Background: Stress has been shown to impact adversely on multiple facets critical to optimal performance. Advancements in wearable technology can reduce barriers to observing stress during surgery. This study aimed to investigate the association between acute intraoperative mental stress and technical surgical performance.

Methods: Continuous electrocardiogram data for a single attending surgeon were captured during surgical procedures to obtain heart rate variability (HRV) measures that were used as a proxy for acute mental stress. Two different measures were used: root mean square of successive differences (RMSSD) and standard deviation of RR intervals (SDNN). Technical surgical performance was assessed on the Operating Room Black Box® platform using the Generic Error Rating Tool (GERT). Both HRV recording and procedure video recording were time-stamped. Surgical procedures were fragmented to non-overlapping intervals of 1, 2 and 5 min, and subjected to data analysis. An event was defined as any deviation that caused injury to the patient or posed a risk of harm.

Results: Rates of events were significantly higher (47–66 per cent higher) in the higher stress quantiles than in the lower stress quantiles for all measured interval lengths using both proxy measures for acute mental stress. The strongest association was observed using 1-min intervals with RMSSD as the HRV measure ($P < 0.001$).

Conclusion: There is an association between measures of acute mental stress and worse technical surgical performance. Further study will help delineate the interdependence of these variables and identify triggers for increased stress levels to improve surgical safety.

Funding information
No funding

Paper accepted 6 August 2018
Published online 27 September 2018 in Wiley Online Library (www.bjsopen.com). DOI: 10.1002/bjs5.104
found to be impacted negatively by increased stress. This is particularly relevant to the surgical environment, where stress and stressors have been shown to adversely impact functions such as decision-making and team performance. Studies have observed stress in the live operating theatre using a variety of methods, but limitations in ECG equipment and stress measurement methods have hindered research. Efforts to investigate this relationship have generally been limited to virtual environments. Although these studies have provided valuable insight, surgical simulations lack realism in comparison with the real surgical environment when imitating a stress scenario. Significant advances in wearable technology, particularly the ability reliably to measure physiological metrics unobtrusively without wired sensors, afford new opportunities to overcome these obstacles.

Standards to assess technical surgical performance and objective methods to estimate acute stress with ECG measurements have existed for decades. Despite this, no studies have employed these tools in the live surgical setting to assess the relationship between intraoperative acute mental stress levels and technical surgical performance. The aim of the present study was to investigate the association between acute intraoperative mental stress and surgical performance.

Methods

Study design

Data collection for this observational study was conducted at Stanford Hospital (Stanford, California, USA) during May and June 2017. Stress and surgical performance data were collected for one attending surgeon during this period. The study was granted institutional review board approval by Stanford University. Patient consent was not required.

Data capture

Physiological data to be used in calculations for proxy measures of acute stress were captured for the attending surgeon during each surgical procedure using the wireless Hexoskin Smart Shirt (Hexoskin, Montreal, Quebec, Canada). The Hexoskin Smart Shirt has built-in sensors that capture 256-Hz ECG data, three-dimensional 64-Hz acceleration data and 128-Hz breathing data, all of which have been validated previously. All raw data are directly downloadable from the Hexoskin online portal, with all data points time-stamped to the Coordinated Universal Time (UTC). The 256-Hz ECG capture satisfies the recommended resolution to capture precise data to allow for HRV interpretations as set forth by the Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology.

Consecutive elective surgical procedures, both laparoscopic and endoscopic, performed between 9 May and 28 June 2017, were included in the study. Each procedure was recorded and entirely anonymized of all markers (hospital information, physician identifiers, patient data). Video recordings were time-stamped to UTC to allow for synchronization with the HRV data captured by the Hexoskin Smart Shirt. Recordings were assessed independently and blindly by a single observer for technical laparoscopic surgical performance using the Operating Room Black Box® platform (SST, Toronto, Ontario, Canada), with a previously validated measurement tool (Generic Error Rating Tool (GERT)).

Stress (heart rate variability) analysis

Although it has been recommended that both time domain and frequency domain statistics be incorporated when analysing HRV, the two statistical methods correlate with one another to such a strong degree that they can be used as surrogates for each other. Further, time domain methods, particularly RMSSD (defined below), have been shown to have greater reliability when measuring short-term HRV than frequency domain methods. The following conventional time domain methods of analysing HRV were therefore selected for this study (for formulas see Appendix S1, supporting information): the square root of the mean sum of squares of RR intervals (RMSSD) and standard deviation of RR intervals (SDNN).

As sympathetic nervous system activation begins to supersede parasympathetic nervous system activation, there becomes less interplay among these two ANS divisions, which manifests as a near-instantaneous decrease in RR-interval variation (lower HRV). As such, lower values for SDNN and RMSSD suggest greater levels of sympathetic nervous system dominance (hereafter referred to as stress).

The Hexoskin ECG processing software provides a filter for data curation of RR intervals that excludes abnormal RR intervals. These may be due to irregular heartbeats or suboptimal ECG signals.

