HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

DEPARTMENT OF BIOLOGICAL CHEMISTRY - MEDICAL SCHOOL

Stress Indices in Surgical Simulation

Konstantinos Georgiou, MD, MSc

PhD Thesis

Supervisors: Prof. Athanassios G. Papavassiliou
Prof. Georgios K. Zografos
Prof. Pantelis P. Karaiskos

Η παρούσα διδακτορική διατριβή εκπονήθηκε υπό τον ελληνικό τίτλο «Δείκτες stress στη χειρουργική προσομοίωση», στο Εργαστήριο Βιολογικής Χημείας και στο Εργαστήριο Ιατρικής Φυσικής της Ιατρικής Σχολής του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών.

ATHENS 2019

Related scientific work produced on the occasion of our Thesis

Publications in international peer-reviewed journals

1. **Konstantinos Georgiou**, Andreas Larentzakis, Athanasios G. Papavassiliou. ***"Surgeons' and Surgical Trainees' Acute Stress in Real Operations or Simulation: A Systematic Review".*** The Surgeon, Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland (SURGE), 2017;15(6):355-365. DOI: 10.1016/j.surge.2017.06.003.

2. **Konstantinos E. Georgiou**, Andreas V. Larentzakis, Nehal N. Khamis, Ghadah I. Alsuhaibani, Yasser A. Alaska, Elias J. Giallafos. ***"Can Wearable devices accurately measure heart rate variability? A systematic review".*** Folia Medica 2018 1;60(1):7-20. DOI: 10.2478/folmed-2018-0012.

3. **Georgiou KE**, Dimov RK, Boyanov NB, Zografos KG, Larentzakis AV, Marinov BI. ***"Feasibility of a new wearable device to estimate acute stress in novices during high-fidelity surgical simulation"***. Folia Medica 2019; 61(1): DOI: 10.2478/folmed-2019-0001.

4. **Konstantinos E. Georgiou**, Ninos Oussi, Dimitrios Thanasas, Lars Enochsson, Andreas Larentzakis, Athanasios G. Papavassiliou. ***"Assessing various noninvasive stress indices to predict novice surgeons' performance during basic skills training in a high-end simulator"***. Journal of the American College of Surgeons 2019, submitted after invitation.

Conference papers in peer-reviewed journals

1. **Georgiou K.**, Larentzakis A., Gialafos I. ***"Non-Invasive Acute Stress Estimation in Novices During High-Fidelity Surgical Simulation".*** Journal of the American College of Surgeons, October 2017, Volume 225, Issue 4, e156. http://www.journalacs.org/article/S107207515(17)31560-0/fulltext, DOI: 10.1016/j.jamcollsurg.2017.07.950,

Presented in the Scientific Forum program at the American College of Surgeons (ACS) 103rd Annual Clinical Congress 2017, San Diego Convention Center, October 22 – 26 2017, San Diego, California, USA.

2. Ninos Oussi, **Konstantinos Georgiou**, Nikolaos Michalopoulos, Andreas Larentzakis, Marcus Castegren, Lars Enochsson. ***"A Randomized Crossover Study In Novices to Evaluate Surrogate Variables for Stress Provoked by Laparoscopic Suture Training in a High Fidelity Simulator Using Different Laparoscopic Handles".*** Journal of the American College of Surgeons, October 2018, Volume 227, Issue 4, Supplement 2, Pages e209-e210, https://www.journalacs.org/article/S1072-7515%2818%2931860-X/fulltext, DOI: 10.1016/j.jamcollsurg.2018.08.567
Presented in the Scientific Forum program at the American College of Surgeons (ACS) 104th Annual Clinical Congress 2018, Boston Convention Center, October 21-25 2018, Boston, Massachusetts, USA.

3. **Georgiou K**, Zografos K, Thanassas D, Giallafos E, Larentzakis A. ***"Cardiac Mental stress related indices can be accurately assessed by a wrist wearable device during basic training laparoscopy skills simulation"***. Accepted oral presentation, E-Poster. International Conference on Healthcare Simulation 2018, October 22-23, 2018, London, UK. J Chem Appl Chem Eng 2018, Volume: 2, DOI: 10.4172/2576-3954-C2-006

4. Ninos Oussi, **Konstantinos Georgiou**, Andreas Larentzakis, Dimitrios Thanasis, Lars Enochsson. ***"Validation of a Novel Needle Holder to Train Advanced Laparoscopy Skills to Novices in a Simulation Environment".*** Scientific Program in Poster Rounds Group B, at the 2019 American College of Surgeons (ACS) Surgical Simulation Summit: An International Multi-Professional Meeting, March 14-16 2019, Swissotel, Chicago, Illinois, USA. https://www.facs.org/~/media/files/education/aei/2019_acs_aei_scientific_abstracts.ashx

5. **Konstantinos E. Georgiou**, Ninos Oussi, Dimitrios Thanasas, Lars Enochsson, Andreas Larentzakis, Athanasios G. Papavassiliou, ***"Assessing various***

noninvasive stress indices to predict novice surgeons' performance during basic skills training in a high-end simulator". *Accepted for oral presentation at the Scientific Forum program of the American College of Surgeons (ACS) 105th Annual Clinical Congress 2019, to be held in October 27-31 2019, San Francisco, California, USA.*

Announcements at international peer-reviewed conferences

1. E-Poster Oral Presentation entitled: **Georgiou K.**, Larentzakis A., Gialafos I. ***"Non-Invasive Acute Stress Estimation in Novices During High-Fidelity Surgical Simulation".*** Journal of the American College of Surgeons, October 2017, Volume 225, Issue 4, e156. http://www.journalacs.org/article/S107207515(17)31560-0/fulltext, DOI: 10.1016/j.jamcollsurg.2017.07.950,
   Presenter of the E-Poster presented in the Scientific Forum program at the American College of Surgeons (ACS) 103rd Annual Clinical Congress 2017, San Diego Convention Center, October 22 – 26 2017, San Diego, California, USA.

2. **Konstantinos Georgiou**. **"Non-Invasive Acute Stress Estimation in Novices during High-Fidelity Surgical Simulation**". Lecture presented at the 1st National Conference on Simulation Training in Medicine, Medical Simulation Training Center (MSTC), Medical University of Plovdiv, 14-16 December 2017, Plovdiv, Bulgaria.

3. **Georgiou K**, Zografos K, Thanassas D, Giallafos E, Larentzakis A. ***"Cardiac Mental stress related indices can be accurately assessed by a wrist wearable device during basic training laparoscopy skills simulation"***. Oral presentation, E-Poster. International Conference on Healthcare Simulation 2018, October 22-23, 2018, London, UK. J Chem Appl Chem Eng 2018, Volume: 2. DOI: 10.4172/2576-3954-C2-006

4. Ninos Oussi, **Konstantinos Georgiou**, Andreas Larentzakis, Dimitrios Thanasis, Lars Enochsson. ***"Validation of a Novel Needle Holder to Train Advanced Laparoscopy Skills to Novices in a Simulation Environment".*** Scientific Program in Poster Rounds Group B, at the 2019 American College of Surgeons (ACS) Surgical Simulation Summit: An International Multi-Professional Meeting,

March 14-16, 2019, Swissotel, Chicago, Illinois, USA.
https://www.facs.org/~/media/files/education/aei/2019_acs_aei_scientific_abstracts.ashx

5. Ninos Oussi, **Konstantinos Georgiou**, Nikolaos Michalopoulos, Andreas Larentzakis, Marcus Castegren, Lars Enochsson. ***"A Randomized Crossover Study In Novices to Evaluate Surrogate Variables for Stress Provoked by Laparoscopic Suture Training in a High Fidelity Simulator Using Different Laparoscopic Handles".*** Journal of the American College of Surgeons, October 2018, Volume 227, Issue 4, Supplement 2, Pages e209-e210, https://www.journalacs.org/article/S1072-7515%2818%2931860-X/fulltext DOI: https://doi.org/10.1016/j.jamcollsurg.2018.08.567 .
E-poster presented in the Scientific Forum program at the American College of Surgeons (ACS) 104th Annual Clinical Congress 2018, Boston Convention Center, October 21-25, 2018, Boston, Massachusetts, USA.

6. **Konstantinos E. Georgiou**, Ninos Oussi, Dimitrios Thanasas, Lars Enochsson, Andreas Larentzakis, Athanasios G. Papavassiliou. ***"Assessing various noninvasive stress indices to predict novice surgeons' performance during basic skills training in a high-end simulator"***. *Accepted for oral presentation in the Scientific Forum program of the American College of Surgeons (ACS) 105th Annual Clinical Congress 2019, to be held in October 27-31 2019, San Francisco, California, USA.*

Ανακοινώσεις σε Ελληνικά Συνέδρια

1. **Κωνσταντίνος Γεωργίου**, Ηλίας Γιαλλάφος, Ανδρέας Λαρεντζάκης, Γεώργιος Ζωγράφος. **«*Αναίμακτη εκτίμηση του οξέος stress σε αρχαρίους κατά τη διάρκεια εκπαίδευσης σε υψηλής πιστότητας χειρουργικό προσομοιωτή*»**. Προφορική ανακοίνωση ΕΑ090, 31ο Πανελλήνιο Συνέδριο Χειρουργικής και Χειρουργικό Φόρουμ 2018 με διεθνή συμμετοχή. 21-24 Νοεμβρίου 2018, Ξενοδοχείο Hilton Αθηνών, Αθήνα.
https://www.execongress2018.gr/articlefiles/31st-Surgical-Final-Program-online-181122-1445.pdf

Table of Contents

1. Contents

Ευχαριστίες.....10

Abstract.....12

Περίληψη 14

Part A: Theoretical Background.....16

Stress: Definition and types 17

Indices measuring stress.....19

1. Stress assessment tests 19
2. Heart Rate Variability 21
3. Electrodermal activity (galvanic skin response) 23
4. Thermal activity.....24
5. Biochemical stress markers (saliva).....26

Stress in Surgery 38

Simulation in Surgery 40

Stress in simulation training.....44

Stress measuring wearable devices 47

Part B: Our Contribution 53

Aim…….....54

Material and Methods 54

Results…. 61

Discussion 91

Limitations.....100

Future Directions 101

Conclusion 102

References.....103

Appendix.....128

Ευχαριστίες

Καθ’ όλη τη διάρκεια εκπόνησης της παρούσας διατριβής είχα την ευκαιρία και την τύχη να γνωρίσω και να συνεργαστώ με πολλούς ανθρώπους, η συμβολή και η συμπαράσταση των οποίων τόσο σε επιστημονικό όσο και σε προσωπικό επίπεδο ήταν καθοριστικές.

Ευχαριστώ πολύ τον κύριο επιβλέποντα αυτής της διδακτορικής διατριβής καθηγητή κ. Αθανάσιο Παπαβασιλείου, τόσο για την ανάθεση του θέματος και την εμπιστοσύνη που έδειξε στο πρόσωπό μου από την πρώτη κιόλας στιγμή, όσο και τη μόνιμη καθοδήγηση που μου προσέφερε. Δεν αποτέλεσε απλώς την επιστημονική πυξίδα που πάντα συμβουλευόμουν, αλλά ανέλαβε και στο ακέραιο την ευθύνη της χάραξης, επίβλεψης και καθοδήγησης στην εκπόνηση της διατριβής αυτής.

Βαθιά ευγνωμοσύνη θα ήθελα να εκφράσω και στον Καθηγητή κ. Γιώργο Ζωγράφο ο οποίος μου επέτρεψε να επωφεληθώ από την γνωριμία ενός σπάνιου ταλέντου, γνώσεων και επιστημονικής επάρκειας ανθρώπου που μου δίνει τη δυνατότητα να διευρύνω τους ερευνητικούς και γνωσιακούς μου ορίζοντες. Η ενθάρρυνση, η συμβολή του και η επιστημονική του αρτιότητα συνέβαλαν τα μέγιστα στην εκπόνηση της παρούσας διατριβής.

Μία ιδιαίτερη και εκ βαθέων ευχαριστία οφείλω στον Καθηγητή κ. Παντελή Καραΐσκο, ο οποίος συνέβαλε αποφασιστικά στην προσπάθειά μου για ολοκλήρωση της παρούσας διατριβής και εξασφάλισε ένα άρτιο και δημιουργικό περιβάλλον, αναγκαία συνθήκη για την επιτυχή ολοκλήρωση μιας ερευνητικής προσπάθειας.

Την αναπληρώτρια καθηγήτρια κ. Χριστίνα Δάλλα καθώς και τον καθηγητή κ. Κωνσταντίνο Πάντο καθώς και το προσωπικό του Εργαστηρίου Φαρμακολογίας της Ιατρικής Σχολής Αθηνών θερμά ευχαριστώ για τις αναλύσεις των βιοχημικών δεικτών στο σίελο σε kits που ευγενικά μας προμήθευσε ο καθηγητής κ. Lars Enochsson.

Ιδιαίτερα θα ήθελα να εκφράσω την ευγνωμοσύνη μου προς τον κ. Αντρέα Λαρεντζάκη, και να τον ευχαριστήσω θερμά για τη συνεχή καθοδήγησή του, την άψογη συνεργασία μας και την αμέριστη συμπαράσταση και βοήθεια που μου προσέφερε, καθώς και τις πολύ ενδιαφέρουσες συζητήσεις μας. Η συνεισφορά του στην ερευνητική διαδικασία διαδραμάτισαν πολύ σημαντικό ρόλο στην εκπόνηση της παρούσας διδακτορικής διατριβής.

