PNAS RESEARCH ARTICLE | NEUROSCIENCE STATISTICS OPEN ACCESS

Monitoring surgical nociception using multisensor physiological models

Sandya Subramanian[a,1], Bryan Tseng[b], Marcela del Carmen[c], Annekathryn Goodman[c], Douglas M. Dahl[c], Riccardo Barbieri[d], and Emery N. Brown[a,b,c,1]

Affiliations are included on p. 9.

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, La Jolla, CA; received November 22, 2023; accepted June 30, 2024

Monitoring nociception, the flow of information associated with harmful stimuli through the nervous system even during unconsciousness, is critical for proper anesthesia care during surgery. Currently, this is done by tracking heart rate and blood pressure by eye. Monitoring objectively a patient's nociceptive state remains a challenge, causing drugs to often be over- or underdosed intraoperatively. Inefficient management of surgical nociception may lead to more complex postoperative pain management and side effects such as postoperative cognitive dysfunction, particularly in elderly patients. We collected a comprehensive and multisensor prospective observational dataset focused on surgical nociception (101 surgeries, 18,582 min, and 49,878 nociceptive stimuli), including annotations of all nociceptive stimuli occurring during surgery and medications administered. Using this dataset, we developed indices of autonomic nervous system activity based on physiologically and statistically rigorous point process representations of cardiac action potentials and sweat gland activity. Next, we constructed highly interpretable supervised and unsupervised models with appropriate inductive biases that quantify surgical nociception throughout surgery. Our models track nociceptive stimuli more accurately than existing nociception monitors. We also demonstrate that the characterizing signature of nociception learned by our models resembles the known physiology of the response to pain. Our work represents an important step toward objective multisensor physiology-based markers of surgical nociception. These markers are derived from an in-depth characterization of nociception as measured during surgery itself rather than using other experimental models as surrogates for surgical nociception.

surgical nociception | autonomic nervous system | anesthesia | multimodal | point process

Significance

Monitoring nociception, the flow of information associated with harmful stimuli through the nervous system even during unconsciousness, is critical for proper anesthesia care during surgery. Currently, this is done by tracking heart rate and blood pressure by eye. Objectively monitoring a patient's nociceptive state remains a challenge, causing drugs to often be over- or underdosed intraoperatively. Postoperative consequences include poor pain control and cognitive dysfunction. In a prospective study of 101 surgeries containing almost 50,000 nociceptive stimuli, we demonstrate objective quantification of nociception throughout surgery. Our findings show that a statistically rigorous multisensor approach identifies a physiologically consistent and reliable nociceptive signature that occurs in the presence of nociceptive stimuli. This signature represents a principled approach to guiding anesthetic dosing.

General anesthesia consists of four simultaneous components: lack of pain processing or antinociception, unconsciousness, amnesia or lack of memory, and muscle relaxation or paralysis while maintaining physiologic stability (1). In the case of general anesthesia, the unconscious processing of pain via primal reflexes at the level of the spinal cord and brainstem is termed nociception (2). One of the primary purposes of general anesthesia is to prevent nociception during surgery (2–4). Throughout this paper, nociception during general anesthesia will be referred to as surgical nociception. The current approach to manage surgical nociception is largely dependent on the individual intuition of the anesthesia caregiver. The only available objective measures involve the anesthesiologist tracking monitored heart rate and blood pressure, while accounting for factors such as patient's age, BMI, the nature of the surgery, and the antinociceptive properties of administered medication (2–4). To complicate matters further, heart rate and blood pressure are highly nonspecific indicators of nociception and are often directly managed through the use of cardiac medications during surgery (2). Altogether, the standard of care results in frequent underdosing/overdosing of intraoperative pain medication (2). Underdosing can lead to the patient waking up in pain, delayed recovery and healing, postoperative cognitive dysfunction, and prolonged hospital stay. Overdosing can delay patient awakening, increase side effects such as nausea, vomiting, and constipation, exacerbate cognitive side effects such as delirium, and prolong hospital stay (2, 4, 5). The bottom line is that improved management of surgical nociception can ensure antinociception during surgery while also reducing the prevalence of undesirable postoperative outcomes including cognitive impairment, postoperative pain (6), and nausea and vomiting.

An important step to improve nociception management is to develop objective measures for tracking surgical nociception. The standard of care for quantification of pain in clinical settings is still limited to subjective 1 to 10 pain scales that can only be used when a patient is conscious and capable of providing meaningful responses (4, 7). The development of objective measures for monitoring surgical nociception could serve as a stepping stone

Competing interest statement: Two utility patent applications have been filed with S.S., R.B., and E.N.B. covering the methods in this manuscript (US 2023-0363702).

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: sandya@mit.edu or enb@neurostat.mit.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2319316121/-/DCSupplemental.

Published September 24, 2024.

toward future assessment of pain in postoperative settings, disorders of consciousness, or even chronic pain. This is a necessary step for future closed-loop nociceptive control systems. However, it is critical that such measures be constructed from data about surgical nociception. It may be tempting to use animal models of pain or nociception or controlled experimental human models of pain, such as cutaneous thermally or electrically induced pain models (8). However, these are nowhere near the magnitude nor intensity of surgical nociceptive stimuli. Subjects are not even anesthetized in the majority of experimental pain models.

A popular approach employed for tracking surgical nociception has been based on autonomic nervous system (ANS) responses. This is justified by the fact that general anesthesia disrupts conscious processing of pain, which leaves the brainstem and spinal cord as the primary processing centers of nociceptive inputs (1, 9). These regions govern the most primal reflexes to nociception, including autonomic responses ("fight or flight"), through several autonomic nuclei of the brainstem, including the nucleus of the solitary tract, the rostral and caudal ventrolateral medulla, and the dorsal motor nucleus of the vagus (the nociceptive medullary autonomic circuit) (1). These nuclei outflow to the heart, blood vessels, sweat glands, etc. Therefore, autonomic outputs can be used to gauge the nociceptive state of the body, if the concurrent contributions of anesthetic and cardiac medications are also considered.

Several previous efforts have been made to define indices that track surgical nociception using autonomic responses. One of these is the Analgesic Nociception Index (ANI—now the HFVI), which uses a proprietary algorithm to extract information from the electrocardiogram (ECG) about heart rate variability (HRV) as a measure of cardiac autonomic activity (10). Other such measures have multiple sensors, relying on more than one physiological signal (11–14). To the best of our knowledge, several of these metrics were neither constructed from prospective surgical datasets focused on nociception nor do they have bases in physiological or statistical models. Most importantly, in validation studies, each surgery was condensed into a few timepoints and aggregated across subjects to obtain statistical significance (15–33). While this suggests the existence of important trends, such models would not be recommended for continuous monitoring of individual nociception during surgery.

In this study, we have collected prospectively a comprehensive observational dataset focused on surgical nociception, including over 100 subjects and recorded during 18,500 min of surgery. Using this dataset, we constructed indices and models able to track nociception successfully throughout the course of surgery. The dataset includes annotations of the timing and type of almost 50,000 nociceptive stimuli, the timing and doses of multiple classes of anesthetic medications, and over five separate physiological signals collected continuously throughout surgery (ECG, EDA, respirations, PPG, and skin temperature). Our dataset includes a variety of drug classes across several types of surgery. While our dataset is observational, it could not have been collected retrospectively. A majority of the nociceptive stimuli we annotated, such as cautery, as well as some of the physiological signals, such as electrodermal activity (EDA), are not recorded or collected as part of the standard of care, rendering the medical record an incomplete information source with respect to surgical nociception. Even the signals that are recorded, such as heart rate and blood pressure, have insufficient temporal resolution in the medical record.