Surgical performance

Technical laparoscopic surgical performance was evaluated with the GERT, a previously validated measurement tool.
The GERT allows for objective and reliable (intrater intraclass correlation coefficient for events 0.85) assessment of operative performance during laparoscopic procedures. Using this framework, both surgical errors and events are identified, with errors defined as ‘the mechanism of unintended or deviated technical task execution’ and events as ‘any deviation that causes injury to the patient or poses a risk of harm’41,45. For the purposes of this study, only ‘events’ were selected as an outcome measure of surgical performance, because they occur less commonly and have a more direct relationship with clinically relevant outcomes. Examples of events include serosal tears arising from grasper slips from bowel, bleeding from needle puncturing vessel, and burns from inadvertent touching of other structures with an energy device41.

**Data analysis**

Software was developed by the authors to derive the HRV statistics from the ECG data files, process the surgical performance data and synchronize all data streams into the prespecified time intervals. The first interval in a procedure commenced the moment after the conclusion of the presurgery ‘time out’ checklist where video capture had begun. Data capture concluded at the end of the surgery when all surgical instruments were removed from the patient.

There is some debate in the literature regarding the optimal time intervals to ensure valid HRV recordings, with the Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology40 recommending that short-term HRV analysis be undertaken in intervals of 5 min, and others46 suggesting that accurate HRV recordings can be taken in intervals as low as 10 s. Consideration was also given to the fact that data streams were quite distinct. For example, although physiological data were measured on a pseudo-continuous timeline (ECG readings at 256 Hz), HRV was measured in discrete intervals and surgical performance on a continuous basis. Thus, for the exploratory purposes of this study, the judgement was made to measure both data streams (HRV measurements and surgical performance) in non-overlapping intervals of 1, 2 and 5 min to best accommodate these differences.

Stress data were divided into quantiles to allow for comparison between low and high stress intervals. Five-minute intervals were presented in quintiles, rather than tertiles, owing to the reduced sample variance resulting from the increased interval length, as stress peaks may last only a fraction of the duration of an interval.

**Statistical analysis**

All statistical analyses were performed using Microsoft Excel® version 16 (Microsoft, Redmond, Washington, USA), SPSS® version 24 (IBM, Armonk, New York, USA) and MATLAB version 9.2 (MathWorks, Natick, Massachusetts, USA).

Stress data were assumed to be normally distributed, so comparisons between groups were performed with independent two-sample Student’s *t* tests. Surgical performance (events) was observed not to be normally distributed; therefore comparisons between groups were performed with Mann–Whitney *U* tests. *P* < 0.050 was considered statistically significant.
Table 1 Changes in heart rate variability measure between intervals with and without a surgical event

<table>
<thead>
<tr>
<th>Interval</th>
<th>HRV measure</th>
<th>0 events</th>
<th>≥ 1 event</th>
<th>Difference (%)</th>
<th>Student's t test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t statistic</td>
</tr>
<tr>
<td>1 min</td>
<td>SDNN</td>
<td>37.0</td>
<td>34.4</td>
<td>−7</td>
<td>2.93 (0.80, 4.44)</td>
</tr>
<tr>
<td></td>
<td>RMSSD</td>
<td>24.8</td>
<td>22.6</td>
<td>−9</td>
<td>2.10 (0.14, 4.30)</td>
</tr>
<tr>
<td>2 min</td>
<td>SDNN</td>
<td>40.4</td>
<td>38.1</td>
<td>−6</td>
<td>2.47 (0.47, 4.13)</td>
</tr>
<tr>
<td></td>
<td>RMSSD</td>
<td>25.3</td>
<td>25.5</td>
<td>+1</td>
<td>0.20 (−2.17, 2.65)</td>
</tr>
<tr>
<td>5 min</td>
<td>SDNN</td>
<td>43.7</td>
<td>41.9</td>
<td>−4</td>
<td>1.58 (−0.43, 4.09)</td>
</tr>
<tr>
<td></td>
<td>RMSSD</td>
<td>27.4</td>
<td>25.9</td>
<td>−5</td>
<td>0.97 (−1.54, 4.49)</td>
</tr>
</tbody>
</table>

Values in parentheses are 95 per cent confidence intervals. HRV, heart rate variability; SDNN, standard deviation of RR intervals; RMSSD, root mean square of successive differences in RR intervals.

Results

Data were captured for gastric bypass (12 patients), sleeve gastrectomy (7) and peroral endoscopic myotomy (6), with a mean(s.d.) duration of 96.7 (35.5) min for all procedures. As the rating framework for assessing surgical performance was tailored for laparoscopic procedures, surgical performance data were observed only for gastric bypass and sleeve gastrectomy. Stress data were observed for all procedures. Mean(s.d.) SDNN values were 38.0 (14.4) ms for 1-min intervals, 40.9 (12.2) ms for 2-min intervals and 43.7 (9.7) ms for 5-min intervals. Respective mean(s.d.) RMSSD values were 25.2 (16.5), 26.2 (14.5) and 27.1 (12.5) ms. The mean(s.d.) number of events per procedure was 23.8 (11.3).

