> <li>- Large fibroids, with hydronephrosis, debilitating pain, bulk symptoms or urinary retention</li> <li>- Other potential GYN malignancy cases</li> <li>- Sterilization procedures for women who are not LARC candidates</li> </ul>	<ul style="list-style-type: none"> <li>- Mesh in a viscus</li> <li>- Advanced stage prolapse with evidence of upper tract obstruction and unable to retain pessary</li> <li>- Obstructed voiding after MUS</li> <li>- incarcerated prolapse</li> </ul>	<ul style="list-style-type: none"> <li>- Myomectomy (for infertility)</li> <li>- Hysteroscopic metroplasty (for infertility)</li> <li>- Transvaginal oocyte Aspiration in infertile woman;</li> <li>- Embryo transfer</li> <li>- Ovarian cyst that is affecting ability to do egg retrieval</li> <li>- Myomectomy, asymptomatic but affecting uterine cavity in infertile woman or affecting ability to monitor and perform egg retrieval</li> <li>- Hysteroscopic polypectomy/myomectomy infertility</li> </ul>	Outpatient hospital with pathologist available  ASC rare but possible
Tier 3b*	<b>High acuity surgery/unhealthy patient</b>				Hospital in-patient

**Table 1**

## Medically-Necessary Time-Sensitive (MeNTS) OR Procedure prioritization worksheet

	1	2	3	4	5
Procedure					
OR time	<30 min	31–60 min	61–120 min	121–180 min	≥181 min
LOS anticipated	Outpatient	23 h	24–48 h	2–3 d	≥4 d
Post-Op ICU need	Very unlikely	<5%	5–10%	11–25%	>25%
Bleeding risk/EBL	<100 mL	101–250 mL	251–500 mL	501–750 mL	≥751 mL
Surgical team size	1	2	3	4	≥5
Intubation needed to perform procedure (Probability)	≤1%	1–5%	6–10%	11–25%	>25%
Surgical site	None of the following	Abdominopelvic MIS Surgery	Abdominopelvic Open Surgery, Infraumbilical	Abdominopelvic Open Surgery, Supraumbilical	OHNS/Upper GI/Thoracic
Disease					
Nonoperative treatment option effectiveness	None available	Available, <40% effective as surgery	Available, 40–60% effective as surgery	Available, 61–95% effective as surgery	Available, equally effective
Nonoperative treatment option resource use/ exposure risk	Significantly worse/ not applicable	Somewhat worse	Equivalent	Somewhat better	Significantly Better
Impact of 2-week delay in disease outcome	Significantly worse	Worse	Moderately worse	Slightly worse	Minimally worse
Impact of 2-week delay in surgical difficulty/risk	Significantly worse	Worse	Moderately worse	Slightly worse	Minimally worse
Impact of 6-week delay in disease outcome	Significantly worse	Worse	Moderately worse	Slightly worse	Minimally worse
Impact of 6-week delay in surgical difficulty/risk	Significantly worse	Worse	Moderately worse	Slightly worse	Minimally worse
Patient					
Age	<20 yrs	21–40 yrs	41–50 yrs	51–65 yrs	>65 yrs
Lung disease (asthma, COPD, CF)	None	-	-	Minimal (rare inhaler)	> Minimal
OSA	Not present	-	-	Mild/moderate (no CPAP)	On CPAP
CV disease (HTN, CHF, CAD)	None	Minimal (no meds)	Mild (1 med)	Moderate (2 meds)	Severe (≥3 meds)
Diabetes	None	-	Mild (no meds)	Moderate (PO meds only)	>Moderate (insulin)
Immunocompromised	No	-	-	Moderate	Severe
ILI Sx's (fever, cough, sore throat, body aches, diarrhea)	None (Asymptomatic)	-	-	-	Yes
Exposure to COVID in past 14 d	No	Probably Not	Possibly	Probably	Yes

CAD = coronary artery disease; CF = cystic fibrosis; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; COVID = coronavirus disease; CPAP = continuous positive airway pressure; CV = cardiovascular; d = day; EBL = estimated blood loss; GI = gastrointestinal; HTN = hypertension; ICU = intensive care unit; ILI Sx = influenza-like-illness symptoms; LOS = length of stay; MIS = minimally-invasive surgery; OR = operating room; OSA = obstructive sleep apnea; OHNS = otolaryngology head and neck Surgery; PO = by mouth.

Based on Prachand et al [6].

made to improve clarity, objectivity, and validity. Specific adaptations made to the original MeNTS score included (Table 2):

(1) Modifying the “need for intubation” to 3 distinct planned anesthesia modalities (local/regional, monitored anesthesia care/conscious sedation, or general).

This eliminated the surgeon estimating risk of intubation when no calculator exists for this.