Ένα πολύ μεγάλο και ειλικρινές ευχαριστώ θα ήθελα επίσης να εκφράσω προς τους κ.κ. Δημήτρη Θανασά και Ηλία Γιαλλάφο για την προσωπική τους βοήθεια στην εκτέλεση των πειραματικών μετρήσεων. Η αλληλεπίδραση που είχαμε απέδωσε επιστημονικούς καρπούς οι οποίοι δεν εντάσσονται μόνο στα πλαίσια της παρούσας εργασίας, κάτι για το οποίο είμαι διπλά ευγνώμων.

Ευχαριστώ θερμά, επίσης τον κ. Μιχάλη Γκατζώνη για την τεχνική βοήθεια που προσέφερε χωρίς δισταγμό όποτε αυτό του ζητήθηκε αλλά κυρίως για την επιμονή και υπομονή του σε μερικές περιπτώσεις.

Τέλος, ευχαριστώ τους γονείς μου που με τη διακριτική παρουσία τους με στηρίζουν και με εμπιστεύονται.

Εύχομαι η διατριβή αυτή να συνεισφέρει, έστω και ελάχιστα, στην αποτελεσματικότερη εκπαίδευση των χειρουργών στην αντιμετώπιση του stress που αντιμετωπίζουν κατά τη διάρκεια των επεμβάσεων.

Abstract

The absence of a single reliable detection method makes estimating surgeons' mental stress during an operation difficult to quantify and interpret. Although several non-invasive stress measuring methods have been proposed, most studies have used only one or two stress estimation parameters and produced conflicting results.

The first objective of this study was to assess the feasibility of a new, easy to wear watch-sized device to noninvasively measure cardiac stress parameters such as heart rate (HR) and heart rate variability (IBI) which are usually measured by means of a Holter monitor, as well as Electrodermal activity (EDA).

The second objective was to concomitantly measure the responses of seven noninvasive stress indices and compare them to the video score (VS) achieved by novices in a stressful simulation environment. Namely, we used one subjective method (STAI) plus the three abovementioned ones derived from the wrist-worn device (HR, IBI, EDA) as well as the salivary levels of cortisol, alpha-amylase and IgA.

Twenty-one male novice trainees were enrolled. After an orientation phase, they wore a wrist device that measures heart rate (HR), interbeat interval (IBI) duration, and electrodermal activity (EDA) as well as a Holter ECG rhythm monitoring system (HM). A saliva specimen was collected for cortisol (sC), alpha-amylase (sAA), and secretory immunoglobulin A (sIgA) measurements (baseline phase, BL). Then the simulation exercise phase (E) started, with the subjects trained on a basic suturing module for 15 minutes. Immediately after, another saliva sample was collected. The whole experiment was videotaped, and VS was calculated. The percentage $(E\text{-}BL)_{diff}$ of each of the six parameters was calculated and compared with VS using Pearson's correlation coefficient as well as Akaike Information Criterion (AIC_c).

Data analysis showed: a) When compared to STAI, electrodermal activity exhibited the best correlation, sensitivity and specificity, b) the device derived cardiac parameters highly correlated with the reciprocal Holter values during all experiment phases and c) EDA_{diff} showed the best correlation with VS, followed by IBI_{diff} and HR_{diff}. Among the saliva biomarkers, sAA_{diff} showed the best correlation in comparison to $sIgA_{diff}$ and sC_{diff}.

It is concluded that this wearable device is an easy to use and well accepted by the participants noninvasive tool, which can provide accurate stress estimation in our simulation setting. Additionally, it can replicate Holter derived stress related heart parameters, thus eliminating the need to wear a rather cumbersome device. In our simulation setting, sympathetic ANS stress parameters (EDA, IBI, HR, sAA) could best describe the novice trainees' performance, but sC and IgA could not.

Περίληψη

Η απουσία μιας καθολικά αποδεκτής και αξιόπιστης μεθόδου μέτρησης του stress καθιστά δύσκολο να εκτιμηθεί και να ερμηνευθεί η ψυχική καταπόνηση των χειρουργών κατά τη διάρκεια μιας επέμβασης. Παρόλο που έχουν προταθεί διάφορες μέθοδοι αναίμακτης μέτρησης, οι περισσότερες μελέτες έχουν χρησιμοποιήσει μόνο μία ή δύο παραμέτρους εκτίμησης του stress και έχουν παράξει αντικρουόμενα αποτελέσματα.

Ο πρώτος στόχος αυτής της μελέτης ήταν να εκτιμηθεί η σκοπιμότητα μιας νέας, εύκολης στη χρήση συσκευής μεγέθους ωρολογίου, για τη μέτρηση των παραμέτρων του καρδιακού stress, όπως η καρδιακή συχνότητα (HR) και η μεταβλητότητα του καρδιακού ρυθμού (IBI), οι οποίες συνήθως μετρούνται με τη βοήθεια ενός Holter ρυθμού, έχοντας επίσης και τη δυνατότητα μέτρησης της ηλεκτροδερμικής ενεργότητας (EDA).

Ο δεύτερος στόχος ήταν να μετρηθεί ταυτόχρονα η ανταπόκριση επτά αναίμακτων δεικτών stress και να συγκριθούν με το βιντεοσκοπημένο αποτέλεσμα (VS) που επιτυγχάνουν οι αρχάριοι σε ένα στρεσογόνο περιβάλλον προσομοίωσης. Συγκεκριμένα, χρησιμοποιήσαμε μία υποκειμενική μέθοδο (STAI) συν τις τρεις προαναφερθείσες που προέρχονται από τη συσκευή που φοριέται στο χέρι (HR, IBI, EDA) καθώς και τα επίπεδα κορτιζόλης, άλφα-αμυλάσης και IgA στον σιέλο.

Στη μελέτη περιελήφθησαν είκοσι ένας άρρενες αρχάριοι. Μετά από μια φάση προσανατολισμού, φόρεσαν μια συσκευή καρπού που μετρά την καρδιακή συχνότητα (HR), τη μεταβλητότητα του καρδιακού ρυθμού (IBI) και την ηλεκτροδερμική ενεργότητα (EDA), καθώς και ένα σύστημα παρακολούθησης καρδιακού ρυθμού Holter (HM). Ένα δείγμα σιέλου συλλέχθηκε για μετρήσεις

κορτιζόλης (sC), άλφα-αμυλάσης (sAA) και εκκριτικής ανοσοσφαιρίνης A (sIgA) (βασική φάση, BL). Στη συνέχεια άρχισε η φάση της άσκησης προσομοίωσης (E), με τους αρχαρίους να εκπαιδεύονται σε μια βασική ενότητα συρραφής λαπαροσκοπικά για 15 λεπτά. Αμέσως μετά, συλλέχθηκε άλλο δείγμα σιέλου. Το όλο πείραμα βιντεοσκοπήθηκε και υπολογίστηκε το VS. Υπολογίστηκε η εκατοστιαία αναλογία της διαφοράς (E-BL)diff για καθεμία από τις έξι παραμέτρους και συγκρίθηκε με το VS χρησιμοποιώντας τον συντελεστή συσχέτισης Pearson καθώς και το Κριτήριο Πληροφοριών Akaike (AICc).

Η ανάλυση των δεδομένων έδειξε ότι: α) Σε σύγκριση με το STAI, η ηλεκτροδερμική ενεργότητα επέδειξε την καλύτερη συσχέτιση, ευαισθησία και ειδικότητα, β) οι παραγόμενες από τη φορητή συσκευή στον καρπό καρδιακές παράμετροι συσχετίστηκαν σε μεγάλο βαθμό με τις αντίστοιχες τιμές του Holter σε όλες τις φάσεις του πειράματος και γ) η EDAdiff έδειξε την καλύτερη συσχέτιση με το VS, ακολουθούμενη από την IBIdiff και την HRdiff. Μεταξύ των βιοδεικτών στον σίελο, η sAAdiff έδειξε την καλύτερη συσχέτιση σε σύγκριση με τις sIgAdiff και sCdiff.

Συμπεραίνεται ότι αυτή η φορητή συσκευή καρπού είναι εύκολη στη χρήση και είναι απόλυτα αποδεκτή από τους συμμετέχοντες, η οποία μπορεί να παρέχει ακριβή εκτίμηση του stress στο περιβάλλον προσομοίωσης που δημιουργήσαμε. Επιπλέον, μπορεί να αναπαράγει αξιόπιστα τις καρδιακές παραμέτρους που σχετίζονται με το stress από το Holter, εξαλείφοντας έτσι την ανάγκη να φορεθεί μια αρκετά δυσκίνητη συσκευή. Στο περιβάλλον προσομοίωσης μας, οι συμπαθητικές παράμετροι του stress που σχετίζονται με το αυτόνομο συμπαθητικό σύστημα (EDA, IBI, HR, sAA) περιγράφουν καλύτερα την απόδοση των αρχαρίων, ενώ οι sC και IgA υστερούν.

Part A: Theoretical Background

Stress: Definition and types

To deal with situations that humans normally do not have the resources to deal with, we have developed a biological and psychological reaction called stress. Stress can be defined as a stability imbalance that occurs when an individual is emotionally activated and perceives that cognitive and/or performance requirements outweigh the available resources [1] or as a physiological arousal, which represents an always present mechanism of coping with perceived or real threats or challenges [2]. Stress is believed that greatly contributes in performance and task execution [3].

It is well known that stress modifies human performance and that many factors can provoke stress. Physiologically, the body must first decide whether or not the situation is stressful. This is based upon sensory input in combination with stored memories. Stressful situations normally contain at least one of the following elements [4]:

- reduced or no control of the situation
- unpredictability, something unexpected is happening, or it is hard to predict what will happen
- novelty, something new that the person has never experienced is happening
- threat of ego, one's skill is put to test and one has doubt about one's capacities
- a threat in general
- time pressure.

The different stressors fall in into one of the following two broad categories: Physical (systemic or reactive) and psychological (emotional or processing), with each one of them undergoing different brain processing [5].

One of the most common stressors is the so called ‘‘occupational stress’’. It is defined as the sum of physical, mental, and physiological responses to work which, when intensified, are transformed into negative emotional reactions that

reflect into loss of productivity and reduction in the quality of the service offered [6].

Much attention in stress physiology has been paid to two physiological systems:

a) The hypothalamus-pituitary-adrenocortical (HPA) axis, with the secretion of the glucocorticoid cortisol
b) The sympathetic-adrenomedullary (SAM) system which induces the secretion of catecholamines.

The parasympathetic nervous system is also involved and combined those two systems, form the autonomic nervous system. Simplified, one can say that the sympathetic nervous system is responsible for “fight or flight” responses, while the parasympathetic nervous system deals with “rest and digest” mechanisms. The short-term effects are produced by the fight or flight response and consist of helping the body to deal with the stressor, e.g. giving the body more energy, but longer exposure can lead to health problems as the organism does not have enough time to recover from the stress. Thus, another division exists namely between acute and chronic stress. Both HPA and SAM are involved to the pathologic consequences of stress (e.g. depression, cardiovascular diseases, immune system suppression etc.), as well as they seem to reflect the intensity of stressful situations [7]. Neuroendocrine hormones have major roles in the regulation of both basal homeostasis and responses to threats [8]. More precisely, it has been shown that the exposure to emotionally stressful situations that differ in intensity results to gradual increases in plasma levels of glucocorticoids, adrenaline, noradrenaline, and prolactin [9, 10, 11].

Indices measuring stress

1. Stress assessment tests

There are several tests aiming to assess stress [12]. Two of those tests have gained popularity: The Trier Social Stress Test (TSST) and the State Trait Anxiety Inventory (STAI).

The TSST

The Trier Social Stress Test (TSST) is among the most popular methods of inducing acute stress in order to experimentally study the stress response in human subjects as it reliably increases hypothalamic-pituitary-adrenal axis activation [13]. For a detailed description of TSST see Birkett [14]. A meta-analysis showed that the TSST is the most useful and appropriate standardized protocol for studies of stress hormone reactivity [15].

Many modifications of the TSST have been suggested and combinations with other physiological stress indices have been proposed. Thus some researchers used simultaneous salivary sampling for stress markers such as cortisol (sC) and alpha-amylase (sAA) which may require additional sampling points and measurements to best capture changes in stress response [14], while others measure heart rate variability (HRV), electrodermal response (GSR), blood pressure, eye blink response, or changes in body temperature in response to the TSST [16].

TSST increases sAA levels, and it is suggested that sAA response to the TSST predicts plasma noradrenaline levels [17] but their responses are not correlated [18] and therefore care must be taken in interpreting results.

Although the TSST consistently increases heart rate, its effects on HRV are controversial, with one study showing no change of HRV between a TSST and a control condition [19], while another study indicated that a number of HRV measures changed significantly between baseline and post-TSST [20].

In conclusion, the TSST remains one of the most important tools available to induce stress and its use in conjunction with other stress biomarkers could be useful in future research [13].