To extract the most physiologically relevant information from the acquired signals, we rely on statistically rigorous point process models that underlie the physiology of heartbeat dynamics and EDA. We develop one model for the generation of heartbeats, which allows for the computation of HRV indices, and another for sweat gland pulses, which cause EDA (34–38). We feed the indices from these models into interpretable supervised and unsupervised frameworks with embedded inductive biases (physiological insight) to characterize surgical nociception. The supervised frameworks include logistic regression (39) and random forest (40), while the unsupervised framework is a state space model (41) assuming that the measurable observations (HRV and/or EDA indices) are driven by two hidden states related to the patient's time-varying nociceptive state. Based on physiology, we postulate that there should be at least two components to the autonomic state (sympathetic and parasympathetic). Therefore, we hypothesized two hidden states. We evaluate the performance of these frameworks, or the hidden states returned by the state space models, using the area under the receiver operating characteristic curve (AUROC) as well as the area under the precision–recall curve (AUPRC), and compare them to the performance of the ANI (42). We also include information about the timing and dosage of anesthetic drugs in some of the supervised models. We use the coefficients estimated in each framework (or frequency of occurrence of different features in the random forest) to assess whether the frameworks are truly capturing physiologic nociception.

The remainder of this paper is organized as follows. In Results, we show that our indices and models can track nociception during surgery with higher accuracy than the ANI. We also show that our analysis demonstrates objective tracking of surgical nociception using both supervised (classification) and unsupervised approaches. Finally, we demonstrate that all of our models, both supervised and unsupervised, recover states that are physiologically consistent with nociception. In the Discussion, we explain the implications of our analysis and the results. In the *SI Appendix*, online Methods, we discuss the details of how we constructed the models of surgical nociception, including the model frameworks, data preprocessing, and results interpretation.

Results

Summary of the Subject Cohort. This study was approved by the Massachusetts General Hospital Human Research Committee. The cohort consisted of 101 subjects, 48 men and 53 women, ranging in age from 29 to 77 y old (written informed consent was obtained) (43). Most surgeries were laparoscopic, with the majority being prostatectomies for the men and hysterectomies for the women. We chose analogous surgeries for men and women, both focused on the lower abdominal area and laparoscopic. A handful of the 101 surgeries were either open or the procedures were converted to open partway through. The age distribution of the subjects was different between men and women (Fig. 1*A*) since the surgeries themselves have a different age demographic. The women who undergo hysterectomies and salpingo oophorectomies are typically younger than the men who undergo prostatectomies. The durations of the surgeries ranged from under one hour to almost 5 h (Fig. 1*B*). The salpingo oophorectomies and hysterectomies were typically shorter in duration than the prostatectomies. ANI data were either unable to be collected or not of sufficient quality to be included (at least 10 min of high-quality data across the full surgery) for 13 surgeries.

Indices Alone are not Sufficient to Detect Nociception. We report all of the physiologic indices computed from the acquired data and used in our models (Table 1) as well as subjects' preprocessed data with annotated times of nociceptive stimuli during surgery (Fig. 2 and *SI Appendix*, Figs. S9–S109). This includes our multisensor

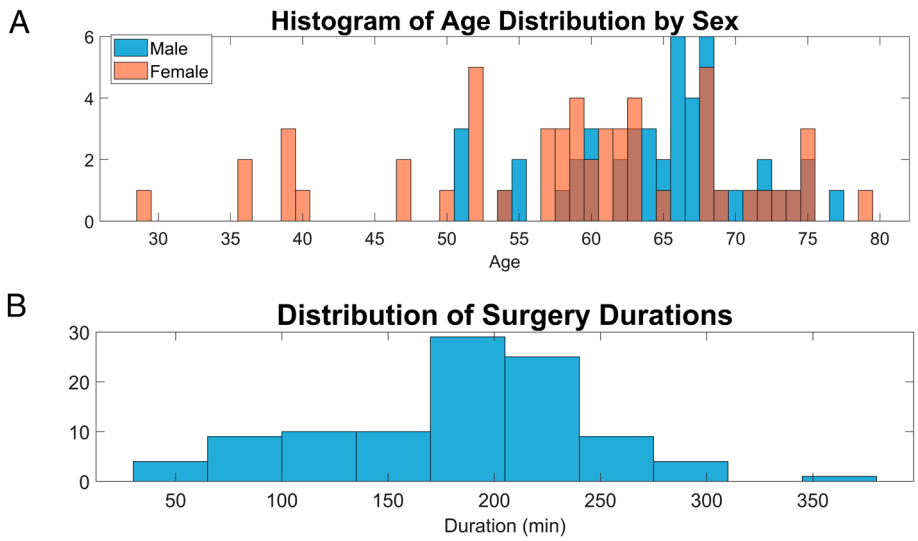

Fig. 1. Characterization of the subject cohort. (*A*) Distribution of age by sex. (*B*) Distribution of surgery durations.

HRV and EDA indices (Table 1) and the values returned by the ANI monitor (for more details on the computation of these indices, *Materials and Methods*). It is worth noting that using just the HRV, EDA, and ANI indices alone, it is not obvious to predict where nociceptive stimuli are occurring (Fig. 2 and *SI Appendix*, Fig. S8). This necessitates the use of more sophisticated models to capture patterns associated with nociception not detectable by eye alone.

Design of Study using Different Model Frameworks. All supervised models were fitted to predict whether or not each of the 223,017 5-s time windows contained a nociceptive stimulus. The raw ANI index was assessed for its ability to do so. For our physiologic indices, two supervised (logistic regression and random forest) and one unsupervised model frameworks (state space model) were used. All indices (Table 1) as well as numerical estimates of their local derivatives over time were used as features for the supervised models. Only the indices themselves were used as observations for the state space model (no estimates of derivatives). For the state space model, the estimated hidden states were assessed as markers of nociception themselves. The performance of the better hidden state is reported throughout. Performance for both supervised and unsupervised models was measured using the AUROC and AUPRC across all 101 subjects after leave-one-subject-out cross-validation (Tables 2 and 3).

Table 1. Summary of all computed physiologic indices used to infer nociceptive state

Modality	Index	Definition
HRV	Mean HR	Mean heart rate
HRV	Sigma HR	SD of heart rate
HRV	Total power	Total power across low, high, and very low frequency bands
HRV	HF	High frequency power (0.15 to 0.4 Hz)
HRV	LF_η	Normalized fraction of total power that is low frequency (LF) power (0.04 to 0.15 Hz)
HRV	HF_η	Normalized fraction of total power that is high frequency power
HRV	LF_η (no VLF)	Normalized fraction of low frequency power out of the sum of low and high frequency power (without very low frequencies)
HRV	LF/HF	Low frequency to high frequency ratio (sympathovagal balance)
EDA	EDA tonic	Tonic component of EDA
EDA	Mean PA	Mean EDA pulse amplitude
EDA	Sigma PA	SD of EDA pulse amplitude
EDA	Mean PR	Mean EDA pulse rate
EDA	Sigma PR	SD of EDA pulse rate

Supervised Models Outperform ANI Monitor. The random classifier expectations were 0.50 for AUROC and 0.22 for AUPRC (Table 2). First, the ANI index performed slightly better than a random classifier (Table 2). The mean AUROC and AUPRC across subjects were 0.53 and 0.25, respectively. These results likely reflect the fact that the ANI was validated using data quite different from ours and that the algorithm is real-time.

For the first set of our multisensor models, those relying only on physiological indices without any drug information, the overall model performance is markedly better than the ANI in terms of both AUROC and AUPRC. They were distinctly better than the random classifier expectation, with a mean AUROC and AUPRC of 0.62 and 0.31 for the logistic regression models, respectively, and a mean AUROC and AUPRC of 0.67 and 0.37 for the random forest models, respectively (Table 2). The performance of the nonparametric model (random forest) exceeds that of the parametric model (logistic regression) (Table 3). Adding drug information to both the logistic regression and random forest models improves classification performance across the board (Table 3). The mean AUROC and AUPRC of the logistic regression models increased to 0.67 and 0.34, respectively. The mean AUROC and AUPRC of the random forest models increased to 0.73 and 0.42, respectively.