(2) Modifying surgical site to reflect gynecologic procedures and potential exposure of operating room staff to smoke plume with planned electrosurgery or laparoscopy. Despite a lack of evidence that infectious viral particles

**Table 2**

Gynecologic Medically-Necessary Time-Sensitive (Gyn-MeNTS) Procedure prioritization worksheet

	1	2	3	4	5
Procedure					
OR Time	<30 min	31–60 min	61–120 min	121–180 min	≥181 min
Estimated LOS	Outpatient	<23 h	24–48 h	2-3 d	≥ 4 d
Post-Op ICU need	Very Unlikely	<5%	5–10%	11–25%	>25%
Anticipated EBL	≤100 mL	101–250 mL	251–500 mL	501–750 mL	≥751 mL
Surgical team size	1	2	3	4	≥5
Anesthesia requirement	Local/Regional	-	MAC/Conscious Sedation	-	General
Surgical site	Vaginal Surgery OR Hysteroscopy IF NO electrosurgery	Hysteroscopy with electrocautery OR Abdominopelvic MIS (LSC) surgery	Vaginal surgery IF electrosurgery	Abdominopelvic open surgery, infraumbilical	Abdominopelvic open surgery, supraumbilical
Disease					
Trial of alternative therapy*	Exhausted reasonable alternatives OR none exist	Treatment requires frequent office visits	2 alternatives tried	1 alternative tried	None
Pain	Severe (7–10), poorly controlled OR requiring ED visit	Moderate to Severe (4–10), and minimally controlled, OR requiring frequent clinic visits	Moderate (4–6), controlled	Mild (1–3), controlled	None
Anemia	Debilitating, Hgb <6.5, OR recent blood transfusion	Severe, Hgb 6.6–8, OR reliance on IV iron	Moderate, Hgb 8.1–10	Mild, Hgb 10.1–12	None, Hgb >12
Impact on desired immediate fertility	Significant/ Probable	-	Mild/ Possible	-	None
Disease impact on GI/GU morbidity <sup>†</sup>	Significant/ Progressive	-	Minimal/ Stable	-	None
Impact of 6-week delay in disease outcome	Significantly worse	Worse	Moderately worse	Slightly worse	Minimally worse
Impact of 6-week delay in surgical difficulty/risk	Significantly worse	Worse	Moderately worse	Slightly worse	Minimally worse
Patient					
Age	≤20 yrs	21–40 yrs	41–50 yrs	51–65 yrs	≥66 yrs
Lung disease (asthma, COPD, CF, smoking)	None	-	-	Minimal OR Active smoking	> Minimal
Obesity/ OSA	Normal BMI (BMI <25)	Overweight (BMI 25–29.9)	Class 1 Obesity (BMI 30–34.9)	Class 2 Obesity (BMI 35–39.9)	Class 3 Obesity (BMI ≥40) or OSA
CV disease (HTN, CHF, CAD)	None	Minimal (no meds)	Mild (1 med)	Moderate (2 meds)	Severe (≥ 3 meds)
Diabetes	None	-	Mild (no meds)	Moderate (PO meds only)	>Moderate (insulin) OR poorly controlled
Immunocompromised	No	-	-	Moderate	Severe
Exposure to COVID in past 14 days	No	-	Possibly	-	Yes

BMI = body mass index; CAD = coronary artery disease; CF = cystic fibrosis; CHF = congestive heart failure; COVID = coronavirus disease; COPD = chronic obstructive pulmonary disease; CV = cardiovascular; ED = emergency department; EBL = estimated blood loss; GI = gastrointestinal; GU = genitourinary; Hgb = hemoglobin; HTN = hypertension; IV = intravenous; ICU = intensive care unit; LOS = length of stay; LSC = laparoscopy; MAC = monitored anesthesia care; MIS = minimally invasive surgery; OR = operating room; OSA = obstructive sleep apnea; PO = by mouth.

\*Trial of alternative therapy examples: expectant management, OCPs, POPs, NSAIDs, TXA, LARCs, GnRH analogues, pelvic floor physical therapy, nerve blocks, pessary, UAE.

<sup>†</sup> Examples include: hydronephrosis, urinary retention, bowel lumen narrowing.

exist outside of the airway, we believed it prudent to incorporate risk of uncontained/unfiltered plume as societies began implementing warnings about the unknown risk of COVID-19 from abdominopelvic surgery,

particularly with laparoscopy and electrosurgery (Society of American Gastrointestinal and Endoscopic Surgeons) [7,8]. At our institution, we have ultralow particulate air filters available for laparoscopy to mitigate this risk.