Stress and surgical performance

For all time intervals there was a trend for an increasing rate of events with increasing stress levels (Fig. 1). For 1-, 2- and 5-min intervals, the highest stress (SDNN) quantiles had event rates that were 55.5, 62.5 and 66.1 per cent respectively greater than those of the lowest stress quantiles, all with statistically significant differences (1 min: \( P = 0.002 \); 2 min: \( P = 0.011 \); 5 min: \( P = 0.028 \)). The same trend was observed with RMSSD as the proxy for stress, with the highest stress quantiles having 57.7, 46.8 and 54.7 per cent greater event rates than the lowest stress quantiles for each respective interval (1 min: \( P < 0.001 \); 2 min: \( P = 0.045 \); 5 min: \( P = 0.046 \)).

When observing the inverse relationship, where mean HRV values were calculated for intervals with and without events, a similar trend was observed (Table 1). With the exception of RMSSD for 2-min intervals, HRV was lower (stress was greater) as measured by both SDNN and RMSSD during intervals where an event was observed versus intervals where an event was not observed. Levene’s test for equality of variances showed equal variances for each test except for 2-min RMSSD.

Discussion

This study has demonstrated a direct relationship between acute mental stress and surgical performance in a live operating theatre. The results indicate a clear association between the occurrence of events and the level of stress experienced, with significant differences between the higher and lower stress quantiles being observed for all measured time intervals. Greater stress responses were also found during intervals where an event was observed, for all measures except the 2-min interval RMSSD.

There are several possible explanations for this finding. One may be that increased stress levels in surgeons occur in response to the occurrence of events. Although this is likely to be true, it does not preclude the possible dependency of surgical performance on stress levels. Not all events are of equal severity and are therefore not necessarily recognized by the surgeon as a deviation that causes injury or poses a risk to the patient. Further, events that are recognized by the surgeon may not manifest in stress responses if they are perceived to be innocuous and easily resolved (for example, minor bleeding). In addition, the possibility of a snowball effect cannot be precluded, whereby an event triggers a stress response that could have a negative impact on surgical performance resulting in subsequent errors and events.

Based on previous research on the effects of acute mental stress and the observations in this study, acute mental stress and surgical performance are likely to be dependent upon each other to some degree. The next step towards better understanding of this relationship would be a study of temporality. For obvious ethical reasons, this cannot be performed in the operating theatre by experimental means (that is, inducing a stress response in a surgeon and observing surgical performance), and therefore would have to be done via novel methods of interval analysis (if incorporating HRV). In this study, the intervals were predetermined from the moment that the video recording commenced, thereby introducing several limitations.
The true relationships may be more pronounced than has been found in this study. Stress responses vary greatly in length depending on the situation and, unless this stressed state fell perfectly within the bounds of an interval, the stressed state would either be diluted with more relaxed HRV data (if the stress response were shorter than the interval length) or be only partly captured (if longer than the interval length). Observing at three different interval times was an attempt to control for this, but these methods could still be improved. In addition, an event occurring near the end of an interval would be unlikely to prompt an increased stress response during that same interval. Conversely, compromised surgical skill because of a raised stress response near the end of an interval may not trigger events until the succeeding interval. Similarly, if the event occurred at the very beginning of an interval, any preceding stress increase that may have been a factor would not be grouped into the same interval.

Despite these limitations, the observed trends of the potential effect of stress on the occurrence of events should trigger further research. In particular, it would be helpful to study stress in intervals on a ‘sliding window’ that moves along a continuous timeline rather than fixed predetermined intervals. The width of this window should be flexible to accommodate varying lengths of stress response. In addition, it would be interesting to study the temporality of the relationship between stress and surgical performance by comparing stress levels during the intervals preceding and succeeding events with baseline periods.

The clinical relevance of exploring variables associated with surgical events is evident. Stress can be induced in surgeons and other members of the operating team for a variety of reasons, many of which are unavoidable. It is therefore beneficial to distinguish between stress that manifests as a result of surgical events that demand increased concentration, and stress that arises as a result of external influences. The latter can be addressed by initiatives that identify and reduce such external events, whereas managing the former, which may or may not be harmful for surgical performance depending on the specific circumstance, is more complex.

It is important to explore methods, either training- or intervention-based, that diminish the occurrence, severity and potential consequences of this unavoidable phenomenon of surgical stress. Previous research20,47 has shown promise in this area and, given the results of the present observational study and the clear association between stress levels and surgical performance, the implementation of intervention- or training-based initiatives aimed at reducing stress in surgeons may be a viable way to reduce surgical events and adverse outcomes. Further study is also necessary to gauge more precisely the triggers for increased stress, their individual effects on surgical performance, and the severity of this relationship.

Disclosure
The authors declare no conflict of interest.

References
15 Böhm B, Rötting N, Schwenk W, Grebe S, Mansmann U. A prospective randomized trial on heart rate variability of the...


**Supporting information**

Additional supporting information can be found online in the Supporting Information section at the end of the article.