The STAI

The STAI is a validated subjective assessment tool for quantifying the stressed state combined with individual traits in the clinical environment [21]. A short version was later introduced [22], i.e. a six-item questionnaire with a 4-point Likert-style rating scale (range 6-24), where a higher number denotes higher stress levels. Usually this questionnaire is completed at baseline and immediately after each session; sessions are categorized as stressful if there is an increase of ≥ 1 in the STAI score from the preoperative to the postoperative measurement.

STAI is usually used in conjunction with other stress performance indices: Thus, a review of surgical stress and performance shows that STAI combined with HRV is a well-validated technique [23]. Additionally, the combination of HRV and STAI provides a feasible method for assessing operative stress in consultant surgeons during elective colorectal resections, since a significant correlation was demonstrated among these two indices [24]. Moreover, in recent years another improvement in the evaluation of surgical stress is the Imperial Stress Assessment Tool (ISAT), which combines both physiological (heart rate and cortisol) and subjective STAI measures of stress [23].

It is evident that the use of stress evaluation questionnaires in conjunction with other established stress measuring indices may be used in future studies to assess the impact of stress on performance in simulation surgical training and surgical techniques**.**

2. Heart Rate Variability

The term heart rate variability (HRV) conventionally describes the beat-to-beat fluctuations in the Heart Rate (HR) or the variations in consecutive R-R intervals.

Due to the development of automated techniques for the quantitative assessment of the rhythmic fluctuations of HR, allowed the detection of tiny changes in the time intervals between heartbeats and thus HRV became a noninvasive useful clinical tool to estimate modulation of autonomic tone [25, 26].

Yang et al., in a recent review characterized HRV as the fourth hotspot in the history of electrocardiogram [27]. In recent years, many studies were carried out in order to investigate the correlation of HRV changes and different body conditions [28, 29, 30].

HRV is traditionally calculated by digital processing of electrocardiograms (ECG). The R-wave peaks of QRS complexes and R-to-R intervals are calculated. Then, HRV parameters are computed using time-domain, frequency-domain, and nonlinear methods [31].

In HRV power spectrum, three components can be found:

1. A high frequency (HF) peak that corresponds to respiratory sinus arrhythmia (>0.15 Hz)
2. A low frequency (LF) peak centered at about 0.1 Hz that is related to arterial pressure control (0.04–0.15 Hz).
3. A very low frequency (VLF, <0.04 Hz; sometimes as a peak) considered to be expression of the peripheral vasomotor regulation.

There is agreement that the HF peak is a good index of vagal activity, whereas the power of the LF peak as an index of sympathetic activity is more controversial since both branches are mediating fluctuations below 0.15 Hz [32]. The most sensitive marker of overall balance of the sympathetic tone is considered to be the LF/HF ratio [32, 33]. However, criticism may be found

among the literature because an agreement on the concept of autonomic balance or interaction does not exist [32].

A lot of HRV parameters are commonly used in the literature. These parameters often refer to a professional term, NN intervals (“normal-to-normal intervals”), which means that only regular heartbeats should be considered [31].

To overcome the traditional limitations of ECG acquisition, a lot of new technologies have been developed to measure HRV without ECG, such as photoplethysmography [34], ballistocardiography [35], a microwave sensor [36], a webcam [37, 38], or even a smartphone [39].

The subject’s gender seems to play a role in HRV: In a recently published meta-analysis, males showed a lower mean HR, a significantly higher mean RR interval and a higher LF/HF ratio [40].

HRV has been used as an indirect measure of the cognitive load and mental strain placed on operating surgeons: In a number of studies it has been shown that HRV was significantly increased as a result of intra-operative stress [41, 42, 43, 44].

Decreased HRV has been shown to correlate with acute stress: In a recent metaanalysis of studies dealing with the mental stress and strain analysis of surgeons, 10 studies were found, 4 of them using only the HR and 6 studies used the HRV as a stress parameter. Control groups have rarely or not been studied. Data analysis showed that stressed surgeons had a higher intraoperative HR and a low expression of the HRV and that the same pattern was experienced in operating surgeons compared to the assistant surgeons or with inexperienced operating surgeons compared to experienced surgeons [45].

A combined objective (HR and HRV) and subjective (STAI) method was used by Jones et al., in order to evaluate the experienced stress of consultant colorectal surgeons performing eighteen elective colorectal anterior resections. During operation, all the surgeons exhibited a significant increase of stress as seen from

the HRV values, and the peaks of stress, according to the operative step, was comparable across the procedures and the surgeons. Moreover, significantly positive correlations were detected between HRV measurements and perceived stress as measured by STAI [24].

Finally, HRV was assessed in two consultants and three surgical fellows during the dissection around the recurrent laryngeal nerve while performing total thyroidectomy. The overall heart rate, time, and frequency domain parameters of HRV, specifically the (LF/HF) ratio, were correlated with the surgical role, particularly while teaching the surgical fellows at critical points. It was found that this type of teaching is associated with a measurable increase in mental strain of consultant surgeons, as determined by HRV. When acting as primary operators, fellows showed increased stress levels as well [44].

3. Electrodermal activity (galvanic skin response)

One of the three types of sweat glands are the eccrine glands which, except the hairy skin, they are also found on the glabrous skin of the palm and sole, where they are not usually activated by heat, but rather by deep respiration, mental stress, and local tactile stimulation, producing the so-called "emotional sweating". It seems that the limbic system plays a role in emotional sweating and therefore the measurement of sweat output on the palm or sole is useful for evaluating sympathetic function and limbic activity [46]. As eccrine glands in these regions fill and drain in response to autonomic stimuli, they produce fluctuations of electrical conductivity and thus galvanic skin response (GSR) aka Electrodermal activity (EDA) is measured. Unfortunately, due to obvious reasons, these two locations are unsuitable for performing measurements in a surgeon while operating or to a trainee performing a simulation task and therefore GSR devices are usually placed in the wrist, or the ankle [1].

GSR is a well-accepted and validated of emotion and cognition and it has been shown that GSR correlates with the number of active sweat glands [47]. Each

device contains 2 silver chloride electrodes that measure and record electrical conductance at 10 times per second (Neumitra Co, Boston, USA). During measurements care must be taken in order to avoid sweating artifact due to changes in ambient temperature.

As large (over 200 times) fluctuations in GSR readings even during a short time period have been typically detected inter- individually, probably the mean GSR could be a more stable estimate of the overall level of physiologic arousal, and maximal GSR readings would be a more sensitive measure for detecting differences between groups and would show better correlation with state data [1].

Studies have demonstrated that functional infrared imaging (an image reporting pixel to pixel the recovery time after a stimulus has been applied) and GSR have similar detection power [48, 49, 2, 50, 51].

In a recent study, the undergraduate students' responses before and after an exam session were assessed, by completing an academic achievement emotions self-report and an interview that paralleled these questions when participants wore a GSR sensor and salivary biomarkers (cortisol, estradiol, progesterone, testosterone and DHEA-S) were collected. Data collected from the three methods showed that GSR responses were unrelated to the exam type, but GSR increased during both emotional activation and cognitive as well as verbal recollection [52].

4. Thermal activity

Under acute stress, sympathetically mediated vasoconstriction occurs which together with stress-induced thermogenesis result in stress-induced hyperthermia which is proportional to the intensity of the stressor [53]. The occurring vasoconstriction is most pronounced in extremities relatively rich in arteriovenous anastomoses, such as the fingers [54] and thus different skin

regions vary in temperature changes under acute stress, for example, cooling occurs on the nose but not the cheeks [55].

The primary assessment of skin temperature is primary performed via contact devices. However, the contact devices may present a number of conductive limitations plus some discomfort which may lead to stress [56, 57, 58].

Recently, non-contact infrared devices for measurement of skin temperature are available and thus infrared thermography (IRT) became widely accepted as a non-invasive alternative to conductive devices without altering stress levels [59, 60, 61].

The impressive advancement of the technology in infrared cameras permits integrating this technology into automated systems for remote and automatic monitoring of physiological activity [62] and thus IRT became a potential system for the quantitative assessment of sympathetic activity in stress research [63].

However, the agreement between IRT and conductive instruments is limited and equivocal as a measuring discrepancy between infrared thermometry and conductive instruments seem to exist [64, 65,58, 66].

The reliability of thermal IR imaging as compared to established stress markers (HR, HRV, finger temperature, sAA, and sC) was assessed in healthy subjects during two standard laboratory stress tests: the cold press or test and the TSST. Thermal responses of several regions of the face were detected during both tests. A weak correlation between the thermal imprints and the stress markers was shown, while the thermal responses correlated with stress-induced mood changes, but on the contrary the established stress markers did not [50].

However, there are several limitations for using thermal IR imaging: Beyond stress, a lot of other confounding parameters such as increased mental workload [67], pain [68], positive experiences [69, 70] and homeostasis changes due to environmental conditions [71], can trigger an acute change in skin temperature. In addition, gender may play a role, as higher facial skin temperature was found

in males possibly due to higher blood circulation and metabolic rate [72].Additionally, emotions are another factor that modifies the thermal response of different body regions, particularly in the face, and thus the emotional status of the assessed subjects must be considered [55, 62]. Finally, from a technical point of view, taking into account how important precision is in measuring human temperature (more than 0.25ºC of side-to-side asymmetry is considered to be abnormal), the accuracy of the best infrared cameras can be more than 1ºC (or 1%) different from the actual temperature. The burden imposed from all these factors is considered as one of the weakest points of infrared thermography [62].

Also, despite the advantages offered by thermal IR imaging, it has to be taken into account that thermal signal development as a result of vascular change, perspiration, or muscular activity is rather slow with respect to other established techniques.

There are a few studies published regarding stress quantification in surgeons by means of an IR camera: Pavlidis et al., quantified stress in the context of surgical training by measuring transient perspiratory responses on the perinasal area through thermal imaging [2].

In conclusion, the current body of research suggests that IRT does not sufficiently agree with traditional conductive methods of skin temperature measurement and proper considerations should therefore be taken when monitoring thermal expression of psychophysiological activity [63].

5. Biochemical stress markers (saliva)

Saliva: Current knowledge

The most comprehensible definition of biomarker is “a cellular, biochemical, molecular, or genetic alterations by which a normal, abnormal, or simple biological process can be recognized or monitored” [73]. Therefore, any molecular substance or compound which demonstrates significant variation in

concentration, as compared to those of control subjects, is a potential biomarker [74].

Whole saliva comprises of a mixture of fluids, secreted from the salivary glands (parotid, submandibular, sublingual, and the minor gland), gingival fold, oral mucosa transudate, and mucous from the nasal cavity and pharynx, with varied composition [75, 76, 73].

In healthy individuals, depending on age and gender, the unstimulated salivary flow rate is between 0.1-2 mL/min [77]. Additional factors influencing unstimulated salivary flow and composition include individual hydration, body posture, lighting, smoking, circadian and circannual rhythms, and medications [78].

Although blood still remains the best body fluid for evaluation of many biomarkers, saliva offers a promising diagnostic alternative for objectively assessing physiological and/or psychological stress. Its use as a diagnostic fluid in stress research has increased exponentially due to the following factors:

- The collection of saliva is a convenient sampling method because it is noninvasive and relatively non-stressful. Been non-invasive, it causes minimal distress to individuals [79] avoiding the stress from the fear of needles pain [80] as well as the ethical concerns associated with invasive measures.
- It is constantly being produced so that samples can be obtained at short intervals and in enough quantities [81].
- Saliva collection does not require any equipment apart from a collection device nor skilled healthcare professionals, thus been suitable in field research in the subjects' natural environment [81, 82].
- A growing number of saliva analytes and collection kits have been recently commercially marketed worldwide [74].

Last but not least, budget estimates incorporating personnel expenses and corresponding collection methods for both blood and saliva collections showed that saliva collection was 48% less costly than blood collection [83]. Therefore, for all the reasons described above, saliva has been established as a cost effective and non-invasive diagnostic tool [84].

Saliva's stress biomarkers

Salivary levels of cortisol, alpha-amylase, secretory Immunoglobulin A and chromogranin A have been associated with stress state [85, 86, 87].

A summary of the available stress measuring salivary biomarkers is shown in Table 1.

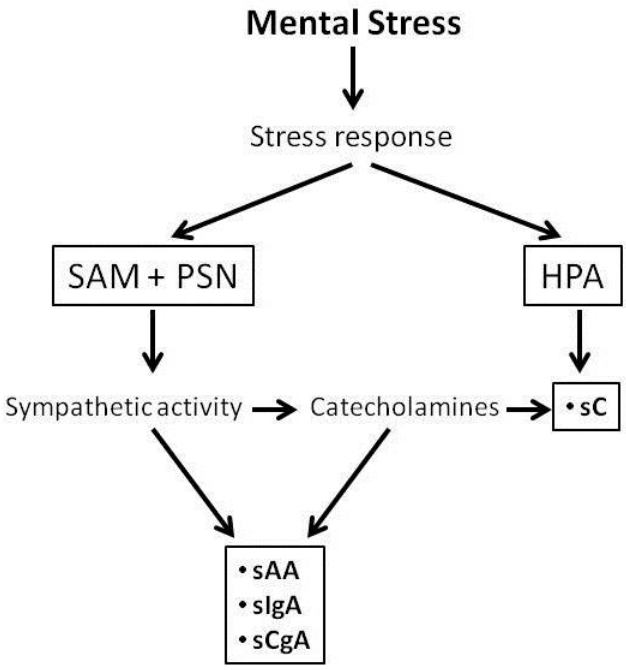
Table 1. Available stress measuring salivary biomarkers (Modified from [88]).