- (3) Instead of calculating “nonoperative treatment effectiveness percentage” and “exposure risk,” we transformed these 2 variables into 5 distinct categories pertinent to the gynecologic surgical patient including whether and how many alternative therapies have been tried, the presence/severity of pain, the presence/severity of anemia, the impact on desired immediate fertility, and the impact on adjacent genitourinary and gastrointestinal systems.
- (4) Removing the variables related to “two-week delay.” Because we could think of no examples within nonemergent benign gynecologic surgery in which a difference in morbidity or surgical difficulty exists between a 2-week and 6-week delay, we dropped this categorization.
- (5) Substituting the obstructive sleep apnea scale with an obesity scale, which included the binary question of the presence or absence of sleep apnea. Given the importance of obesity severity on postoperative outcomes, acknowledging this formally seemed prudent. In addition, given a large number of patients who are morbidly obese have undiagnosed sleep apnea and many diagnosed patients are noncompliant with therapy, we believed making obstructive sleep apnea a binary yes/no question was valid.
- (6) Removing the variable “influenza-like illness symptoms.” Any patient demonstrating these symptoms should continue to be delayed until resolution.
- (7) Limiting the options for “exposure to known COVID positive patient in the last 14 days” to improve reproducibility. At our institution, we implemented universal COVID-19 testing for all patients undergoing emergent and scheduled surgery.

Three authors (C.Q.M., J.S.K., C.Z.W.) adapted and applied the modified Gyn-MeNTS scoring system to all patients. All 5 authors were blinded to each other’s scores to reduce bias. The authors could not be blinded from the urgent surgeon self-prioritization (SSP) because these surgeries were performed as scheduled and could be elicited from chart review.

### Statistical Analysis

#### Variables

Gyn-MeNTS scores were calculated by 3 reviewers and analyzed as continuous variables with possible scores ranging from 21 to 105. The lower the score, the more prioritized the surgery would be. The mESAS score was calculated by 2 different reviewers and analyzed as a 3-level ordinal variable (1/2/3) with a higher score indicating more urgency/prioritization. Finally, the SSP score was made into a binary variable with those considered highest

priority (urgent SSP) separated from those considered lower priority.

#### Reliability

The 3 Gyn-MeNTS raters’ scores were examined using Spearman  $r$  to determine the level of monotonic association and paired  $t$  tests to determine whether there were systematic differences between raters. A relevant systematic difference was indicated by a mean difference  $>0.1$  along with a significant difference on the paired  $t$  test. If all Spearman  $r$ ’s were  $>0.80$ , it indicated a strong interrater reliability. To determine which Gyn-MeNTS items had the worst reliability, agreement between raters’ Gyn-MeNTS item scores were examined using percent exact agreement, rather than  $\kappa$ , because the raters used different numbers of categories on several items.

#### Concurrent Validity

We took the mean of the Gyn-MeNTS scores and examined the Spearman  $r$  of this score with the mean of the mESAS scores and with urgent SSP as measured by actual scheduling (a binary variable, yes/no). Concurrent validity for the Gyn-MeNTS score was indicated by strong Spearman  $r$  with both the mESAS score and urgent SSP. We also examined whether urgent SSP could be predicted independently using both the Gyn-MeNTS score and the mESAS rating in a multivariable logistic regression model. If both were significant independent predictors, we then used the log-linear equation produced by the regression model to calculate each patient’s probability of being classified as urgent SSP and examined the association of probability quartile with urgent SSP status using chi-square. We examined the distribution of mESAS rating with the Gyn-MeNTS scores achieving the highest 67%, 75%, 80%, 85%, 90%, and 95% of urgency levels using chi-square.

SAS (version 9.4, SAS Institute, Cary, NC) was used for data analysis, with  $p < .05$  considered significant.

### Results

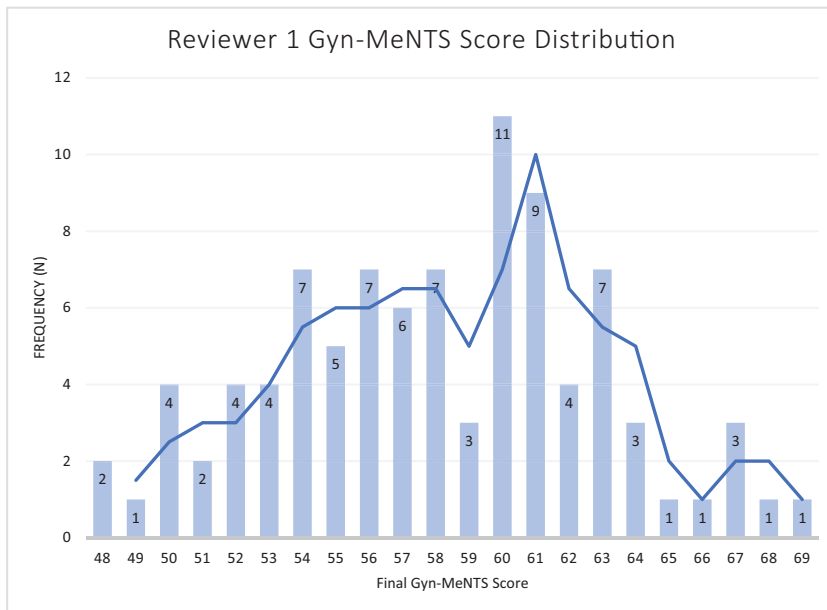
There were 93 patients in the sample. The mean distribution of patients with the mESAS ratings were 8 (9%) level 3, 33 (35%) level 2, and 52 (56%) level 1. Of the 93 patients, 12 had been deemed highest priority by the SSP scheme, and their surgeries were performed without delay. An additional 12 patients were chosen as the next level of prioritization to be scheduled as soon as allowable. These 24 patients make up the “urgent SSP”. Individual Gyn-MeNTS scores ranged from 48 to 70 points. Average scores that fell into the most urgent quartile ranged from 49 to 55. The distributions of Gyn-MeNTS score assignments for the 3 reviewers can be seen in [Figs. 2 to 4](#).