Unsupervised Model Nearly Matches Supervised Performance. For the unsupervised model formulation, the performance of the best hidden state from the state space models as a marker of nociception is comparable to the performance of the multisensor

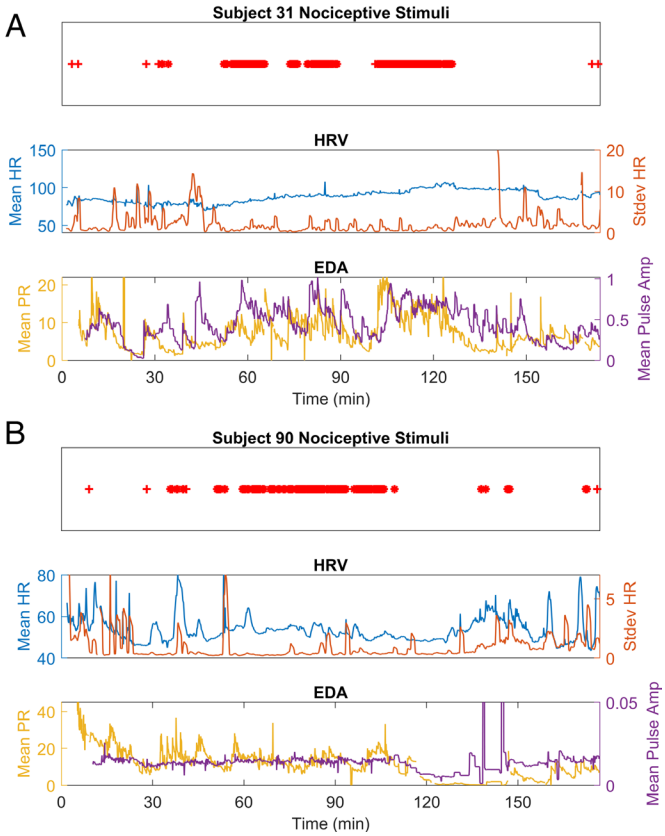

Fig. 2. Two example subjects (*(A)* Subject 31 and (*B*) Subject 90) along with some of the physiological features for both HRV and EDA and the ANI index for each. In each panel, the *top* row contains the information about the timing of nociceptive stimuli, depicted in red.

logistic regression models, which were trained on information about the timing of nociceptive stimulation (Table 2). The mean AUROC and AUPRC of the best hidden state across the dataset were 0.59 and 0.28, respectively.

Model Predictions Visually Track Nociceptive Stimuli. Fig. 3 *A* and *B* provides an intuitive visual representation of the results in Tables 2 and 3. Fig. 4 shows all of the model predictions throughout the surgery for the same two example subjects whose data are shown in Fig. 2. In contrast to the indices in Fig. 2 alone, our model predictions in Fig. 4 clearly track with nociceptive stimuli even by eye. This suggests that the models characterize the signature of nociception better than the raw indices themselves.

Learned Signatures of Nociception Align with Known Physiology. Table 4 summarizes the nociceptive signature learned by the all of the models, in terms of 95% CI of coefficients of the models across cross-validated runs for logistic regression and state space models, and the computed feature importance for random forest. This information is also further summarized using bold to denote significant coefficients and feature importance. Significant coefficients were defined as those for which the entire 95% CI was at least 0.05 away from 0 for logistic regression (classification) and state space models. Significant feature importance was defined as the entire 95% CI greater than 1e-5. The signs of the coefficients for each hidden state were remarkably consistent across all rounds of leave-one-out cross-validation, so there was no need to shuffle the labeling of hidden states by subject to match similar states (the assignment of State "1" and State "2" in a state space model can be arbitrary). For supervised models, the coefficient signs and feature importance for the derivatives of each index are also included.

Table 2. A summary of the AUROC and AUPRC of all of the models constructed using physiological indices alone, supervised and unsupervised, to classify time windows when nociceptive stimuli occurred, without including any information about drug dosing or timing

	AUROC	AUPRC
Random classifier	0.50	0.22
ANI index	[0.51, 0.55]	[0.23, 0.27]
Multisensor (Logistic regression)	[0.60, 0.64]	[0.28, 0.33]
Best Hidden State (State space)	[0.56, 0.61]	[0.25, 0.30]
Multi-sensor (Random forest)	[0.64, 0.69]	[0.34, 0.40]

Performance is reported as the 95% CI of the mean AUROC and AUPRC across the full dataset after leave-one-subject-out cross-validation. AUROC and AUPRC were computed based on the model's ability to classify time windows with and without nociceptive stimulation during surgery.

Additional Results. The supplementary information contains additional analyses and results. *SI Appendix*, Fig. S1 shows the results of replicating one of the evaluation methodologies from a previous ANI study on our ANI data, aggregated across the entire dataset. *SI Appendix*, Fig. S2 shows the distribution of model performance across the dataset. *SI Appendix*, Tables S1 and S2 summarize hyperparameters used during preprocessing. *SI Appendix*, Figs S3–S7 summarize the breakdown of model performance by demographic factors. It is worth noting that in our dataset, there is a strong correlation between age, gender, and surgery duration because the men and women had different surgeries which each had their own age and duration distributions. At this point, there is no convincing evidence of a correlation between model performance and any of the demographic factors examined. *SI Appendix*, Fig. S8 shows the boxplots of each feature including ANI compared in timepoints with and without nociceptive stimuli. It is clear from *SI Appendix*, Fig. S8 that no single feature (including ANI) is clearly different when there is a nociceptive stimulus. Even if there were single feature differences, there can be additional information and value in aggregating information across multiple indices. *SI Appendix*, Table S3 summarizes the performance of the random forest model compared to ANI when varying the window length from 2 to 15 s. The model performance is very consistent regardless of window length, which suggests that there is not significant information loss occurring due to compression into windows.

Discussion

In this study, we used data collected in the operating room during surgery to measure surgical nociception and derive models to predict it using noninvasive physiological indices which incorporate

Table 3. A summary of the AUROC and AUPRC of the supervised models, allowing for comparison between the types of classification models and the effect of including drug information (*Materials and Methods* for how drug information was encoded)

	AUROC	AUPRC	AUROC	AUPRC
	Logistic regression		Random forest	
Multisensor without drug info	[0.60, 0.64]	[0.28, 0.33]	[0.64, 0.69]	[0.34, 0.40]
Multisensor with drug info	[0.65, 0.68]	[0.31, 0.36]	[0.71, 0.75]	[0.39, 0.45]

Performance is reported as the 95% CI of the mean AUROC and AUPRC across the full dataset after leave-one-subject-out cross-validation. AUROC and AUPRC were computed based on the model's ability to classify time windows with and without nociceptive stimulation during surgery.

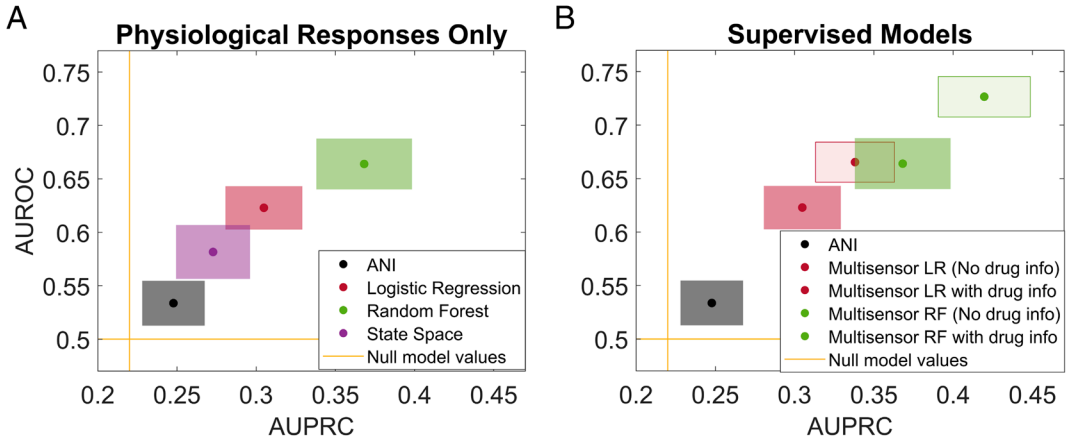

Fig. 3. A visual summary of the performance of all the models. (*A*) A summary of the performance of all of the constructed from physiological responses alone, (*B*) A summary of the performance of the supervised models for each of the feature sets. In each plot with x- and y-axes defined by AUPRC and AUROC, respectively, the horizontal and vertical lines that cut across the whole plot mark the random classifier expected AUROC and AUPRC. Each model is represented by a single color-coded rectangle, centered at the two-dimensional mean of the model performance and with dimensions given by the 95% CI in each dimension. LR = logistic regression, RF = random forest.

appropriate inductive biases. We collected a multisensor dataset in the operating room during surgery, including physiologic time series and manual annotations of nociceptive stimulation and drug dosing.