Salivary Biomarker	Advantages	Limitations
Cortisol (sC)	• Level corresponds well to that of plasma • Short lag time • Relatively stable & can be stored for several weeks	• Only measures free fraction of cortisol • Influenced by conditions that affect CBG levels • Inter-individual variability
Salivary a-Amylase (sAA)	• High sensitivity to stress related changes, with dose-response relationship	• Salivary flow rate also affected by PSN • Production is dependent on numerous salivary glands that have different responses to stimulation • Inter-individual variability
Secretory Immunoglobulin A (sIgA)	• Sensitive marker of fatigue	• Huge biological variation & inter-individual variability • Not as sensitive as other stress biomarkers
Salivary Chromogranin A (sCgA)	• Reflects sympathetic nervous activity	• No correlation between sCgA & serum CgA
Total protein	• Used to adjust concentrations of sIgA	• Can be influenced from gum diseases

CBG: Cortisol binding globulin, PSN: Parasympathetic nervous system

For a better understanding of how stress influences the salivary secretion of those stress biomarkers, it is necessary at first to recall their physiology. A summary of the secretion physiology of the salivary stress biomarkers is shown in Figure 1.

Figure 1. Secretion mechanism of salivary biomarkers after mental stress (adapted from [86].



From a physiology point of view, these salivary markers can be classified into two groups:

1. Plasma circulating markers able to enter into saliva (e.g. cortisol)
2. Locally produced saliva markers (e.g. α-amylase, secretory IgA).

This classification is relevant to methodology and interpretation of salivary findings, as each source involves different mechanisms.

A reliable measurement of blood-borne constituents requires a constant saliva/plasma ratio (SPR), assuming that the concentration in saliva follows intra-

and inter-individual variations in plasma [89]. This is the case of cortisol which has an excellent SPR, entering saliva through passive diffusion and thus its salivary concentrations reliably reflect those found in plasma [90].Unfortunately, cortisol is somewhat of an exception as most of the plasma derived molecules cannot be reliably assessed in saliva.

In contrast, most saliva proteins are released upon activation of the innervating sympathetic nerves and thus the amount secreted per time unit corresponds to the level of local sympathetic activation [91]. Thus, salivary protein concentration reflects the combined effect of SFR (which is largely parasympathetic) and protein secretion (which, in sympathetically innervated glands, is largely sympathetic). Therefore, the PNS is mainly responsible for the secretion of ‘high fluid-low protein saliva’, while the SNS mainly elicit secretion of ‘low fluid-high protein saliva’ [92].

sAA levels are not related to α-amylase levels in blood, which are derived from pancreatic secretion [93]. Parotid and palatinal glands are very rich in sAA, whereas sublingual and submandibular glands secrete much less sAA. Passively, most saliva is secreted by the submandibular glands and only about 20% derives from the parotid gland [93]. However, during mechanical (chewing) or gustatory stimulation about 50% of all saliva derives from the parotid glands. Because parotid saliva contains a 4- to 10-fold higher sAA concentration than submandibular saliva [94], sAA concentrations in whole saliva are likely to drastically change independently of any higher CNS regulation [95].

Some sAA-rich salivary glands, like the sublingual and minor glands, are strongly under PNS control [95] and it has been found that the sympathetic effects on sAA release are strongly moderated by concurrent PNS activity, a phenomenon denoted as ‘augmented secretion’. Additionally, it is known that over the course of an acute stressor, the PNS tends to exhibit a faster off and onset than the SNS, while the PNS withdrawal rapidly restores immediately post-stress, while sympathetic activation still lingers [96]. It is exactly at that period that the largest discrepancy between sAA concentration and secretion can be anticipated with

obvious implications for sample timing [95]. Furthermore, a PNS rebound immediate post-stress has been reported, whereby PNS activity overshoots baseline levels, causing a transient sympathetic-parasympathetic co-activation [97, 98].

sIgA is also not derived from blood; instead it is produced by β-lymphocytes adjacent to the mucosal cells, then transported through the cell, and secreted into saliva as secretory IgA [99].The production of sIgA from the various salivary glands is not the same, with the highest levels found in the minor saliva glands [99, 100]. Saliva sIgA levels are affected by flow rate, and therefore it is suggested to measure flow rate as well in order to express sIgA secretion as a function of time [101].

Similarly to cortisol, it is known that sIgA levels increase under mental stress conditions [82, 102]. It appears that while chronic stress down-regulates sIgA, acute stress elevates sIgA [103, 104].

Finally, sCgA is produced by the submandibular glands and secreted into saliva [105]. The correlation between saliva and serum CgA is not well elucidated [106, 107].

Furthermore, on a local cellular level, the salivary glands are capable of producing highly differentiated protein responses to different stressors [91, 89]. This differential activation can cause a myriad of additive, synergistic or antagonistic intracellular responses between and within glands. A typical example of overruling central regulation (e.g. stress) is reflex secretion which is a mechanism of glandular secretion that is independent of the central neural effects of stress [108].

It is therefore regrettable that the abovementioned physiology knowledge is generally not taken into account in the published literature regarding stress biomarkers evaluation during surgery/simulation.

Cortisol (sC)

The most well studied salivary stress biomarker is cortisol [109, 110, 111, 112]. Salivary cortisol is widely considered as a valid indicator of free cortisol and hypothalamus-pituitary-adrenocortical-axis [113, 112]. It has been shown that the rise of sC levels start within 5 min of increases in plasma cortisol, with the highest levels occurring at 31–40 min after onset of stressors and they are strongly correlated with plasma cortisol concentrations [15, 111, 114, 115].

There are several sC kits available commercially, which commonly use an enzyme-linked immunosorbent assay (ELISA) and electrochemiluminescence (ECL) both been validated and accepted for sC measurement [116].

Many studies demonstrate that acute stress elevates salivary cortisol [14]: In army nurses exposed to a combat casualty simulation a marked cortisol response to real-life and simulated emergency situations have been documented [117], as well as in female medical students during an ambulatory patient consultation in their 6th year of study [118].

However, several other studies have failed to find elevations in sC following acute stressful situations [119, 120, 121]. This may be due to the many structures that regulate the HPA-axis (hippocampus, hypothalamus, pituitary, adrenals), and the complexity of their respective modulators, receptors, or binding proteins, as well as their various stressors that may differently affect sC [113]. Thus, the link between mental stress and cortisol reactivity has been shown to be plausible [122].

Additionally, other factors such as medical skills and knowledge, the participants' awareness of their actual performance and the enhancing effects of endocrine stress responsiveness, might also interfere. Furthermore, coping strategies have been shown to be associated with medical performance and thus may also have modulated stress responses and the association between stress response and medical performance, respectively [123, 124]. Therefore, the relationship

between endocrine stress response and medical performance remains speculative and should be addressed in future studies [125].

Saliva Alpha-Amylase (sAA)

Saliva Alpha-Amylase (sAA) is a digestive enzyme that that breaks down insoluble starch into soluble maltose and glucose [93].

The enzymatic activity of sAA (in Units/ml) can be measured with commercially available kits based on chromogenic or fluorogenic substrates that utilize its enzymatic activity as a proxy for sAA concentration [126].

sAA salivary concentration increases rapidly during acute stress and thus it has been proposed as a means to monitor ANS activity [92, 127, 128, 129].

However, it was found that a stressor eliciting SNS-PNS co-activation (viewing a surgical video) caused a marked sAA release, whereas a cognitive stressor causing SNS activation in conjunction with PNS inhibition (a memory search task) showed no significant change in sAA release. Importantly, the latter stressor caused a much stronger sympathetic activation than the stressful video [91]. Therefore, it seems that sAA is not a reliable index of SNS activity, since it also co-represents a moderating effect of PNS activity [92].

It seems that sAA responds to stress quicker than sC: It has been shown that in lay people who watched a video about cornea transplantation, sAA activity rose within 3 min and decreased to baseline values shortly after the end of the video while sC concentration rose within 6 min and decreased slower as compared to sAA activity [130]. Additionally, it has been shown that repetitive training in a high-fidelity simulator in simulated crisis situations reduces the increase of sAA concentration but not that of sC [4]. This difference between sAA and sC may be due to the fact that sAA is a marker of the SNS system and thus may be differently affected by stress as compared to cortisol, which is a marker for the HPA axis [131].

In a recent study, 34 participants underwent an 8-minute memory task (MT) as well as a cold press or task (CPT) under continuous monitoring of cardiovascular SNS (pre-ejection period, blood pressure) and PNS (HR variability) activity. Unstimulated saliva was collected repeatedly during and after each laboratory stressor, and sAA concentration (U/ml) and secretion (U/minute) was determined. It was concluded that the sAA interpretation as a measure of SNS activity, or as a physiological marker of stress, is less straightforward than often assumed, depending on sample timing and flow rate, likely involving both SNS and PNS influences [132].

Finally, in a recently published study, sC and sAA were measured in response to a trauma reminder. The sAA reactivity was associated with neural reactivity in the salience network in response to novel negative scenes and neural hypervigilance as indexed by reactivity to novel neutral scenes as measured using functional magnetic resonance imaging. sAA, but not cortisol, increased in response to the trauma reminder [133].

Due to the reasons above, much attention has to be paid regarding collection timing and SFR of sAA measurements in future stress studies [132].

Secretory Immunoglobulin A (sIgA)

Secretory Immunoglobulin A (sIgA) is a subclass of Immunoglobulin A, being the main immunoglobulin found in mucous secretions from many mucous systems and organs including the salivary glands [134]. sIgA exhibits a diurnal rhythm, decreasing from the highest levels in the morning to the lowest in the evening [135].

sIgA is sensitive to physical and psychological stress through interactions with the ANS system [101, 136].

Campisi et al., using the TSST found significant increases in heart rate, systolic blood pressure, and sC plus a non-significant increase in sIgA levels following

the TSST, which significantly decreased during the recovery period. They thus suggested that exposure to acute psychosocial stress in the form of a TSST activates the physiological stress response [137].

Salivary Chromogranin A (sCgA)

Chromogranin A (CgA), or parathyroid secretory protein 1, is a major soluble glycoprotein occurring into many types of neuroendocrine tissues [138]. CgA is involved in the intracellular storage of catecholamines and is co-secreted from sympathetic nerves along with the catecholamines into the blood [139]. Thus, CgA has been proposed as a promising biomarker of ANS activity [140, 141, 142].

It has been shown that sCgA changes sensitively and rapidly in response to various psychological stressors, such as psychosomatic stress [143], computer operation psychological stress [144], academic assessment stress [145] and psychological tension stress before surgery or anesthesia [142, 145]. Additionally, Kanamaru et al., found a relatively rapid increase in sCgA levels, from just before to within 20 min after the start of a cognitive test battery, thus concluding that the changes in sCgA secretion may indicate psychological stress in subjects exposed to a cognitive task [146]. Furthermore, Nomura et al., found that sCgA concentration increased during the mental stress tasks and decreased during the intermissions, demonstrating the usefulness of sCgA as a biomarker for a short-term mental workload [147].

It seems that sCgA changes more rapidly and more sensitively to psychological stressors than sC [143]. The same group also found a prompt elevation in sCgA levels and a delayed increase in sC levels when psychosomatic stress was induced by a test involving an oral presentation in front of an audience or a driving situation [148]. All these studies suggest that sCgA may be a sensitive and promising index for psychosomatic stress [149].

However, others have reported that the measurement of sCgA may not be a good parameter to evaluate the ANS activity [142]. Dimsdale et al., found that within the normal physiological range, as exemplified by small postural changes and mental stress (orthostatic challenge and mental arithmetic), plasma CgA level is stable or else slow to respond and that basal CgA levels were unrelated to other manifestations of sympathetic nervous system activity [150]. Similarly, Noto et al., examined the relationship between STAI score (STAI-s) in subjects exposed to the 15 min mental arithmetic task stress and sCgA. The STAI-s was not significantly correlated to sCgA and they concluded that sCgA was not a good marker of stress in their study. They also described that a mental arithmetic task may change autonomic tone, but this change may not be strong enough to increase sCgA [141].

Total protein

Although total protein is a non-specific stress biomarker, it is used to examine changes in overall protein secretion in saliva and thus to normalize concentrations of various salivary proteins since concentrations can vary significantly in response to stimulation of saliva flow [151, 152, 153, 154, 155, 156]. However, it has been suggested that this practice may cause misleading results due to differences that exist in the control of secretion of individual salivary proteins among the different salivary glands [99, 101, 157].

Stress in Surgery

A limited amount of studies examined the psychobiological stress response in highly stressful professions such as policemen, fire fighters, emergency care, paramedical, and military [121]. These professionals showed increased endocrine, cardiovascular and inflammatory markers during real life emergency situations [118, 158].