#### Interrater Reliability

The mean  $\pm$  standard deviation Gyn-MeNTS scores for raters 1, 2, and 3 were  $58.0 \pm 4.8$ ,  $59.2 \pm 4.3$ , and  $58.0 \pm$

**Fig. 2**

Reviewer 1 distribution of Gyn-MeNTS score. Gyn-MeNTS = Gynecologic Medically-Necessary Time-Sensitive.

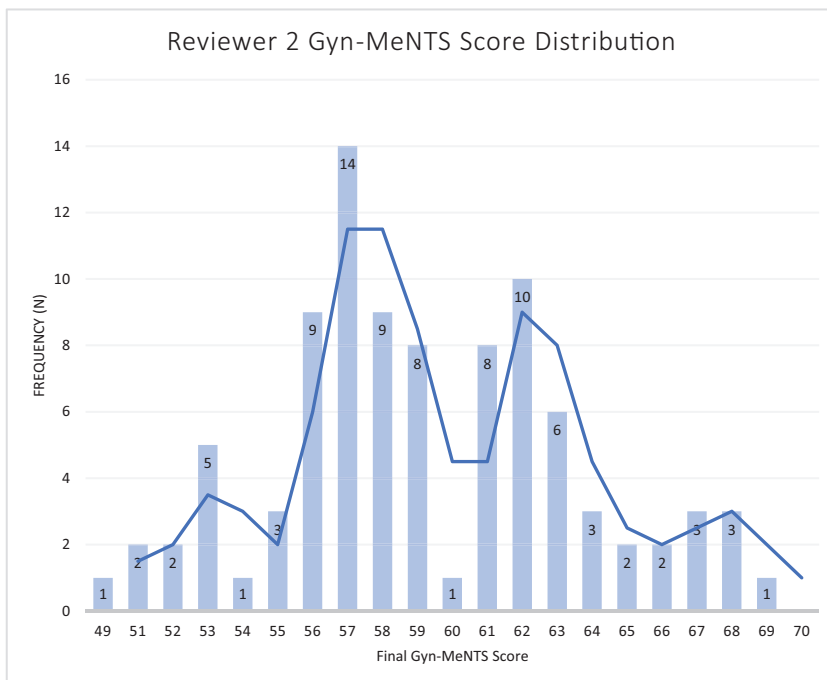


5.0, respectively. The Gyn-MeNTS scores for raters 1 and 3 did not differ (mean difference 0.03,  $p = .91$ ), whereas the scores for rater 2 differed significantly from both rater 1 (mean difference  $-1.2$ ,  $p < .001$ ) and rater 3 (mean

difference 1.2,  $p < .001$ ). Spearman  $r$ 's for all 3 rater pairs were  $>0.80$  (0.84 for 1 vs 2, 0.82 for 1 vs 3, and 0.82 for 2 vs 3, all  $p < .001$ ) indicating strong agreement in the patient rankings, with rater 2 having a constant added to