We report three principal findings from our work. First, the performance of our models shows that we can track surgical nociception accurately during surgery using multisensor ANS responses combined with statistically and physiologically rigorous computational models to extract the relevant information from the raw data (Fig. 3 and Tables 2 and 3). Including information about drug dosing and timing improves model performance as expected (Table 3 and Fig. 3*B*). It is important to note that when specific timepoints are examined and aggregated across all the subjects (*SI Appendix*, Fig. S1), the ANI performs as reported. However, this is not how the metric is used clinically. Significant results in group aggregate may not translate to individual or timepoint-by-timepoint validity. We specifically measured performance in terms of ability to track nociception throughout each individual surgery (for each subject) for this reason.

Our second principal finding is that for both supervised and unsupervised models, the nociceptive signature learned by the model was largely physiologically consistent with known responses to nociception, especially with respect to EDA responses (3, 4). According to Table 4, the nociceptive signature consists of increasing or high tonic EDA, EDA pulse amplitude, and EDA pulse rate. Even though the coefficients were smaller in magnitude for more of the HRV indices, the logistic regression signature also included increasing heart rate. The EDA indices are overall higher in importance than the HRV indices (Table 4), but this is not surprising given that cardiovascular drugs directly affect HRV compared to peripheral sweating activity. This could also explain some of the ambiguous HRV coefficients, such as those for mean heart rate in the logistic regression. Therefore, the EDA indices were likely more reliable as indicators of underlying nociception even when certain drugs that affected HRV (e.g., beta blockers) were given. With the exception of EDA pulse rate variability, the definitions of the EDA nociceptive signature from the classification and state space models agree entirely and are consistent with physiology (Table 4). The implicitly defined nociceptive state captured high tonic EDA, pulse rate, and pulse amplitude even without information about the timing of nociceptive stimuli (Table 4). This suggests that our models incorporated the correct inductive biases which allowed them to learn to identify a true nociceptive state, rather than a confound in the data.

Our third principal finding is that there is an underlying nociceptive state that can be discerned from the physiologic responses even without knowledge of when the nociceptive stimulus is having its effect in training. Based on the results of the state space modeling, there is the potential to define an implicit 'nociceptive

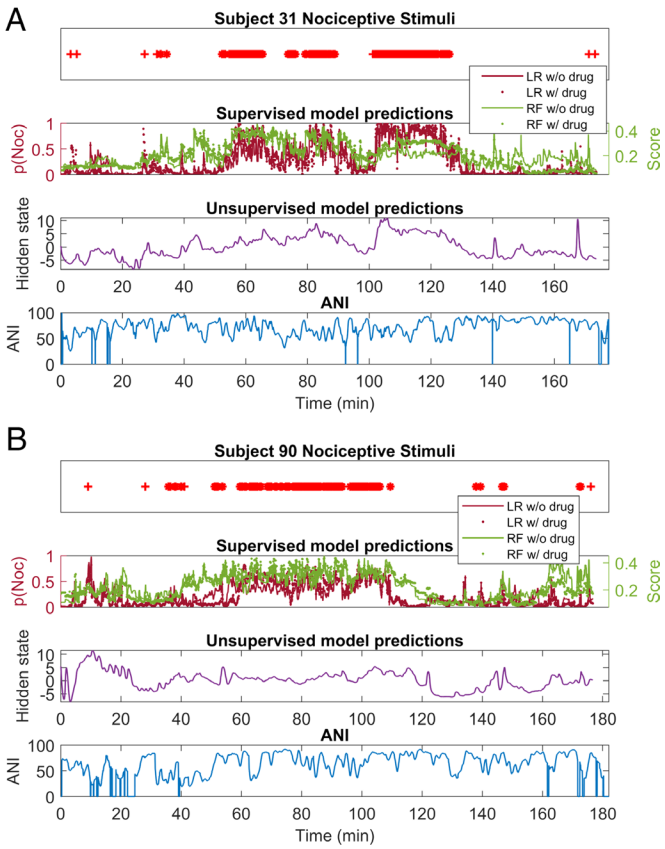

Fig. 4. The same two example subjects as in Fig. 2 ((*A*) Subject 31 and (*B*) Subject 90) with some of the model predictions along with the annotations of nociceptive stimuli. For each subject, the *top* row contains information about the timing of nociceptive stimuli, depicted in red. The *second* row for each subject shows our multisensor supervised model predictions with and without drug information in addition to the physiological information. LR stands for logistic regression, and RF for random forest. The *third* row for each subject shows the unsupervised model predictions. The *bottom* row for each subject shows the ANI index. p(Noc) = estimated probability of nociception.

Table 4. A summary of the fitted coefficients or feature importance of the various models, shown as the 95% CI of each coefficient or feature importance across leave-one-subject-out cross-validation

		Logistic regression coefficients	Random forest feature importance	State space model coefficients
Indices	**Mean HR**	**[−0.34, −0.34]**	**[8.0e-5, 8.2e-5]**	[0.02, 0.02]
	Sigma HR	**[0.11, 0.11]**	[6.2e-6, 6.5e-6]	**[0.13, 0.13]**
	Total power	[−0.04, −0.04]	[8.5e-6, 8.7e-6]	**[0.07, 0.07]**
	HF	[−0.02, −0.02]	[1.2e-5, 1.3e-5]	**[0.07, 0.08]**
	LF_η	[−0.05, −0.05]	[7.8e-7, 8.3e-7]	[0.02, 0.02]
	HF_η	**[0.07, 0.07]**	[6.8e-6, 7. 2e-6]	[0.03, 0.03]
	LF/HF	[0.01, 0.01]	[3.3e-6, 3.5e-6]	[−0.002, −0.002]
	EDA tonic	**[0.11, 0.11]**	**[1.1e-4, 1.1e-4]**	**[0.20, 0.20]**
	Mean EDA pulse amplitude	**[0.15, 0.16]**	**[1.0e-4, 1.1e-4]**	**[0.08, 0.08]**
	Sigma EDA pulse amplitude	**[0.07, 0.07]**	**[5.1e-5, 5.3e-5]**	**[0.09, 0.09]**
	LogN mean pulse rate	**[0.19, 0.19]**	**[2.9e-5, 3.0e-5]**	**[0.12, 0.12]**
	LogN sigma pulse rate	**[−0.19, −0.19]**	**[1.1e-5, 1.2e-5]**	**[0.08, 0.08]**
	IG mean pulse rate	**[0.09, 0.09]**	**[3.9e-5, 4.0e-5]**	**[0.14, 0.14]**
	IG sigma pulse rate	[0.02, 0.02]	**[1.4e-5, 1.5e-5]**	**[0.13, 0.13]**
Numerical derivatives	**Mean HR**	[0.04, 0.04]	[8.2e-7, 8.9e-7]	
	Sigma HR	[0.05, 0.05]	[1.1e-6, 1.2e-6]	
	Total power	[0.05, 0.05]	[1.9e-6, 2.0e-6]	
	HF	[0.003, 0.003]	[1.1e-6, 1.1e-6]	
	LF_η	[0.02, 0.02]	[2.2e-7, 2.4e-7]	
	HF_η	[0.05, 0.05]	[5.6e-7, 6.0e-7]	
	LF/HF	[0.01, 0.01]	[3.3e-7, 3.6e-7]	
	EDA tonic	[0.01, 0.01]	**[1.0e-5, 1.1e-5]**	
	Mean EDA pulse amplitude	**[0.07, 0.07]**	**[2.5e-5, 2.5e-5]**	
	Sigma EDA pulse amplitude	[0.03, 0.03]	**[1.2e-5, 1.2e-5]**	
	LogN mean pulse rate	**[0.06, 0.06]**	[1.7e-6, 1.8e-6]	
	LogN sigma pulse rate	[−0.01, −0.01]	[7.8e-7, 8.3e-7]	
	IG mean pulse rate	**[0.08, 0.08]**	[8.8e-6, 9.2e-6]	
	IG sigma pulse rate	[−0.02, −0.02]	[2.1e-6, 2.2e-6]	

state' that is physiologically consistent and can track nociception dynamically and with comparable performance to supervised models, without requiring information about the timing of nociceptive stimulation. A state space framework using multisensor physiological observations is effective in uncovering this implicit nociceptive state with a consistent definition across multiple subjects. This is an important step toward defining a metric to track nociception without including nociceptive "ground truth" information, most practical for scalability and implementation in clinical settings. This hidden state definition is a precursor to a future metric to track surgical nociception noninvasively in real-time for a closed-loop system to control surgical nociception.