It is well known that surgeons work in a stressful environment due to the abundant surgical challenges, the technical requirements, and the time pressure they face every day. It is also known that acute stress has a direct impact on surgical performance and patient safety [159, 160]. Moreover, surgical teams frequently encounter highly complex crisis situations, which can cause considerable stress and can, in turn, directly contribute to patient outcomes [161, 162]. Additionally, due to the evolving nature of surgical techniques, surgeons will continue to experience increasing acute stress, thus, making stress assessment more relevant and necessary than before [163].

An in-depth understanding of how surgeons perceive operative stress is still missing [24]. Several factors have been suggested to affect surgeons' operative practice including mental and environmental stressors [4] such as bad perioperative sleep quality [164], engagement levels and emotional exhaustion [165], complex or rarely performed cases, lack of experience in the performance of new surgical procedures, such as laparoscopy [41], and poor assistance [160]. Therefore, the ability to implement a coping strategy to deal with stress seems to be important for enhancing performance [124]. In order to address these issues, formal intraoperative stress management training becomes a necessity [160, 166].

A battery of stress measuring tools, including objective and subjective methods have been used in the literature to estimate stress in surgery [24, 45, 166].

A rise in sympathetic tone was observed in surgeons, whilst performing a range of operations [42, 44, 167].

Regarding stress biochemical markers, only few studies have measured the changes of specific stress biomarkers: It has been shown that the occurrence of stress in surgeons may be associated with their lack of experience as well as from advanced surgical procedures such as laparoscopy [41]. It was also shown that monitoring saliva cortisol secretion and heart rate (HR) variations in operating surgeons is crucial in order to be able to prevent the consequences that typically arise due to the lengthened and repetitive stress states [7, 168].

In another study, salivary levels of cortisol, alpha-amylase, IgA and chromogranin A (CgA) and cardiovascular (HR and systolic blood pressure) stress responses of oral and maxillofacial surgeons, engaged in different surgical procedures with increasing degrees of technical difficulty, were analyzed. It has been found that higher stress management ability was present in the more experienced surgeons as compared to those with less experience [5].

Associated incidents in critical anesthesia, can be quite stressful for anesthetists, but not many studies have been carried out that address the stress levels experienced by anesthetists and their responses during surgery [169, 170, 171].

Apart from individual involvement, surgical teams encounter frequently highly dynamic and complex crisis situations, which can cause considerable stress upon the individuals of the team, which in turn can directly contribute to patient outcomes [161, 162, 172, 173]. It has also been shown that stress can affect the management of crisis situations in the intensive care due to a diminished perspective from the team [174].

Simulation in Surgery

Simulation-based training (SBT) has long been utilized for hundreds of years in both medical and surgical education [175]. Leadership in the medical professions has embraced simulation and surgery pioneered in SBT to teach residents and students [176]. The first SBT specific to surgeons can be traced to Sushruta, an Indian scholar from approximately the fourth century BC. His Sanskrit language text was translated in 1907 and contains descriptions and lessons for students to practice inserting their knives and tools into natural objects mimicking body parts [175]. Over the recent years, a great concerted effort took place in order to develop organized SBT curricula with assessments to teach and evaluate the cognitive, technical, and decision-making skills of the trainees trying to enhance education and patient care [177].

For the last century, surgical educators have modeled the Halsted's classic model of surgical education. According to this surgical training model, also known as "see one, do one, teach one", the trainees receive a significant amount of their education from the surgical faculty, by observing/participating in the operating room and caring for patients on the ward. Through this experience, students were expected to learn the knowledge, skills, and attitudes needed for efficient patient care [178, 179].

However, during the last two decades we witnessed significant changes in health care delivery and these changes have required surgical educators to adapt their approach to education [177]. The main reasons for these changes are mainly due to:

1. The ever-increasing amount of new medical knowledge that the trainees are expected to learn in putting huge pressures to reform medical education [177, 180].
2. Changes in patient safety, quality, and expectations on surgical faculty [180].
3. Work hour limitations and loss of surgical teams [180, 181].
4. A move toward competency-based education and the creation of entrustable professional activities [182].

SBT combined with assessment tools appears to be an answer to these changes. It has been shown that SBT:

- is a validated method
- increases safety
- improves performance
- allows an objective skills assessment
- is financially appealing since it diminishes costs of training - both direct costs as well those due to extending time in the OR.

Therefore, SBT became a teaching method of choice, since it provides a safe environment for a student to fail, practice, and learn before a patient encounter. In prospective, randomized studies, learners who trained to a benchmarked proficiency level on the simulator performed significantly better than learners who were traditionally trained [183].

The recognition of simulation as a highly effective tool to teach a variety of technical and nontechnical skills to students, residents, and practicing surgeons has been critical in this transformation. The value of simulation is now widely accepted and integrated into health care training, from undergraduate medical education through continuing medical education, across and among professions, and within health systems. Therefore, the new model of medical education which replaced the Halsted model can be summarized as: ”See one, practice many, do one”. We have moved beyond validating simulators and now understand that the process of gathering validity evidence for an educational activity is an ongoing process grounded in similar scientific principles to the evidence a clinician would consider when choosing a treatment for a patient. Academics are now applying more rigorous criteria to the study of simulation and its impact on the knowledge, skills, and attitudes, as well as the health care outcomes at the patient and population level. In addition, we have a broader perspective of the applications of

simulation throughout health care. Simulation-based training has helped uncover the importance of human factors engineering in designing systems that facilitate desired behaviors or outcomes and that actively discourage unsafe or inappropriate actions. We are learning more about the impact of teaching methods and developing the skills of those educators who teach using simulation [184].

As a result of more rigorous scientific study of how technical skills are acquired, and the accumulation of validity evidence around specific assessment tools, simulation is used in several high-stakes examinations in undergraduate medical education and for professional certifications. This is truly an educational achievement, to test a person's skills through the performance of those skills, instead of just their ability to answer multiple choice questions. With simulation, the skills being taught and learned can be uncoupled from the stress of the clinical environment, to improve learning and cognition for the student/novice learner [183, 185, 186].

Surgical SBT over the last 20 years has gone from being used by a few educators teaching traditional skills to a main focus of surgical education. Simulation has always been used to teach surgical skills, but the technique is rapidly becoming the standard for early surgical education. Key SBT benefits include its ability to compartmentalize education and combine immediate assessment and feedback, all in an effort to accelerate skills acquisition for the trainee. SBT value is demonstrated in its being adopted in multiple national medical student surgical educational initiatives [176].

Simulators can be divided into multiple classification schemes such as animate/inanimate, virtual/real, or partial task/procedural. These classifications can be combined to describe the various characteristics of a simulator. A simulator, however, does not in and of itself constitute a curriculum. It is but a tool and it must be properly paired with the learning objectives of what is being taught and valid assessment and feedback for the learner. In SBT, many types of simulators are currently employed (Table 2).

Table 2. Simulator classification (modified from ref #2)

Simulator	Description	Uses	Cost
Abstract partial task trainers	Knot tying boards Foam suture models Simple laparoscopic box trainers	Teaching a repetitive simple task such at suturing or knot tying Peg transfer for laparoscopy	$ Most are very inexpensive & reusable, or students can take them home to practice
Physical anatomic trainers	Silicone, plastic or foam replicas of humans or portions of humans	Teaching a complete procedure with multiple smaller steps –Foley, nasogastric tube, central line placement, intubation	$$-$$$ Commercially made products can get expensive for an entire school
Human simulators	Replicas of humans, may be computerized, may be simple or more complex	Team training, cardiopulmonary resuscitation, airway & intubation, codes	$-$$$$ Computerized mannequins can be expensive & take training to run but are very realistic
Computer Based	Programmed clinical scenarios that can be repeated over and over	Very useful for teaching clinical decision making and pathways, allows for retesting across time &assessing knowledge acquisition	$-$$$ Initial purchase or writing of scenarios Subscription rates can build & increase
Standardized patients	Professional patients to volunteer students. Can assist with the assessment & adds realism	Physical examination skills, team training, immediate feedback	$-$$$ Professional patients can cost, volunteer students require training
Animate models	Live animals or food grade portions of animals	Performing complete operations or adding realism to tissue consistency for partial task trainers	$-$$$$$ Store bought models are reasonably priced, but live animal models are expensive due to housing, anesthesia & disposal costs
Virtual anatomic trainers	Image-guided computer-assisted specific task trainers	Laparoscopic, endoscopic & vascular procedures	$$$$$ Initial purchase & maintenance
Virtual reality	Completely interactive computer-based environment	Usually very realistic Performance scores can be delivered	$$$$$ Very high upfront cost; hard to use with large student volumes

The end result of a good simulator with well-designed metrics is a training system where trainees can learn both what TO do and what NOT to do when operating on patients. The errors must be quantified so as to be completely unambiguous. Simulation training is optimal with metric-based feedback, particularly formative trainee error assessments, proximate to their performance. Without robust metrics the simulator is at best an expensive video game and at worst an adverse outcome waiting to happen. Ii is through those metrics that we can distinguish the different stages of skill development approximate to a ‘traditional’ learning curve and set benchmark criteria for any surgical curriculum [187].

Finally, surgical education has adopted simulation as a compulsory component of the specialty exams in different countries. The revolution of using simulation is happening now and it’s all about improved patient care through advanced surgical education.

Stress in simulation training

In contrast to real surgery ever changing cases, simulators provide a standardized milieu that allows inter- and intra- individual comparisons among participants [163]. Furthermore, simulation training is a well-established method for improving technical and non-technical skills in a controlled and quantitative manner [188]. Training in medical simulators focuses on improving the skills and performance of the trainees, so that they are able to tackle more routinely critical situations, thus resulting to lower stress levels after their training. It is assumed that a high-fidelity simulator can replicate real psychological challenges and stress [123] and yet it can facilitate lower stress loads in real environments [189, 190]. A simulation-based study showed that surgeons with insufficient stress-coping strategies demonstrated poor performance in virtual laparoscopic procedures [123].

Among the several methods proposed to measure and assess a trainee’s competence in a simulated environment, it is evident that technology-based

measures can avoid bias and should be preferred to quantify performance and provide objective metrics [191]. Today, various performance parameters derived from technology-based assessment tools have been proposed and used, such as motion tracking, visual attention, stress responses, and video analysis. Each one of them has advantages and disadvantages and corresponds to a specific objective interpretation of performance [192]. The complexity of stress mechanisms and the involved factors make acute stress estimation and measurements difficult to interpret.

However, only limited data exist in the literature regarding surgeons' acute mental strain measurement methods in the surgical operating and simulation field, while several stress-measuring tools, including subjective and objective methods, have been used for this purpose.

Arora et al., demonstrated through direct correlation that stress (measured using questionnaire, heart rate and salivary cortisol) impairs the surgical performance of novice surgeons assessed on a simulator [159].

Additionally, during simulated procedures, the effect of stress on the performance of the operation was investigated. It was shown that the experience of the surgeon was the single best predictor of performance [124].

Keitel et al., assessed salivary cortisol and psychological responses (visual analogue scales, VAS) in medical students every 15 min from 15 min prior to until 60 min after been exposed in three conditions: rest, laboratory stress (LS; public speaking), and simulated emergency situation (SIM; myocardial ischemia and ventricular fibrillation). As compared to rest, cortisol increased significantly in both stress conditions but at different time courses. They concluded that SIM is a profound stressor. Since they found a positive relationship between salivary cortisol and medical performance as measured in a simulated emergency situation, they suggest that high stress responsiveness might be a predictor of good performance, while the high stress response might counteract educational efforts associated with training using high-fidelity patient simulation [125].

In one study, 29 intensivists took part in one of six 1-day simulator courses randomized to either crew resource management (CRM) training, which contains psychological teaching and simulator scenarios, or classic simulator training. Before and after the course each participant took part in a 10-min test scenario. Saliva amylase and cortisol were measured before the scenario, immediately after, and 15 min later. The Anaesthetist's Non-Technical Skills (ANTS) was used to evaluate the non-technical skills and a checklist for assessing clinical performance. Both cortisol and amylase showed a significant increase during the test scenarios, however, in the post-intervention scenario the stress response of salivary alpha-amylase was significantly smaller but not that of cortisol. Both ANTS scores as well as clinical performances were significantly better in the post-intervention scenario. Neither stress nor performance differed between the two groups [4].

Phitayakorn et al., tested the practicality of a watch-sized device to measure galvanic skin response (GSR) in operating room (OR) interprofessional members during a "cannot intubate/cannot ventilate" simulation scenario. Participants wore sensors (Neumitra Inc, Boston, USA) on the wrist (all) and ankle (surgeons and scrub nurses/technicians) during the orientation, case, and debriefing phases of the scenario and completed the STAI test. Mean wrist GSR levels significantly increased from baseline and remained elevated even after the simulation, while no correlation was found between wrist and ankle data. Surgery residents were the only group that demonstrated continued increases in wrist GSR levels throughout the entire simulation. Large intraindividual differences (#200 times) were found in both wrist and ankle GSR. Analysis of the STAI scores and GSR levels showed that residents had significantly higher trait anxiety than the nurses and surgical technicians and, surprisingly, senior practitioners had significantly higher trait anxiety than junior practitioners. They concluded that continuous GSR monitoring during OR simulation is feasible but would be difficult to implement in an actual OR environment. The large variation in individual levels of physiologic activation suggests that additional studies are required to better understand how OR members respond to stressful situations [1, 193].