**Fig. 3**

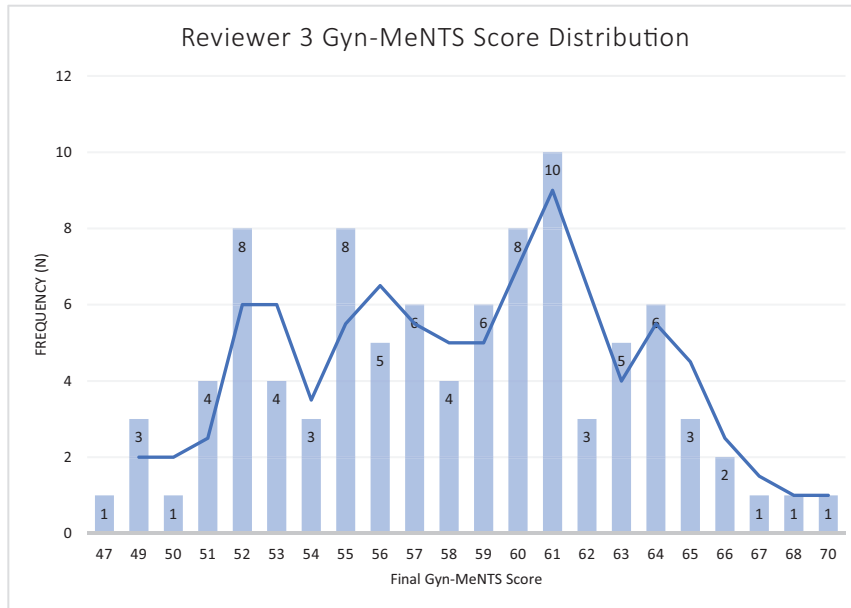
Reviewer 2 distribution of Gyn-MeNTS score. Gyn-MeNTS = Gynecologic Medically-Necessary Time-Sensitive.





**Fig. 4**

Reviewer 3 distribution of Gyn-MeNTS score. Gyn-MeNTS = Gynecologic Medically-Necessary Time-Sensitive.



each patient’s score. The Gyn-MeNTS items with the worst interrater agreement were operating room time, anticipated estimated blood loss, trial of alternative therapy, and pain severity (Table 3).

The mean ± standard deviation mESAS ratings were 1.40 ± 0.61 for rater 4 and 1.41 ± 0.61 for rater 5 (not different, p = .84). The weighted κ for the 2 mESAS raters was 0.57 (95% confidence interval [CI], 0.40–0.73) indicating

**Table 3**

Percent of patients with perfect interrater agreement on Gyn-MeNTS items between raters 1, 2, and 3

Variable	1 vs 2	1 vs 3	2 vs 3	Average
OR time	42	53	62	52
Estimated LOS	95	99	98	97
Postoperative ICU need	96	96	100	97
Anticipated EBL	72	74	73	73
Surgical team size	99	89	87	92
Anesthesia requirement	97	99	98	98
Surgical site	95	89	92	92
Trial of alternative therapy	28	41	32	34
Pain	58	63	49	57
Anemia	100	100	100	100
Impact on desired immediate fertility	97	97	97	97
Disease impact on GI/GU morbidity	96	94	95	95
Impact of 6-week delay in disease outcome	87	84	80	84
Impact of 6-week delay in surgical difficulty/risk	99	99	100	99
Age	100	100	100	100
Lung disease	100	100	100	100
Obesity/OSA	100	100	100	100
CV disease	100	100	100	100
Diabetes	100	100	100	100
Immunocompromised	100	100	100	100
Exposure to known COVID + person in past 14 days	100	100	100	100

COVID = coronavirus disease; CV = cardiovascular; EBL = estimated blood loss; GI = gastrointestinal; GU = genitourinary; Gyn-MeNTS = Gynecologic Medically-Necessary Time-Sensitive; ICU = intensive care unit; LOS = length of stay; OSA = obstructive sleep apnea; OR = operating room.

only fair agreement between raters. Specifically, there was 73% agreement in those categorized as level 1, 38% agreement in level 2, and 50% agreement in level 3.

**Concurrent Validity**

Gyn-MeNTS scores showed fair correlation with mESAS (Spearman  $r = -0.31$ ,  $p = .003$ ), and a moderate correlation with urgent SSP (Spearman  $r = -0.46$ ,  $p < .001$ ). However, mESAS had a slightly stronger correlation with urgent SSP (Spearman  $r = 0.53$ ,  $p < .001$ ). When used together to predict urgent SSP, both mESAS and Gyn-MeNTS had significant independent contributions, and the overall model had an area under the curve of 0.89, indicating excellent discrimination. For each 1-unit increase in mESAS, the odds of being labeled as urgent SSP increased by 12.78 (95% CI, 3.65–44.76;  $p < .001$ ), after adjusting for the Gyn-MeNTS score. For each 1-unit increase in the Gyn-MeNTS score, the odds of being urgent SSP decreased by 0.75 (95% CI, 0.62–0.90;  $p = .002$ ), after adjusting for the mESAS score. When the probabilities derived from this model were coded into quartiles, the incidence of being

urgent SSP was 75% in the highest priority quartile, 14% and 12% in the middle 2 quartiles, and 0% in the lowest quartile ( $p < .001$ ). The equation for calculating the probabilities is in the Supplemental Appendix. The 67th, 75th, 80th, 85th, 90th, and 95th percentiles of the Gyn-MeNTS score are shown in Table 4.