This study has some limitations. The most significant limitation is that we do not have a known ground truth for when nociception is experienced. Therefore, we use a superset of ground truth, which is the collection of all nociceptive stimuli during surgery. It is true that every nociceptive stimulus will not result in nociception, especially at times when adequate antinociceptive medication has been given. However, in this study, we observed (and the models corroborated) many instances in which clear nociceptive responses were occurring but passing undetected by the ANI. We hypothesize that the better performance of our models was due to their increased ability to detect these true nociceptive responses. If our models had been detecting spurious responses not tied to nociception, the timing of these "false positives" would have been more randomly distributed, and the model signature associated with nociception would not have been as physiologically consistent as it was. A second limitation of our study is that while our models can detect nociceptive responses associated with inadequate pain control, they cannot detect responses associated with overdosing of pain medication during surgery. A third limitation is that our dataset consisted primarily of laparoscopic surgeries of the lower abdomen, which is not representative of all surgeries. In our future work, it is our intention to expand our data collection in the perioperative period and to different types of surgery to overcome both limitations. A fourth limitation is that our study involved fully post hoc data analysis, in which our algorithms benefit from having access to the full course of surgery. In contrast, the ANI is an online real-time algorithm. In our future work, we will investigate online implementations of our algorithms as well. This study serves as proof of concept that a signature of nociception exists within the data described.

This study also presented several atypical challenges. First, working within a clinical setting presents several logistical challenges since it is not an ideal controlled research setting. Heavy-duty surgical equipment caused interference with most wireless sensor capabilities and occasional limitations in signal quality which we attempted to resolve in preprocessing. The demands of surgery allowed limited access for adjustments and troubleshooting after surgery started, including when the patient is moved or positioned. Most importantly,

working within the requirements and needs of the clinical staff, which by necessity are the highest priority, required constant adaptation, sometimes limiting access, rushing study procedures such as instrumentation, or cutting off portions of study procedures altogether. Second, conducting an observational study in a clinical setting is drastically different from a controlled experimental setting. Each surgery had different anesthesia staff, additional surgical staff, and nursing, which leads to wide variation in anesthetic strategy and general workflow. The duration of surgery was also variable, even for the same surgery type. Many variables of each surgery were, for obvious reasons, not under our control.

Despite these challenges and limitations, the operating room is the only place to study surgical nociception. The nociceptive responses associated with surgical stimuli cannot be studied in humans in any other controlled experimental setting. In an observational study paradigm, we can overcome these challenges by carefully keeping track of all of the relevant potential confounders and covariates, despite the lack of a controlled setting, and then accounting for them during data analysis and in model frameworks. In our future work, as we expand our patient cohort, we will continue to investigate additional patient-specific factors that can influence surgical nociception, including not only demographic factors like age and gender but also comorbidities and outpatient medications. Our preliminary analysis of demographic factors within this dataset yielded no significant or consistent correlations (*SI Appendix*, Figs. S3–S7). In addition, because we have such detailed annotations of the nociceptive stimuli, we can eventually compare the magnitude and latency of the physiologic response to different types and locations of nociceptive stimuli. This can contribute to a future definition of "intensity" for nociceptive stimuli.

Our study is an important step toward developing objective markers to track surgical nociception. These markers will enable objective assessment of nociception in other complex clinical settings, such as the ICU, as well as catalyze future development of closed-loop control systems for nociception.

Materials and Methods

Data and code to replicate this study are available on Physionet (https://www.physionet.org) (43, 44).

Data Collection. We collected data from 101 subjects during surgery under a study protocol approved by the Massachusetts General Hospital (MGH) Human Research Committee (Protocol # 2017P002591). The Human Research Committee is the Institutional Review Board for MGH. Subjects were between the ages of 28 and 77 and included 53 females. The majority of procedures were laparoscopic, consisting mainly of prostatectomies, hysterectomies, and Salpingo-oophorectomies. After obtaining informed consent from subjects the morning of their surgery, we instrumented them with several sensors: 3-lead electrocardiogram (ECG) from the chest, electrodermal activity (EDA) from two digits of the left hand, pulse plethysmography (PPG) from one digit of either the right or left hand, respirations from an elastic strap around the chest, skin temperature from the surface of the hand, and the Analgesic Nociception Index (ANI) sensors on the chest. All except for the ANI were obtained using the Thought Technology Neurofeedback System (45). Data collection was started prior to induction of anesthesia and continued until after extubation. All sensor data were sampled at 256 Hz, except for the ANI, which was sampled around 1 Hz.

During the course of the surgery, nociceptive stimuli, such as intubation, incision, insufflation, electrocautery, desufflation, and extubation were manually annotated. The timing and dosage of anesthetic medication delivered was also manually annotated and verified using the electronic health record. Anesthesiologists were told to follow what they would normally do with no interference due to the study. Surgeries were chosen to be laparoscopic lower abdominal surgeries, so as to not interfere with the placement of sensors on the chest and to allow for precise annotation of nociceptive stimuli using the video monitors throughout the operating room. Participation in this study in no way changed routine surgical or anesthetic care.

HRV Preprocessing. For ECG data, the goal of preprocessing is to accurately extract R peaks, because HRV measures are highly sensitive to errors in even a single beat (46). The extraction of R peaks from clean ECG has been accomplished robustly and efficiently already using the Pan Tompkins algorithm (47, 48). The main challenge in data collected from the operating room is the presence of artifact that interferes with the ability to cleanly extract R peaks. Even when there is lesser artifact, ectopic beats or temporary motion artifact can be misidentified by the Pan Tompkins algorithm as an R peak. Therefore, methods must be employed to check and correct for incorrectly identified R peaks.

Stage 1: After a 20 Hz low-pass filter, we identified large sections of artifact by thresholding the derivative of the ECG where the absolute value of the derivative was at least 10 SD from the mean. We isolated the 0.5 s before and after each such instance and set the ECG signal to zero. In addition, we measured the maximum and minimum values of the ECG signal for each minute and used 1.25 times the median of each distribution as a threshold for the bounds of the signal. Any portion of the ECG that was outside these bounds was also set to zero. We kept a record of all such "zeroed" time points. The Pan Tompkins algorithm was run after the identified portions of the ECG had been set to zero. It was important to zero out sections of large artifact before running the Pan Tompkins algorithm because the presence of such artifact can interfere with the identification of R peaks in neighboring artifact-free regions of ECG.

Stage 2: The use of the Pan Tompkins algorithm on ECG data with zeroed out sections led to the extraction of R peaks with gaps corresponding to those sections. Therefore, to fill in the gaps, we interpolated R peaks using the peaks extracted from the pulse plethysmography (PPG) signal. Peaks can be extracted from the PPG signal using any local maximum or peak finding method (such as findpeaks in Matlab). However, there is naturally a delay between the R peak in the ECG and the corresponding PPG peak to account for the distance between the heart and the finger. We estimated this delay by taking a 5-min segment of "clean" ECG (no artifact) and computing the median relative delay between ECG R peak and PPG peak for all of the beats within the 5-min segment. We then used the estimated relative delay as a backward shift for any segment of PPG peaks used to fill in a missing section of R peaks. This method of estimating the delay between signals will be improved upon in our future work using techniques such as generalized linear models (GLMs). This technique addressed broad sections of missing R peaks. For four of the surgeries, the extent of ECG artifact was such that the PPG peaks were used instead throughout the surgery.