Stress measuring wearable devices

Acute stress affects the autonomic nervous system (ANS) that modulates heart rate (HR), electrodermal activity (EDA) and thermal activity levels [194].

HR and Heart Rate Variability (HRV) are the most commonly used parameters. HRV is calculated by digital processing of electrocardiograms (ECGs) and corresponds to the beat-to-beat changes in heart rate [195] and is a non-invasive tool to estimate heart rate modulation by the ANS in a variety of settings including stress [196, 197]. Several HRV parameters are computed using time-domain, frequency-domain and nonlinear methods and emerged as a non-invasive tool to estimate the vagal activity in several conditions, including monitoring of athletic responses to training [39, 40]. Decreased HRV has been reported as a predictive factor for adverse outcomes in disease states and has been found to be associated with fatigue, stress, and even burnout during athletic performance [198, 199, 200].

Despite being accepted as gold standard methods for HR and HRV interval monitoring and analysis, both the classical ECG and the ambulatory Holter monitoring have several drawbacks regarding the proper and accurate detection of RR intervals. For example, patients with tremor or elderly patients with fragile skin have bad quality of recordings with a lot of noise and artifacts [201]. Similarly, other factors such as surface electromyography, increased electrode impedance, respiration induced baseline drift, and electrode contact movement can cause noise and motion artifacts especially when the subject is motion [202].

Additionally, morphological variations in the ECG waveform and heterogeneity in the QRS complex can often make difficult to identify the RR interval [203]. Another limitation can be the need for the presence of a specialized technician/doctor, thus increasing the cost and accordingly decreasing the wide applicability. Finally, a reported drawback in ECG wearable devices that do not

record standard ECG derivations is their inability to distinguish some arrhythmias and ectopic beats [204].

In order to compensate with these limitations, especially during exercise, several wearable devices measuring cardiac parameters have been introduced. The general concept of operation of these small, robust and user-friendly devices is that they contain sensors which reliably monitor minor changes in the intensity of light from high intensity light emitting diodes (LEDs) that is transmitted through or reflected from the human tissues. Although they have obvious advantages over the classical ambulatory ECG recording, the fact that they use photoplethysmography (PPG), i.e. a different detecting approach raises the question of how much accurate and reliable are when compared to the gold standard ECG method [194].

It is important to realize that the basic difference between PPG and ECG is the captured signal per se: The electrical activity of the heart is depicted by ECG, whereas the PPG is a mechanical signal measuring the propagation of the peripheral pulse wave. Therefore, the time of propagation of the PP wave from the heart to the distal arterioles is called pulse transit time (PTT). It is a measure of the time that elapsed between the R-wave of QRS complex in the ECG and the arrival point to PPG device [56]. Several studies have shown that PTT seems to be a surrogate marker of ANS in parallel to HRV [205] and that PTT is dependent on the properties of the pulse wave velocity, the vascular path from the heart to the location of the detector and is negatively correlated with blood pressure, arterial stiffness and age [202].

However, several parameters have to be considered when interpreting PPG measurements. These include:

1. *Motion artifact*: Special attention must be exercised during data PPG acquisition to eliminate motion-induced artifacts [202]. The contact force between the site and the sensor should be considered as PPG is vulnerable to such type of artifacts. However, despite the importance of this factor, we found only one study in 16 male ischemic patients measuring the accuracy of a

smart phone derived pulse rate versus ECG, where an excellent correlation was found in rest, which was slightly deteriorated during exercise [206].

2. *Respiration*: Since respiration alters the intrathoracic pressure and causes blood flow variations in both the veins the DC component of the PPG waveform shows minor changes with respiration [207]. Thus, it has been shown that the short-term variability (RMSSD, SD1, and HF) and LF/HF agreement between PRV and HRV deteriorates as a result of the vulnerability to respiratory changes [208].

3. *Age, gender and environmental factors*: Normal HRV values for various age and gender groups are still not available in the literature. However, it is well known that the elderly has increased arterial stiffness which leads to faster pulse transmission in the periphery and thus pulse transit time (PTT) differences observed between HRV and PRV could be attributed to aging [34].

   Regarding gender influence, in our review, just one study showed that measuring HRV at rest was age and gender dependent, the correlation with ECG being lower in women than men and further decreasing in older women [204]. As there is no strong evidence provided so far if age and/or gender can play a role in the studied correlations, further studies are needed to investigate these two variables on different populations while using appropriate sampling and prospective study design with a longitudinal follow-up [209].

   Environmental factor effects such as temperature was investigated in one of the studies. This study concluded that ambient temperature could induce a difference in the short-term variables that reflect the parasympathetic activity between HRV and PRV [210].

4. *Software analysis*: Some proprietary software systems for collection and analysis of RV data exist like the PPT5 or the ithlete™ software application [211, 212, 213] or freely available software may be used e.g. Kubios (http://www.kubios.com) [214]. According to guidelines [31], manual editing should be preferred instead of automated data analysis as automatic filters are

known to be unreliable and may potentially introduce errors. In our review only six studies used automated analysis only whether the rest used both manual and automated analysis.

5. *Statistical analysis*: The Bland–Altman plot must be used to compare the agreement among a new measurement technique with a gold standard, as even a gold standard does not imply to be without error. This plot allows the identification of any systematic difference between the measurements [215]. In our review only 17 studies used this technique, while four studies did not apply the Bland–Altman analysis and therefore only the correlation, but not agreement between the two methods, could be determined from these publications.

6. *Sampling rate*: The sampling rate is a matter of difference between the two approaches. Sampling rate of PPG is usually 20 Hz much less than that of ECG which is 125 to 250 Hz. This obviously implies weaker ability of the PPG devices for events detection [216].

In summary, Georgiou et al in a systematic review revealed that HR and/or HRV wearable devices, especially those using PPG, may provide a promising alternative solution for measuring HRV. However, it is evident that more robust studies in non-stationary conditions are needed with appropriate methodology in terms of number of subjects involved, acquisition and analysis techniques implied, before being able to recommend any of the commercially available devices. Therefore, so far wearable devices can only be used as a surrogate for HRV at resting or mild exercise conditions, as their accuracy fades out with increasing exercise load [194].

The first company offering a wearable EDA measuring device was Neumitra (https://www.neumitra.com) but it is not anymore available. The last years, a couple of more sophisticated wearable devices are commercially available, which except HR and Inter-Beat Interval (IBI), they concomitantly measure several other parameters such as EDA, acceleration and body temperature [194]. Thus, recently, a Dutch company (http://www.zephyr-

technology.nl/en/product/71/zephyr-bioharness.html) introduced the Zephyr BioHarness^{TM}3, which enables the capture and transmission of comprehensive physiological data (ECG, heart rate & heart rate RR as well as breathing data, IR skin temperature, posture, activity, skin conductance).

Another company (https://www.empatica.com/) introduced the E4 wristband which is a wearable wireless device designed for continuous, real-time data acquisition. The device has a photoplethysmography sensor (measures Blood Volume Pulse (BVP), from which heart rate, HRV, and other cardiovascular features may be derived), a 3-axis accelerometer, a GSR sensor, and an infrared thermopile.

Unfortunately, for obvious reasons, some locations are unsuitable for performing measurements on a surgeon while operating on an individual or on a trainee performing a simulation task [2]. Therefore, EDA devices are usually placed on the wrist or ankle. Usually, the device contains two stainless steel or silver electrodes that measure electrical conductance at sample frequency of 4Hz (Empatica S.R.L., Milano, Italy). During the measurements, care must be taken to avoid sweating artefacts due to changes in ambient temperature [1]. It is important to note that HRV, EDA and thermal activity all require the use of specific equipment that must be worn by the subject. These contact devices may have several limitations with respect to data collection and may involve some discomfort, thus compromising performance [56].

In our study we used such a wearable device, the E4 wristband (https://www.empatica.com/) which is a wearable wireless device designed for continuous, real-time data acquisition in daily life (Figure 2).

Figure 2. The Empatica4 wristband used in our study

The device has a photoplethysmography Sensor (measures Blood Volume Pulse (BVP), from which HR, HRV, and other cardiovascular features may be derived), a 3-axis accelerometer (captures motion-based activity), an electrodermal activity (EDA) sensor (used to measure sympathetic nervous system arousal and to derive features related to stress, engagement, and excitement) and an infrared thermopile (reads peripheral skin temperature).

Part B: Our Contribution

Aim

Because of the lack of a “gold standard” method to assess surgeons’ mental stress, several noninvasive measuring methods have been proposed with conflicting results as most studies used only one or two of them.

The first objective of this study was to assess the feasibility of a new, easy to wear watch-sized device to noninvasively measure cardiac stress parameters such as heart rate and heart rate variability which are usually measured by means of a Holter monitor. Additionally, this device can also measure EDA thus enhancing stress estimation with one additional parameter.

The second objective of our work was to concomitantly measure the responses of seven noninvasive stress indices and compare them to the video score achieved by novices in a reproducible stressful simulation environment. Namely, we used one subjective method (STAI) plus the three abovementioned ones derived from the wrist-worn device (HR, IBI, EDA) as well as the salivary levels of cortisol, alpha-amylase and IgA.

Practically, it might be of value if such an easy to wear device combined with a noninvasive saliva collection might be used to objectively evaluate stress and avoid the use of other motion restricting, uncomfortable devices and invasive methods in our simulation setting as it might open a potentially practical window to conveniently train surgeons to manage their stress using a surgical simulation training system.

Material and Methods

The study protocol (ISRCTN No 10985546) was approved by the Institutional Ethical Committee and all participants gave informed consent.

Subjects

Twenty-one male novice trainees aged 23 - 26 years with body mass index (BMI) ranging from 18.5 to 24.9 were included in the study. All of them were medical students in their last year or PGY1 without any previous simulation experience. Table 3 shows the gender and age of the subjects included in the study.

Table 3. Gender and age of the 21 subjects

ID	Gender	Age	ID	Gender	Age
1	M	24	15	M	23
2	M	23	16	M	23
3	M	23	17	M	23
4	M	24	18	M	26
5	M	23	19	M	23
6	M	25	20	M	24
7	M	26	21	M	25
8	M	26			
9	M	23			
10	M	24			
11	M	23			
12	M	25			
13	M	23			
14	M	24			

Experimental Protocol

Prior to the simulation task, all participants completed a baseline questionnaire, which included items on demographic characteristics, prior laparoscopic or simulator experience, and conformity with the inclusion criteria listed below.

The inclusion criteria were: a) no current prescribed medication or nonprescription drugs, b) no signs of cold or upper respiratory tract infection, c) nonsmokers, and d) refrain from drinking alcohol, coffee, and exercising for 12 hours before testing and from eating and brushing teeth for one hour prior to the experiment. Experiments were set between 14:30 and 16:30 to minimize circadian effects [217].

They also completed the short, six-item Spielberger State-Trait Anxiety Inventory (STAI). An orientation phase was then implemented (approximately 15 minutes) during which they were introduced to the simulator and briefed on the tasks to follow.

Thereafter they wore the Empatica E4 wristband (E4WB) (EmpaticaS.r.l, Italy) on their nondominant hand. Additionally, all subjects wore an ambulatory Holter ECG rhythm monitoring (HM) and electrodes were positioned in predetermined thorax positions (ELA Medical - Syneflash MMC-24hour Rhythm - 3 channel lead- Synescope ELA Medica, France). Throughout the whole experiment, the subjects wore both the E4WB and the HM. An ambient temperature of 27°C was kept in the simulation room in order to avoid any sweating artifacts.

A baseline recording phase (BL phase) of 10 minutes was initiated with the subjects engaged in leisurely reading and a saliva specimen was collected using an unstimulated passive drool technique (Base line-BL phase).

Immediately after, the simulation exercise started (E phase) in which the subjects were trained on a basic skills module (Lap mentor, 3DSystems) for 9 minutes.

Immediately after, another saliva sample was collected from all participants in a similar manner as before which was followed by a 6 minutes recovery period (R phase). During this phase all subjects repeated the short, six-item STAI.

Each phase was tagged by triggering concomitantly the markers on both devices. E4WB recorded photoplethysmography (PPG) BVP data were used to estimate HR and inter-beat interval duration (IBI). Additionally, EDA, 3-axishand motion activity (Acc), and skin temperature (ST) data were recorded from the E4WB. Furthermore, R-R interval data obtained from the Holter monitor were extracted in order to measure HR and calculate HRV indices.

The whole experiment was videotaped and stored for further analysis. An ambient temperature of 27^{o}C was kept in the simulation room to avoid any sweating artifacts.

Saliva samples where refrigerated and, subsequently, frozen at -20^{0}C within four hours of collection until assayed.

Data analysis

When each session ended, the E4WB was connected to a PC and its internal storage data were uploaded to a secure cloud web site. Through this site the sessions' data can be visualized and exported as CSV files for analysis of each one of the recorded stress-related parameters: HR, IBI duration, and EDA.