To make it easier to visualize these relationships, we coded the Gyn-MeNTS into 3 levels: low, medium, and high priority, on the basis of quartiles: the lowest quartile of scores was highest priority, the middle 2 quartiles were medium priority, and the highest quartile of scores was the lowest priority. We then compared these 3 levels of Gyn-MeNTS urgency and the 3 levels of SGS priority with the urgent SSP (Table 5).

Comparing Gyn-MeNTS with urgent SSP, we found that from the highest priority quartile on Gyn-MeNTS to the lowest, 52%, 23%, and 4% of the patients were urgent SSP ( $p = .001$ ). For patients with mESAS levels of high, moderate, and low priority, the percentages found within the urgent SSP category were 88%, 36%, and 10%, respectively ( $p < .001$ ).

When looking only at the 12 of 24 patients who were deemed the highest priority by the SSP scheme and their surgeries were performed without delay, the mESAS

**Table 4**

Association of Gyn-MeNTS percentiles with SGS mESAS rank

Percentile	n	Gyn-MeNTS score	SGS mESAS rank			p
			1 (%)	1.5–2 (%)	2.5–3 (%)	
95th	5	≤51	1 (20)	3 (60)	1 (20)	.23
90th	11	≤53	3 (27)	7 (64)	1 (9)	.10
85th	14	≤53.7	3 (21)	10 (71)	1 (7)	.009
80th	19	≤54.4	6 (32)	11 (58)	2 (11)	.05
75th	23	≤55	6 (26)	13 (57)	4 (17)	.02
67th	33	≤56	12 (36)	16 (48)	5 (15)	.044

Gyn-MeNTS = Gynecologic Medically-Necessary Time-Sensitive; mESAS = modified Elective Surgery Acuity Scale; SGS = Society for Gynecologic Surgeons. SGS mESAS score is average score between raters 4 and 5. Gyn-MeNTS score is average score between raters 1, 2, and 3.

**Table 5**

Association of Gyn-MeNTS priority scores and SGS mESAS rank versus SSP rank

Rating scale	Urgent SSP		p
	Lower priority	Highest priority	
Gyn-MeNTS			.001
Top priority quartile (%)	11 (48)	12 (52)	
Middle priority 2 quartiles (%)	36 (77)	11 (23)	
Lowest priority quartile (%)	22 (96)	1 (4)	
SGS mESAS			<.001
Priority 3 (top) (%)	1 (13)	7 (88)	
Priority 2 (middle) (%)	21 (64)	12 (36)	
Priority 1 (lowest) (%)	47 (90)	5 (10)	

Gyn-MeNTS = Gynecologic Medically-Necessary Time-Sensitive; mESAS = modified Elective Surgery Acuity Scale; SGS = Society for Gynecologic Surgeons; SSP = surgeon self-prioritization.

system was able to capture 92% of these patients in its most urgent quartile, whereas the Gyn-MeNTS system captured only 67% in its most urgent quartile.

## Discussion

Despite finding overall high interrater reproducibility in the Gyn-MeNTS scoring system (Spearman  $r$  0.82–0.84), it does not seem that this scoring system strongly discriminates the most urgent cases as determined by either the mESAS system (Spearman  $r$  0.31) or when surgeons proceed with using their instinct alone, the SSP (Spearman  $r$  0.46). The interrater reproducibility of the mESAS–tiered system was moderate (weighted  $\kappa$  0.57), and it appears to perform slightly better in discerning how surgeons instinctively prioritize (Spearman  $r$  0.53). The mESAS system identified 92% of the patients classified as most urgent SSP, whereas the Gyn-MeNTS found only 67%. However, when used together, the 2 scoring systems had high discrimination in capturing clinicians' instinctive beliefs about urgency and each contributed independently, suggesting that (a) they capture distinct issues related to urgency and (b) their combined use may provide the optimal system for objectively rating surgical urgency.

### Rationale for Specific Gyn-MeNTS Alterations

Despite seeing merit in the original scoring system, we believed that additional steps could be taken to make the MeNTS model more objective, yielding higher interobserver reliability. Our goal was to create a modified gynecologic MeNTS that would still score in comparable ranges with the original in the event our institution later decided to prioritize by this route. Most modifications were made within the disease factors category. We believed that the category should be given more weight of importance (with more scored items) and tailored to the disease burden that gynecologic patients uniquely incur. Thus, we expanded the number of scoring categories and created objective criteria for quantifying pain, anemia, fertility impact, and impact on adjacent organ systems such as the genitourinary and gastrointestinal systems.