Stage 3: To address the single beat discrepancies, Luca Citi and Riccardo Barbieri have previously developed an algorithm rooted in the same point process framework to identify and correct irregular or ectopic beats in RR-interval traces (46). We used this algorithm on the extracted R peaks after Stage 2 for the next stage of correction.

Stage 4: We identified any sections of lingering irregularity in the RR-interval trace that were at least three R peaks in length and searched through all of the redundant versions of signal we had (between beat-corrected and uncorrected versions of both the ECG and PPG peaks) for the optimal placement of each R peak in those sections.

For each subject's R peaks, the point process HRV model was fitted (34). The point process model has two hyperparameters: p, which is the model order for the autoregressive portion of the model, and w, which is the window length in seconds for the maximization of local likelihood. The values of p and w for each subject were chosen by screening across 6, 8, 10, and 12 for p and 60, 90, and 120 for w. The optimal values were chosen based on the combination of hyperparameter values that minimized the Kolmogorov–Smirnov (KS) distance (49) of the fit. The final hyperparameter values used for each subject are in *SI Appendix*, Table S1.

EDA Preprocessing. The methods we developed for identifying and removing electrocautery-related artifact from the EDA data have been detailed in refs. 50, 51. EDA is commonly divided into tonic and phasic components (52). We used the methods previously described (35, 36, 53, 54) to separate the phasic component and extract pulses from each subject's EDA data after preprocessing and removing artifact. (The tonic component for each subject was also retained for later use.) Then, we fit 4 dynamic models for each subject using the extracted EDA pulses.

We have previously demonstrated the point process structure in the intervals between EDA pulses and their amplitudes (35, 36). The 4 dynamic models are detailed below.

History-dependent inverse Gaussian model: We fit a history-dependent inverse Gaussian (HDIG) model first, analogous to that for HRV (54). The hyperparameter values screened were larger, however, since EDA pulses are more spaced. To maximize local likelihood, we screened p (autoregressive model order) between 1 and 3 and w (window length) between 180 and 900 seconds in increments of 60. This model estimates the mean and SD of pulse rate continuously across time.

History-dependent lognormal model: Second, we fit a history-dependent lognormal (HDL) model that was identical in structure to the HDIG model except that a lognormal density was used instead of the inverse Gaussian for the interpulse intervals. This directly follows from the results discussed in refs. 35, 53 which showed that both inverse Gaussian and lognormal models are appropriate fits for interpulse interval information in EDA data. The same hyperparameter values were used for screening. This model provides another estimate of mean and SD of pulse rate continuously across time.

History-dependent inverse Gaussian model with adaptive window length: One drawback of the HDIG model is that it does not return estimates for the first window length w of the data. For example, if w is 600, then there are no estimates of mean and SD of pulse rate for the first 600 s of the data.

To work around this, we built a modified version of the HDIG model with an adaptive window length. The window length is constrained in length by needing to contain at least $P + 2$ pulses within it at all times, where p is the autoregressive model order. Therefore, for a fixed value of w, the sections of data with the sparsest pulses determine the value of w, even if other sections of data have a high density of pulses. We built a HDIG model in which the value of w is adjustable dynamically, such that w becomes larger specifically in regions with sparse pulses and reduces in length when pulses are more frequent. Then, we set the minimum length of w and minimized the lost time at the beginning in which there are no model estimates. Because of the loss in model fit, we only used the adaptive model estimates for the beginning portion of each subject's EDA data, until the original HDIG model produced estimates. We screened the minimum window length w of the adaptive model to values up to 300 s only (60, 120, 180, 240, 300). Therefore, we could guarantee the maximum of lost time (for EDA indices) to the first 5 min only.

History-dependent inverse Gaussian model for pulse amplitudes: Finally, we fit a history-dependent inverse Gaussian model for the pulse amplitudes. This directly follows from the results discussed in ref. 35 which showed that the inverse Gaussian is acceptable simplification for the amplitude structure in EDA. The hyperparameter values screened were the same as the HDIG and HDL models for temporal information because the spacing of pulses was the same. This model estimates the mean and SD of pulse amplitude continuously across time. The final threshold for extracting pulses as well as the final hyperparameter values for each of the 4 dynamic models for all 101 subjects are in *SI Appendix*, Table S2.

ANI Preprocessing. From the ANI monitor, the ANI index over time for each surgery was extracted. It is provided directly by the output of the monitor at roughly 1 Hz frequency. ANI data were either unable to be collected or not of sufficient quality to be included (at least 10 min of high-quality data across the full surgery) for 13 surgeries. For all surgeries for which ANI was included, ANI was affixed according to instructions provided with device and signal quality was verified at the beginning of surgery. However, in some cases, signal quality declined or wavered during the course of surgery, when we could no longer access the electrodes on the patient to perform manual adjustments, which is a challenge of the ANI.

Annotation Preprocessing. Nociceptive annotations: For nociceptive annotations, first, we split them into cautery and noncautery annotations. For noncautery annotations (e.g., incision, intubation, insufflation, desufflation, and extubation), we assumed a 15-s window of uncertainty on either side of the annotation to account for the duration of the event itself and the variability in when it was documented. For cautery annotations, we annotated the start and stop of each event. To account for any mistake in annotation, we added 5 s after the annotated end of each instance of cautery.

Annotations about drug timing and dosing: For drug-related annotations, we were concerned with 8 classes of drugs typically given throughout surgery: sedative, antinociceptive, muscle relaxant, pressor, beta blocker, alpha agonist, NMDA antagonist, and inhaled sedative. Then, for each annotation, we extracted information about whether it was a bolus or an infusion and the dose. We then normalized the dosing levels between 0 and 1, where 1 referred to the largest dose of that class of drug given throughout the surgery. In some cases where more than one drug is in a class and they are dosed at very different levels (e.g., fentanyl and hydromorphone), we normalized each drug's dosing separately. Infusion dosing was typically normalized separately from bolus dosing since they are not always comparable. All boluses were divided across 15 s total to account for uncertainty in exact timing. Finally, we compiled two dynamic summaries of the drug information, one that only contained binary timing information of whether a drug was administered or onboard in each class and one that also contained normalized dosage information.

Modeling Framework. For each subject, the duration of their surgery was divided into 5-s nonoverlapping windows. For ground truth, each 5-s time window was classified as nociceptive or not based on the presence of a nociceptive annotation within that time window. Model performance was measured based on the ability to accurately classify these 5-s windows as nociceptive or not.

The median of all of the indices in Table 1 within each window was computed as part of the feature vector for each window (15 features). To account for the direction of change of an index as a possible feature, for each feature in Table 1, its 8th-order central finite difference derivative was also computed as a feature. Each feature in Table 1 was smoothed with a 12-point weighted linear least squares (with 1st-degree polynomial) moving filter before the derivative was computed so that the derivative captured overarching trends rather than temporary fluctuations. This accounted for an additional 15 features for each time window, for a total of 30.

When it was included, drug-related information consisted of timing information and dose information as described in the Drug Annotations and added an additional 14 features to each time window, raising the total to 44.

All features were Z-score normalized within each subject before being concatenated across subjects to account for different baseline values across subjects. For the dependent variable, each 5-s window was coded either 0 or 1 based on whether it contained a nociceptive stimulus. The annotations were adjusted for uncertainty in exact timing as described in Annotations Section.

To compare the ability of the ANI to detect nociception, the ANI index was also divided into 5-s time windows and the median within each time window was computed.

Metrics of Performance. To test the generalizability of the model, model performance was always evaluated using leave-one-subject-out cross-validation. This means that the model was always validated on an entirely new subject's data compared to the training data.