In order to compare the R-R with the IBI data derived respectively from the two devices, some pre-processing was necessary before any further analysis takes place. This was necessary due to two reasons: a) the E4WB starts recording automatically 40 sec after pushing the mark button, while the Holter starts to record immediately after pressing the mark, and b) E4WB heart beats are recorded in respect to initial time which in turn is expressed in UTC time. Therefore, all the time points had to be converted to absolute local time.

Additionally, since both devices did not share the same time settings from a reliable third-party source, their recorded data needed synchronization.
In order to avoid multi factorial external stress bias, the percentage differences between base line and exercise of all parameters measured were derived for each subject. Additionally, pre- and post-exercise STAI scoring as well as their differences was calculated.

The Bland-Altman method was used to plot the difference for each subject against HRV and IBI values in order to compare Holter derived HRV versus PPG derived IBI.

All sets of synchronized intervals data extracted from both Holter and E4WB were analyzed by using the Kubios HRV® software. Six Indices of HRV were computed in the frequency domain (high frequency [HF]; low frequency [LF] and LF/HF ratio) for all phases of the study.

Commercially available kits (SME-1-3002 Salivary Cortisol Research ELISA kit), a-amylase (SME-1-1902 Alpha-amylase Kinetic Reaction Kit Research), and sIgA (SME-1-1602 Salivary Secretory IgA Research ELISA kit) were used. A saliva specimen's concentrations were determined following the manufacturer's procedures (www.salimetrics.com). Minimum, maximum and mean normal values according to the manufacturer are listed in the Table below:

Table 4. Normal values of measured saliva parameters.

			Minimum	Maximum	Mean
Amylase (U/ml)			3,1	423,1	92,4
IgA (µg/ml)			93,2	974,03	379,39
Cortisol (µg/dl)	males	am	0,112	0,743	
		pm	0	0,308	
	females	am	0,272	1,348	
		pm	0	0,359	

To avoid multifactorial external stress bias, for each subject, the percentage difference value ($Value_{diff}$) of each parameter was calculated from their reciprocal baseline ($Value_{pre}$) and post-exercise ($Value_{post}$) values, where $Value_{diff}=100\cdot(Value_{post}-Value_{pre})/Value_{pre}$. Therefore, six predictor variables were derived.

Video scoring

A blinded expert reviewer assessed the videos using a 7-point procedure-specific checklist and summed the reciprocal scores to a total score achieved (using a 7-point procedure-specific checklist and summed the reciprocal scores to a total score achieved (see Table 5). As can be deducted from this Table, the maximum total score is 21.

Table 5. Video Scoring system

Score	Knot tying	Continuous stitches	Dominant handle movement (needle holder)		Non-Dominant handle movement		Task completion time (tct) in minutes
			Instrument outside of view	Needle drops	Instrument outside of view	Needle drops	
0	No tie	No stitch	>5 times	>5 times	>5 times	>5 times	Failure (tct> 15)
1	1 tie	1 stitch	3 times	3 times	3 times	3 times	13 ≤ tct ≤ 15
2	2 ties	2 stitches	1 time	1 time	1 time	1 time	10 ≤ tct< 13
3	3 ties	3 stitches	None	None	None	None	tct< 10

Results

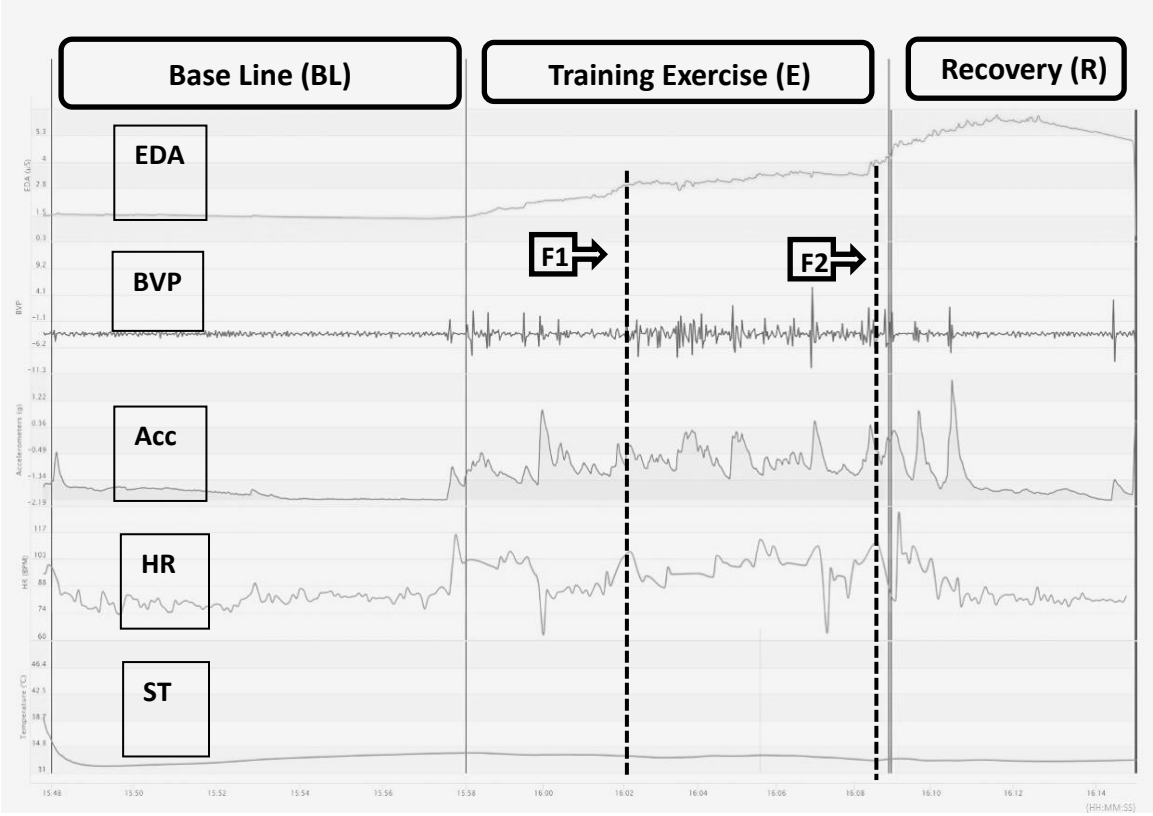
None of the subjects reported any problems / discomfort related to the wearing equipment.

A) Exploiting the best stress detector among the E4WB parameters.

A1. Descriptive observation

A representative output of the E4WB parameters of one subject is shown in Figure 3.

Figure 3. A representative output of the Empatica E4 parameters of a subject. F1 & F2 represent detected and time stamped failures.



EDA: Electrodermal activity, BVP: Photoplethysmography, Acc: 3-axis hand motion activity, HR: Heart rate, ST: Skin temperature.

As it can be seen, EDA started to increase as soon as the E phase starts and remained elevated even during the R phase. EDA was further increased right after each failure by more than 2% ($p<0.001$) per second (2.5% at F1 and 2.1% at F2). IBI increased about 0.9% ($p<0.10$) per second (at F1 and F2). BVP did not seem to have a trend during the E phase. Acc and HR increased during E phase but with many fluctuations, while ST remained almost stable during the whole experiment.

A2. Failures detection

Large intra-individual differences were observed in baseline values of all parameters studied. Therefore, in order to examine which parameter is the best indicator for failure detection, the differences between BL and E phase of all parameters were analyzed in respect of failures detection in each subject.

During the E phase a total of 32 failures occurred and were recorded. For each parameter, the mean percentage change and SD values per second along with their corresponding p values are presented in Table 6. As it can be seen, only BVP and Acc did not reach statistical significance.

Table 6. Failure detection: Mean percentage change (per second) and its SD of each recorded parameter, along with reciprocal P values.

	Mean (%) ± SD (%)(sec^{-1})	P value
EDA	2.21 ± 0.24	<0.001
IBI	0.917 ± 0.026	<0.10
ST	-0.007 ± 0.004	<0.001
HR	0.032 ± 0.096	<0.001
BVP	1.005 ± 0.178	>0.60
Acc	-0.015 ± 0.003	>0.90
R-R interval (Holter)	0.910 ± 0.144	<0.001

A3. Individual Video score, STAI and E4WB parameters

Table 7 shows the video score, the STAI value during baseline ($STAI_{pre}$), the STAI value after exercise ($STAI_{post}$), and the percentage difference between the two STAI values ($STAI_{diff}$).

Table 7. Video score, $STAI_{pre}$, $STAI_{post}$ and $STAI_{diff}$ of the 21 subjects.

ID	video score	$STAI_{pre}$	$STAI_{post}$	$STAI_{diff}$
1	10	36.7	63.3	72.48
2	17	36.7	56.7	54.50
3	9	40	66.7	66.75
4	13	40	66.7	66.75
5	12	30	50	66.67
6	9	33.3	56.6	69.97
7	10	40	66.7	66.75
8	14	26.7	46.7	74.91
9	20	36.7	56.7	54.50
10	12	40	63.3	58.25
11	11	36.7	60	63.49
12	13	30	53.3	77.67
13	18	36.7	60	63.49
14	10	33.3	56.7	70.27
15	15	30	50	66.67
16	13	33.3	56.7	70.27
17	19	33.3	53.3	60.06
18	18	36.7	56.7	54.50
19	17	30	50	66.67
20	18	33.3	53.3	60.06
21	10	36.7	63.3	72.48

$STAI_{diff}=100\cdot (STAI_{post}-STAI_{pre})/STAI_{pre}$

The range of video score achieved by the 21 subjects was from 9 to 20, (mean= 13.7, SD= ± 3.6).

Table 8 shows the individual values of each of the measured parameters from the E4WB before($_{pre}$), and after exercise ($_{post}$), as well as their percentage differences ($_{diff}$).

Table 8. EDA_{pre}, EDA_{post}, EDA_{diff}, HR_{pre}, HR_{post}, HR_{diff}, IBI_{pre}, IBI_{post} and IBI_{diff} of the 21 subjects.

ID	EDA_{pre} (μS)	EDA_{post} (μS)	EDA_{diff}	HR_{pre} (min^{-1})	HR_{post} (min^{-1})	HR_{diff}	IBI_{pre}(s)	IBI_{post}(s)	IBI_{diff}
1	0.95	15.50	1531.58	91.47	93.64	2.37	0.62	0.60	-3.23
2	0.9	6.00	566.67	75.00	70.00	-6.67	0.65	0.70	7.69
3	0.95	10.00	952.63	87.67	94.37	7.64	0.64	0.58	-9.38
4	3	12.00	300.00	108.85	98.10	-9.88	0.55	0.61	10.91
5	1.1	7.50	581.82	89.79	82.30	-8.34	0.65	0.70	7.69
6	0.9	10.50	1066.67	85.00	89.00	4.71	0.65	0.61	-6.15
7	1.2	11.00	816.67	81.00	91.70	13.21	0.59	0.53	-10.17
8	0.17	2.00	1076.47	76.00	66.00	-13.16	0.70	0.80	14.29
9	1	5.00	400.00	80.00	70.00	-12.50	0.60	0.68	13.33
10	1	8.00	700.00	75.00	82.14	9.52	0.62	0.56	-9.68
11	0.9	14.00	1455.56	84.69	82.44	-2.66	0.78	0.80	2.56
12	0.6	8.50	1316.67	77.00	86.00	11.69	0.72	0.63	-12.50
13	2	15.00	650.00	96.80	92.30	-4.65	0.56	0.60	7.14
14	1	13.00	1200.00	67.30	71.31	5.96	0.65	0.61	-6.15
15	0.6	5.50	816.67	68.00	70.00	2.94	0.77	0.74	-3.90
16	0.55	3.00	445.45	76.00	81.00	6.58	0.80	0.75	-6.25
17	0.4	1.50	275.00	99.15	87.90	-11.35	0.60	0.66	10.00
18	1.3	6.00	361.54	100.50	95.90	-4.58	0.53	0.56	6.34
19	1.3	7.50	476.92	80.60	78.40	-2.73	0.63	0.65	3.17
20	0.45	3.50	677.78	100.40	85.50	-14.84	0.60	0.68	13.33
21	0.8	9.00	1025.00	89.00	84.00	-5.62	0.67	0.72	7.46

$EDA_{diff}=100\cdot (EDA_{post}-EDA_{pre})/ EDA_{pre}$

$HR_{diff}=100\cdot (HR_{post}-HR_{pre})/ HR_{pre}$

$IBI_{diff}=100\cdot (IBI_{post}-IBI_{pre})/ IBI_{pre}$

The correlations of Video score with STAI as well as each one of the wristband derived parameters (namely EDA, HR, and IBI) expressed as their percentage differences ($_{diff}$) is shown at the following figures (Figures 4, 5, 6 and 7).