Although we agree that exposure to known COVID-19 should be considered, we recommend that, if available, all patients undergoing scheduled surgery be tested for COVID-19 within 48 hours of their planned surgery. If a patient's result is positive, the surgery should be delayed. If testing is unavailable, we recommend screening all patients undergoing elective surgery for influenza-like symptoms, and the surgery should be postponed if present. Given the reports of significantly worsened morbidity and mortality when surgery is unwittingly performed in patients who are presymptomatic for COVID-19, all attempts to identify these patients should be made [9,10]. In addition, universal COVID-19 testing protects healthcare workers from unnecessary exposure, as well as creates a binary personal protective equipment triaging system to “standard precautions” or “transmission-based precautions” to protect scarce resources.

### Points of Consideration in Implementation of Gyn-MeNTS

Despite our attempts to make the scoring system as objective as possible, many categories are open to wider interpretation than it may first seem. We recommend, before implementing the scoring system that an initial dialogue to “lay the ground rules” occurs to improve reviewer reproducibility and thus reliability across the cohort. For example:

- (1) Lung Disease. Attempt to define what constitutes “minimal” disease. Our system did not capture smoking history other than a binary yes/no. Quantifying risk of disease in a pack/year history calculation may prove beneficial as smoking affects risks of both respiratory disease and wound healing with a direct impact on outcomes. It is also becoming increasingly known that lung disease, such as chronic obstructive pulmonary disease) and asthma, can worsen patient outcomes in the setting of COVID-19, independent of their usual perioperative risks [11].
- (2) Surgical team size. We agreed that team size would be calculated by the *minimum* number of surgeons needed to perform a procedure safely and efficiently. In an academic teaching institution, the size of the surgical teams can easily be twice the actual number needed. Thus, all hysteroscopies (that did not require intraoperative sonography) were scored a 1, and almost all laparoscopies a 2. When constant uterine manipulation was needed, a score of 3 was given to laparoscopy.
- (3) Operating room time. At our institution, we are asked to provide estimates of “wheels-in” to “wheels-out” rather than “incision-to-closure” time when posting cases. As long as consistency is applied across graders, this variable has the potential to be reliable. Given surgeon notoriety around being poor predictors of needed surgical time, it is not surprising that this variable performed among the worst in interrater agreement.

### Limitations of Gyn-MeNTS

- (1) Pain. Despite our attempt to objectively define pain using a visual analog scale system and incorporating the ability to control this pain on the basis of frequency of office or emergency room visits, our interrater agreement remained fair (agreement = 57%) (Table 3). This was due, in large part, to provider differences in the documentation of pain, its impact on quality of life, and the ability to control pain with medical management.
- (2) Alternative therapy. We created objective parameters for measuring alternative therapy that included an extensive list of examples (Table 2 footnotes) and accounted for number of therapies trialed and burden of resource

utilization/exposure risk. Despite this, our interrater agreement was poor (agreement = 34%). We believe this was due to the difficulty in abstracting this information from charts.

- (3) Prediction of perioperative transfusion. We agree that the presence of anemia can be inconsequential when it comes to procedures with a risk of low blood loss. The counter is true as well; high blood loss procedures can be tolerated when the patient has no baseline anemia. The most important question, particularly during a time of blood shortage, is whether and to what degree the patient will need perioperative blood products. Finding an objective way to quantify this risk was difficult, a problem compounded by the surgeons' inability to reliably predict estimates of blood loss (agreement = 73%).
- (4) Cardiovascular disease. While simplifying the determination of heart disease to represent the number of medications it takes to control it, care should be taken to avoid this simple assumption. Patients with untreated hypertension may carry significantly more heart disease than patients who are well controlled on 3 medications owing simply to their access to healthcare and their compliance with recommended therapy.
- (5) Emphasis on hospital resource use with a bias toward the young, healthy patient. Perhaps the biggest limitation to the Gyn-MeNTS scoring system is its favor toward quick procedures on patients who are young and healthy. Despite our attempts to increase the weight of disease burden by adding more graded variables, patients who are young and healthy getting quick, elective procedures (e.g., tubal ligation or polypectomy) were consistently prioritized in this grading system. Although not an invalid conclusion when a hospital system is severely limited in its capacity to do anything but the most quick and simple of procedures on patients who are healthy, it would be difficult to justify, for example, elective sterilization over treatment of debilitating pain from endometriosis, especially if patients have access to alternative contraception. In our institution, despite receiving some of the lowest Gyn-MeNTS scores, patients requesting sterilization are not being prioritized at this time.
- (6) We also must acknowledge that many of our changes to the original MeNTS scoring system could be specific to the study institution, which thus affects the generalizability of the Gyn-MeNTS system.