Model performance was measured using two metrics. The first is area under the receiver operating characteristic, or AUROC. However, because the data are unbalanced where only ~20% of time windows contain nociceptive stimuli, the second metric was the area under the precision–recall curve, or AUPRC. The AUPRC can be a more reliable measure of model performance as the data tend to be more imbalanced (42). The random classifier performance for AUROC is 0.5, whereas for AUPRC, it is equal to the proportion of the data that are positives (in this case, ~20%).

ANI Accuracy. We also tested the accuracy of classifying nociceptive time windows using the ANI index alone by computing the AUROC and AUPRC for the ANI index.

Classification Models. The two classification models used were logistic regression (39) and random forest (40). Each classification model was applied to the physiological features with and without drug information. For the logistic regression models, LASSO regularization was used to penalize the inclusion of redundant or uninformative features in the model and minimize overfitting (55). The optimal model was always selected by minimizing Akaike's Information Criterion (AIC) (56). For the random forest models, the random forest was bagged to minimize overfitting using the square root of the total number of features (40). Each

random forest had 100 trees, and each tree had a maximum of 50 splits. Ninety percent of the data were randomly sampled at each split to reduce overfitting.

Finally, with respect to interpretation of the models, for the logistic regression models, we examined the magnitude and sign of the coefficient for each feature across the final models. Since leave-one-out cross-validation results in one model per subject, we included the coefficients from all models and kept track of the strongest and most consistent trends across models. For the random forest approach, we scored the importance of each feature based on how many times it is selected for a split throughout the forest. We kept track of features that were consistently most important across all final models.

State Space Models. The state space framework assumes that the measurable observations in the system are driven by a dynamically evolving latent or hidden state that cannot be directly observed (41). We treated the physiological indices as measurable observations governed by a latent nociceptive state that evolved over the course of surgery for each subject. We did not include any information about nociception, such as the nociception annotations, in the framework of the model, making it fully unsupervised. This is in contrast to the classification models, which were trained with the ground truth information about when nociceptive stimulation occurred. Our goal was to test the hypothesis that an implicit definition of nociceptive state can be learned from the physiological observations alone, without explicitly defining nociception. If the state space framework yields a hidden state that tracks with nociception, this would constitute an implicit nociceptive state.

Our state space framework was defined as follows (as a linear Gaussian state space model):

State evolution equation:

$$X_k = FX_{k-1} + v_k, \quad v_k \in N(O, Q).$$

Observation equation:

$$Y_k = A + BX_k + \varepsilon_k, \qquad \varepsilon_k \in N(0, R).$$

Here, X_k is the hidden state at time k, and Y_k is vector of observations at time k. $\{F, Q, A, B, R\}$ are the model parameters. Both the hidden state and observations are assumed to have Gaussian noise. The observations that made up the Y vector were the indices from Table 1, the 8 HRV indices and 7 EDA indices, Z-score normalized for each subject in 5-s windows. We defined a state space model with all 15 indices as observations and a two-dimensional hidden state. This was to account for different sets of governing dynamics, as we know the ANS has two arms: the sympathetic and parasympathetic.

The overall process of fitting a linear Gaussian state space model using the expectation–maximization (EM) algorithm is covered in ref. 41. We used the SPXEM package to fit linear state space models (57). At each iteration of the EM algorithm, the E-step (expectation) computes the expected value of the hidden state given the current estimates of the parameters $\{F, Q, A, B, R\}$. Then, the M-step (maximization) updates the estimates of the parameters by maximizing the joint likelihood of the observations and the hidden state (41). We defined a leave-one-subject-out cross-validation scheme that would allow the model to train on 100 subjects' data and be truly blind to the test subject's data. This consisted of several steps:

1. First, we ran the EM algorithm on each individual subject's data backward (58), starting at the last set of observations running to the first set of observations. We saved the value of the last backward hidden state as the initial value of the forward hidden state for each subject.
2. We fitted a single-state space model to 100 of the 101 subjects by concatenating their data together. However, during each iteration of the Kalman filter, when a new subject's data were reached, the saved initial value of the hidden state for that subject was used to reinitialize that subject. These measures were employed to avoid false assumption of continuity when data were concatenated across subjects.
3. Finally, for the held-out test subject, we used their saved initial value of the hidden state along with the model parameters from the model fitted on the training data and just performed one iteration of the E-step of the EM algorithm to arrive at the estimated trajectory of the hidden state for that subject. This allowed the model fitting process to be truly blind to the test subject since each subject's state initialization could be done using that subject's data alone.

By treating each hidden state as a dynamic prediction of nociception, we tested the ability of each of the two hidden states to classify nociceptive timepoints. We used the state space model coefficients to interpret the model.

Data, Materials, and Software Availability. Anonymized Features from physiological signals data have been deposited in Physionet (https://doi.org/10.13026/gzmm-5h49) (43).

ACKNOWLEDGMENTS. We would like to thank the Department of Anesthesia at Massachusetts General Hospital and the many anesthesiologists who helped with our study.

Author affiliations: [a]Harvard-Massachusetts Institute of Technology Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Picower Institute of Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Massachusetts General Hospital, Boston, MA 02114; and [d]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy 20133

Author contributions: S.S., R.B., and E.N.B. designed research; S.S., B.T., M.D.C., A.G., and D.M.D. performed research; S.S., R.B., and E.N.B. contributed new reagents/analytic tools; S.S. analyzed data; and S.S., R.B., and E.N.B. wrote the paper.

1. E. N. Brown, R. Lydic, N. D. Schiff, General anesthesia, sleep, and coma. *New England J. Med.* **363**, 2638–2650 (2010).
2. E. N. Brown, K. J. Pavone, M. Naranjo, Multimodal general anesthesia. *Anesth. Analg.* **127**, 1246–1258 (2018).
3. J. P. Desborough, The stress response to trauma and surgery. *Br. J. Anaesth.* **85**, 109–117 (2000).
4. E. N. Brown, L. A. Santa Cruz, S. Subramanian, Multimodal general anesthesia in practice. *Anesth. Analg.* **127**, 1246–1259 (2018).
5. H. Kehlet, J. B. Dahl, The value of "multimodal" or "balanced analgesia" in postoperative pain treatment. *Anesth. Analg.* **77**, 1048–1056 (1993).
6. L. A. Santa Cruz Mercado, R. Liu, K. M. Bharadwaj *et al.*, Association of intraoperative opioid administration with postoperative pain and opioid use. *JAMA Surg.* **158**, 854–864 (2023), 10.1001/jamasurg.2023.2009.
7. M. Haefeli, A. Elfering Pain assessment. *Eur Spine J.* **Suppl 1**, S17–S24 (2006) 10.1007/s00586-005-1044-x.
8. N. S. Gregory *et al.* An overview of animal models of pain: Disease models and outcome measures. *J Pain.* **14**, 1255–1269 (2013), 10.1016/j.jpain.2013.06.008.
9. E. N. Brown, P. L. Purdon, C. J. Van Dort. General anesthesia and altered states of arousal: a systems neuroscience analysis. *Annu Rev Neurosci*, **34**, 601–628, 2011.
10. R. Logier, M. Jeanne, B. Tavernier, J. De jonckheere "Pain/analgesia evaluation using heart rate variability analysis" in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (2006) pp. 4303–4306.
11. N. Ben-Israel, M. Kliger, G. Zuckerman, Y. Katz, R. Edry, Monitoring the nociception level: A multi-parameter approach. *J. Clin. Monit. Comput.* **27**, 659–668 (2013).
12. D. T. J. Liley *et al.*, Propofol and remifentanil differentially modulate frontal electroencephalographic activity. *Anesthesiology* **113**, 292–304 (2010).
13. M. Rantanen *et al.*, Novel multiparameter approach for measurement of nociception at skin incision during general anaesthesia. *Br. J. Anaesth.* **96**, 367–376 (2006).
14. J. L. Rhudy, C. R. France, E. J. Bartley, K. M. McCabe, A. E. Williams, Psychophysiological responses to pain: Further validation of the nociceptive flexion reflex (nfr) as a measure of nociception using multilevel modeling. *Psychophysiology* **46**, 939–948 (2009).
15. T. Ledowski *et al.*, Analgesia nociception index: Evaluation as a new parameter for acute postoperative pain. *Br. J. Anaesth.* **111**, 627–629 (2013).
16. M. Jeanne, R. Logier, J. De Jonckheere, B. Tavernier, "Validation of a graphic measurement of heart rate variability to assess analgesia/nociception balance during general anesthesia" in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, (2009), pp. 1840–1843.
17. M. Jeanne, C. Clément, J. De Jonckheere, R. Logier, B. Tavernier, Variations of the analgesia nociception index during general anaesthesia for laparoscopic abdominal surgery. *J. Clin. Monit. Comput.* **26**, 289–294 (2012).
18. M. Gruenewald *et al.*, Influence of nociceptive stimulation on analgesia nociception index (ANI) during propofol–remifentanil anaesthesia. *Br. J. Anaesth.* **110**, 1024–1030 (2013).
19. R. Treister, M. Kliger, G. Zuckerman, I. G. Aryeh, E. Eisenberg, Differentiating between heat pain intensities: The combined effect of multiple autonomic parameters. *Pain* **153**, 1807–1814 (2012).
20. H. Storm, Changes in skin conductance as a tool to monitor nociceptive stimulation and pain. *Curr. Opin. Anaesthesiol.* **21**, 796–804 (2008).
21. H. Storm *et al.*, Skin conductance correlates with perioperative stress. *Acta Anaesthesiol. Scand.* **46**, 887–895 (2002).
22. H. Storm, M. Shafiei, K. Myre, J. Raeder, Palmar skin conductance compared to a developed stress score and to noxious and awakening stimuli on patients in anaesthesia. *Acta Anaesthesiol. Scand.* **49**, 798–803 (2005).
23. T. Ledowski *et al.*, New parameters of skin conductance compared with bispectral index monitoring to assess emergence from total intravenous anaesthesia. *Br. J. Anaesth.* **99**, 547–551 (2007).
24. T. Ledowski, M. J. Paech, H. Storm, R. Jones, S. A. Schug, Skin conductance monitoring compared with bispectral index® monitoring to assess emergence from general anaesthesia using sevoflurane and remifentanil. *Br. J. Anaesth.* **97**, 187–191 (2006).