### **Limitations of mESAS**

Despite the extensive list of surgical examples given by major gynecologic surgical societies in the mESAS table,

we found inconsistencies with application between reviewers. It is clear there is still room for interpretation of disease severity, and, thus acuity, making assignment of tiers prone to significant variation. In addition, at the end of this exercise, a high-volume institution could still have large numbers of patients within each cohort that must be prioritized further, and this system provides no guidance as to how to perform this. Finally, the mESAS system appears to contain some inconsistencies within their surgical examples. For instance, endometriosis with poorly controlled pain and desire for fertility is categorized as a level 2, whereas myomectomy for a patient who is asymptomatic and is experiencing infertility is categorized as a level 3. A hysteroscopic polypectomy in the patient who is infertile is categorized more urgently (level 3) than a hysteroscopic evaluation or polypectomy in patients older than 50 years with an inability to sample in the office (level 2) that have higher risks of malignant potential. Although these surgical assignments have been agreed on by major societal stakeholders, these discrepancies deserve attention.

### **Limitations of this Study**

To a certain extent, both scoring systems depend on accurate and elaborate chart documentation. Ideally, surgeons would grade their own patients to improve accuracy and overcome this obstacle. Difficulty was encountered when reviewers graded each other's patients in categories such as efficacy of alternative therapies, immediacy of fertility desire, and severity of pain. This clearly affected the interrater reliability and could lead to the false assumption that poor interrater reliability means that a scoring system is invalid. Finally, further evaluation in a setting with clinicians who did not develop the scoring system is warranted.

### **Strengths of this Study**

To our knowledge, this is the first study to assess reliability and validity of previously published surgical scoring systems. In addition, we are the first to report application of these scoring systems in gynecologic patients and have made recommendations for implementation in this arena.

### **Proposals for Future Research**

More robust prospective data are needed to either confirm or refute our retrospective findings. Further study should also evaluate the efficacy of using both the Gyn-MeNTS and mESAS systems together to triage nonemergent procedures. We also believe strongly that a system that provides more emphasis on the disease variables of the Gyn-MeNTS scoring system would yield an even more valid triaging system.

## References

1. American College of Surgeons: COVID-19: recommendations for management of elective surgical procedures. Available at: <https://www.facs.org/covid-19/clinical-guidance/elective-surgery> Accessed May 2, 2020.
2. U.S. Surgeon General: Hospitals & healthcare systems, please consider stopping elective procedures until we can #FlattenTheCurve. Available at: [https://twitter.com/Surgeon\\_General/status/1238798972501852160](https://twitter.com/Surgeon_General/status/1238798972501852160) Accessed May 6, 2020.
3. U.S. Centers for Medicare & Medicaid Services: CMS releases recommendations on adult elective surgeries, non-essential medical, surgical, and dental procedures during COVID-19 response. Available at: <https://www.cms.gov/newsroom/press-releases/cms-releases-recommendations-adult-elective-surgeries-non-essential-medical-surgical-and-dental> Accessed May 7, 2020.
4. American College of Surgeons: COVID-19: guidance for triage of non-emergent surgical procedures. Available at: <https://www.facs.org/covid-19/clinical-guidance/triage> Accessed May 3, 2020.
5. Society of Gynecologic Surgeons: Joint statement on re-introduction of hospital and Office-Based Procedures in the COVID-19 climate for the practicing gynecologist. Available at: [https://www.sgsonline.org/index.php?option=com\\_content&view=article&id=638;joint-statement-on-re-introduction-of-hospital-and-office-based-procedures-in-the-covid-19-climate&catid=20:site-content](https://www.sgsonline.org/index.php?option=com_content&view=article&id=638;joint-statement-on-re-introduction-of-hospital-and-office-based-procedures-in-the-covid-19-climate&catid=20:site-content) Accessed May 2, 2020.
6. Prachand VN, Milner R, Angelos P, et al. Medically necessary, time-sensitive procedures: scoring system to ethically and efficiently manage resource scarcity and provider risk during the COVID-19 pandemic. *J Am Coll Surg.* 2020;231:281–288.
7. Joint statement on minimally invasive gynecologic surgery during the COVID-19 pandemic. *J Minim Invasive Gynecol.* 2020;27:1027–1029.
8. Society of American Gastrointestinal and Endoscopic Surgeons. SAGES and EAES recommendations regarding surgical response to Covid-19 crisis. Available at: <https://www.sages.org/recommendations-surgical-response-covid-19/> Accessed May 7, 2020.
9. Lei S, Jiang F, Su W, et al. Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19 infection. *EClinicalMedicine.* 2020;21:100331.
10. Nahshon C, Lavie O. Author's reply: hazardous postoperative outcomes of unexpected COVID-19 infected patients: a call for global consideration of sampling all asymptomatic patients before surgical treatment. *World J Surg.* 2020;44:3194–3195.
11. Fang X, Li S, Yu H, et al. Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis. *Aging (Albany NY).* 2020;12:12493–12503.

### APPENDIX I. Equation for calculating probability of high urgency by clinician behavior

$$Y = 11.628 + 2.548 * SGS - 0.29 * Gyn-MeNTS$$

$$\text{Probability} = \exp(Y) / (1 + \exp(Y))$$