25. T. Ledowski *et al.*, The assessment of postoperative pain by monitoring skin conductance: Results of a prospective study. *Anaesthesia* **62**, 989–993 (2007).
26. T. Ledowski, B. Ang, T. Schmarbeck, J. Rhodes, Monitoring of sympathetic tone to assess postoperative pain: Skin conductance vs surgical stress index. *Anaesthesia* **64**, 727–731 (2009).
27. C. Ilies *et al.*, Evaluation of the surgical stress index during spinal and general anaesthesia. *Br. J. Anaesth.* **105**, 533–537 (2010).
28. M. Huiku *et al.*, Assessment of surgical stress during general anaesthesia. *Br. J. Anaesth.* **98**, 447–455 (2007).
29. D. Harrison *et al.*, Skin conductance as a measure of pain and stress in hospitalised infants. *Early Hum. Dev.* **82**, 603–608 (2006).
30. M. Gruenewald *et al.*, Influence of different remifentanil concentrations on the performance of the surgical stress index to detect a standardized painful stimulus during sevoflurane anaesthesia. *Br. J. Anaesth.* **103**, 586–593 (2009).
31. M. Eriksson, H. Storm, A. Fremming, J. Schollin, Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants. *Acta Paediatr.* **97**, 27–30 (2008).
32. V. Bonhomme *et al.*, Comparison of the surgical pleth indextm with haemodynamic variables to assess nociception–anti-nociception balance during general anaesthesia. *Br. J. Anaesth.* **106**, 101–111 (2011).
33. I. Bergmann *et al.*, Surgical pleth index-guided remifentanil administration reduces remifentanil and propofol consumption and shortens recovery times in outpatient anaesthesia. *Br. J. Anaesth.* **110**, 622–628 (2013).
34. R. Barbieri, E. C. Matten, A. A. Alabi, E. N. Brown, A point-process model of human heartbeat intervals: New definitions of heart rate and heart rate variability. *Am. J. Physiol. Heart Circ. Physiol.* **288**, H424–435 (2005).
35. S. Subramanian, R. Barbieri, E. N. Brown, Point process temporal structure characterizes electrodermal activity. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26422–26428 (2020).
36. S. Subramanian, P. L. Purdon, R. Barbieri, E. N. Brown, Elementary integrate-and-fire process underlies pulse amplitudes in electrodermal activity. *PLOS Comput. Biol.* **17**, e1009099 (2021).
37. D. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes* (Springer, New York, 2003).
38. D. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. (Springer, New York, 2007).
39. C. R. Shalizi, *Advanced Data Analysis from an Elementary Point of View, chapter 11* (Cambridge University Press, 2017), pp. 234–261.
40. L. Breiman, Random forests. *Machine Learning* **45**, 5–32 (2001).
41. R. H. Shumway, D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples, chapter 6* (Springer, New York, NY, ed. 3, 2011) pp. 319–404).
42. J. Davis, M. Goadrich "The relationship between precision-recall and ROC curves" in *Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh* (2006).
43. S. Subramanian, B. Tseng, R. Barbieri, E. Brown, Multimodal Physiological Indices During Surgery Under Anesthesia (version 1.0). *PhysioNet.*, (2024), https://doi.org/10.13026/gs4v-4q80. Deposited 23 August 2024.
44. A. Goldberger *et al.*, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).
45. Neurofeedback expert system. Thought Technology Ltd. https://thoughttechnology.com/neurofeedback-expert-system/. Accessed 1 October 2017.
46. L. Citi, E. N. Brown, R. Barbieri, A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats. *IEEE Trans. Biomed. Eng.* **59**, 2828–2837 (2012).
47. J. Pan, W. J. Tompkins, A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **32**, 230–236 (1985).
48. H. Sedghamiz. Matlab Implementation of Pan Tompkins ECG QRS detector, ResearchGate (2014). https://doi.org/10.13140/RG.2.2.14202.59841. Accessed 14 March 2018.
49. E. Brown, R. Barbieri, V. Ventura, R. Kass, L. Frank, The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.* **14**, 325–346 (2002).
50. S. Subramanian, B. Tseng, R. Barbieri, E. N. Brown, An unsupervised automated paradigm for artifact removal from electrodermal activity in an uncontrolled clinical setting. *Physiol. Meas.* **43**, 115005 (2022), 10.1088/1361-6579/ac92bd.
51. S. Subramanian, B. Tseng, R. Barbieri, E. N. Brown. "Unsupervised machine learning methods for artifact removal in electrodermal activity" in *Proc. 43rd IEEE International Conf on Eng in Biol and Med (EMBC)*, (2021). https://doi.org/10.1109/EMBC46164.2021.9630535
52. W. Boucsein, *Electrodermal Activity* (Springer, New York, NY, 2012).
53. S. Subramanian, P. L. Purdon, R. Barbieri, E. N. Brown, A Model-Based Approach for Pulse Selection from Electrodermal Activity. *IEEE Trans. Biomed. Eng.* **68**, 2833–2845 (2021), 10.1109/TBME.2021.3071366.
54. S. Subramanian, P. L. Purdon, R. Barbieri, E. N. Brown, Quantitative assessment of the relationship between behavioral and autonomic dynamics during propofol-induced unconsciousness. *PLOS ONE* **16**, e0254053 (2021), 10.1371/journal.pone.0254053.
55. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B (Methodol.)* **58**, 267–288 (1996).
56. Y. Pawitan, *In All Likelihood* (Clarendon Press, Oxford, United Kingdom, 2013).
57. Sebastien Blais. EM algorithm for linear state-space models. *MATLAB Central File Exchange* (2017). (https://www.mathworks.com/matlabcentral/fileexchange/64585-em-algorithm-for-linear-state-space-models). Retrieved April 16, 2020.
58. A. C. Smith *et al.*, State-space algorithms for estimating spike rate functions. *Comput. Intell. Neurosci.* **2010**, 426539 (2010).