Figure 4. The correlation of Video score with $STAI_{diff}$

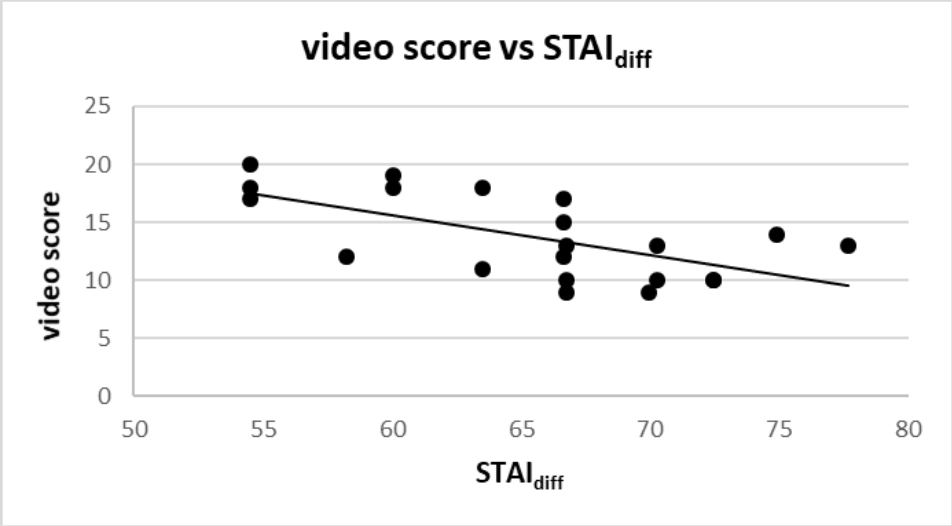


Figure 5. The correlation of Video score with EDA_{diff}

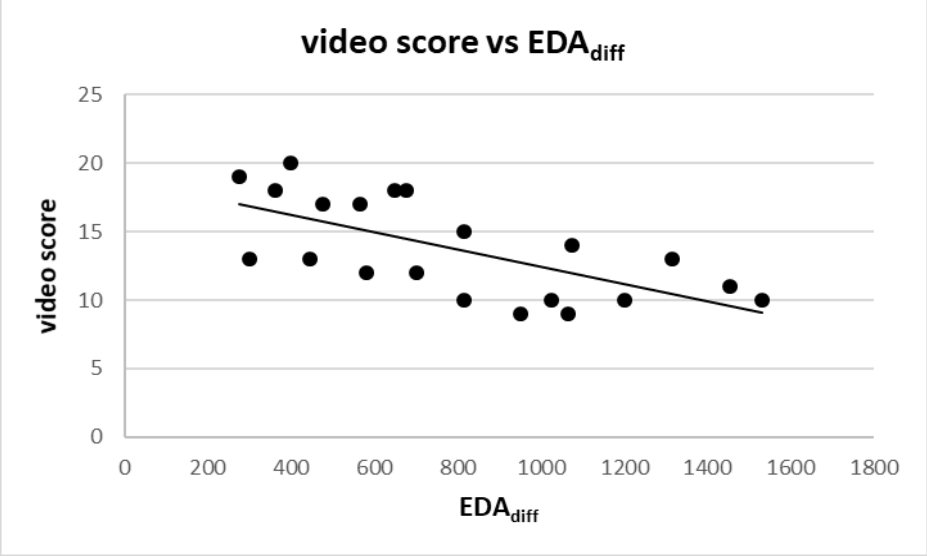


Figure 6. The correlation of Video score with HR_{diff}

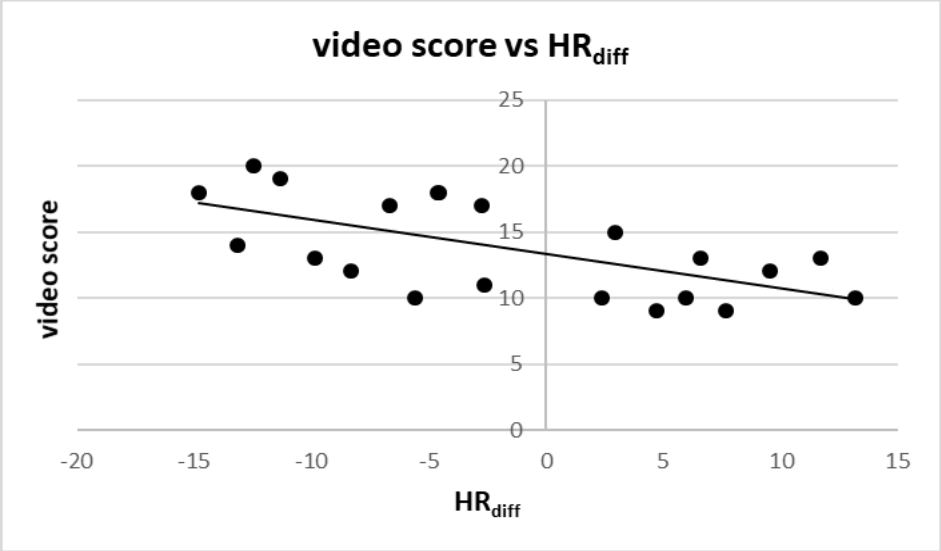
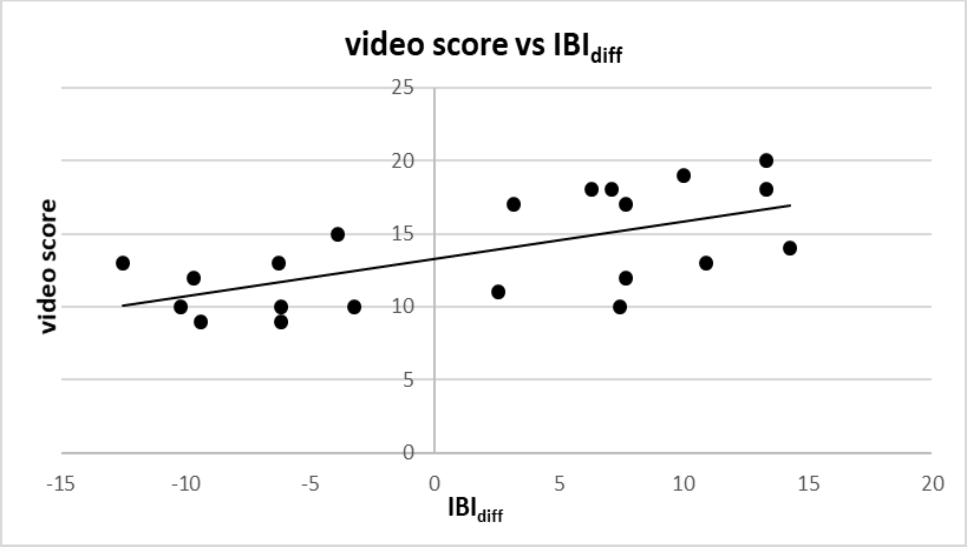


Figure 7. The correlation of Video score with IBI_{diff}

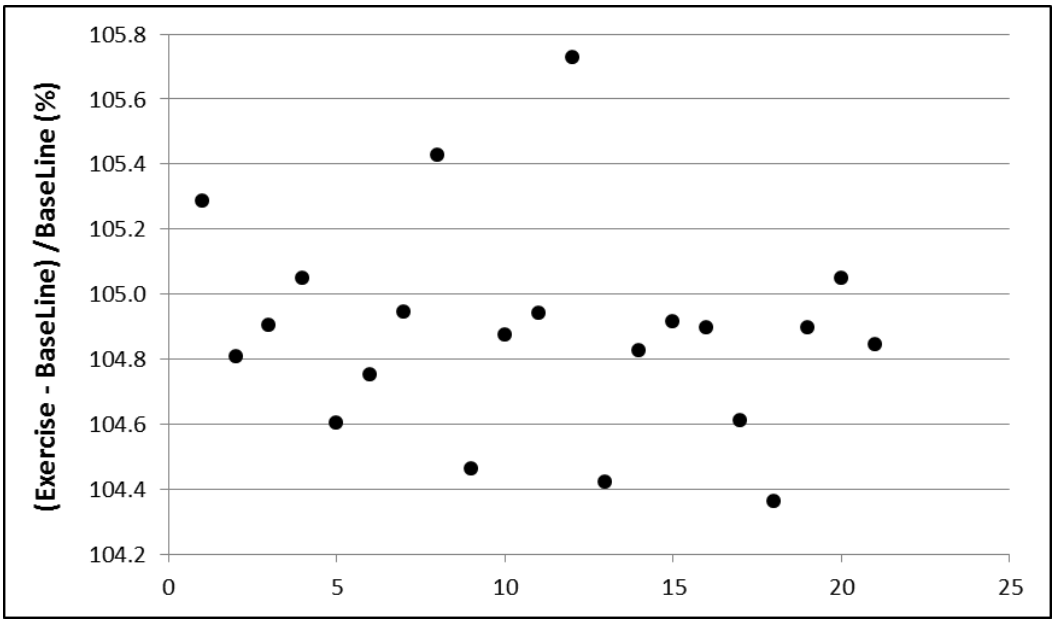


A4. E4WB parameters response during E phase

Of all parameters studied, only the mean EDA levels for all subjects combined, significantly increased from BL phase to the E phase (from 1.47 to 3.00 μS; p<0.001), and further increased during the R phase as compared with BL levels (from 1.47 to 5.62 μS; p<0.001).

Figure 8 shows the percentage increase of the mean value of EDA during E phase to the one during BL phase for each individual.

Figure 8. Average percentage increase of the mean value of EDA (Y axis) during Training Exercise phase to the one during Base Line phase for each individual (X axis).



EDA: Electrodermal activity.

It can be noticed that stress during E phase provoked about 105% increase in the EDA values. IBI change was smaller as it changed from 730 ms during BL phase to 607 ms during E phase (16% decline) and finally to 716 ms during R phase (p<0.001). Similar non-significant variations were observed in HR

irrespectively if it was derived from Holter or E4WB. In all cases during the E phase, the Acc values where within the measurement limits since none exerted the force of 2 g. Temperature data didn't change significantly in all subjects at any phase ($p<.001$).

A5. Correlation of E4WB parameters with STAI

A positive post- to pre-exercise STAI scoring difference (STAIdif) was seen in 15 (71%) subjects while in the remaining six (29%) it was negative. Percentage differences of the average values of EDA (EDAdif%), IBI (IBIdif%), Holter R-R (HolterRRdif%) during exercise and baseline phase as well as STAI (STAIdif%) differences between post- and pre- exercise phase were calculated.

Table 9 summarizes the correlation coefficient r between all the above parameters. As can be seen from this table, the EDA difference exhibited the best correlation with STAI difference, while IBI and Holter ones showed similar r values.

Table 9. The correlation coefficient r between STAIdif%, HolterRRdif%, IBIdif% and EDAdif%.

	STAIdif%	HolterRRdif%	IBIdif%	EDAdif%
STAIdif%	1	-0.78	-0.74	0.83
HolterRRdif%	-0.78	1	0.87	-0.71
IBIdif%	-0.74	0.87	1	-0.67
EDAdif%	0.83	-0.71	-0.67	1

The sensitivity and specificity of Holter's R-R, IBI and EDA are seen in Table 10. As can be seen from this table, the EDA difference exhibited the best sensitivity (equal to Holter R-R differences) and the best specificity.

Table 10. The sensitivity and specificity of Holter's R-R, IBI and EDA.

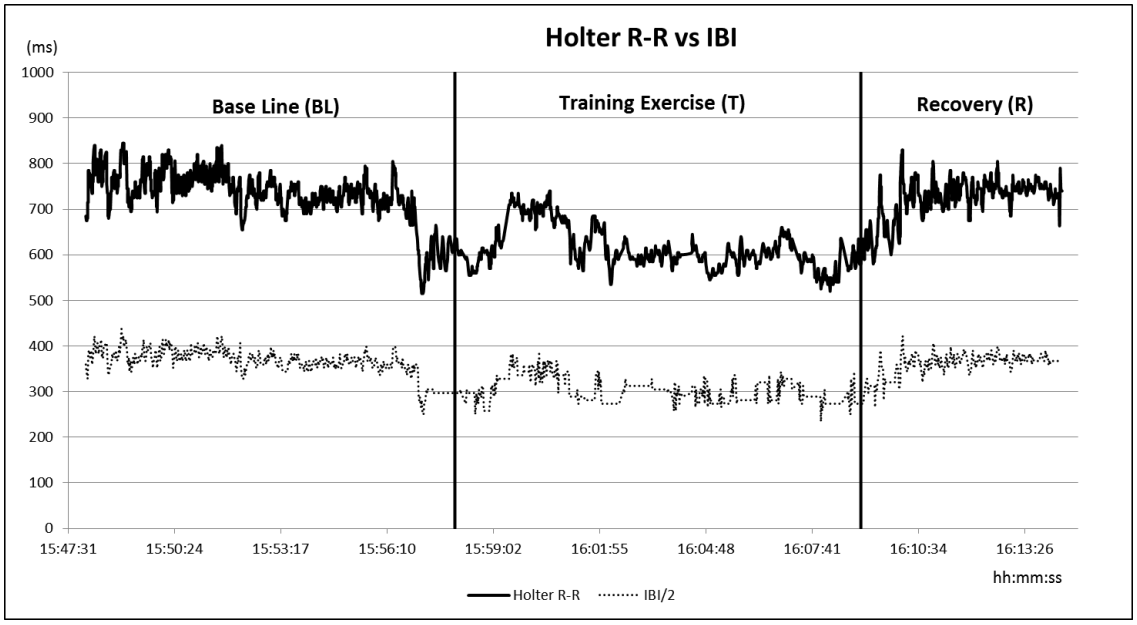
	Sensitivity	**Specificity**
Holter R-R	67%	93%
IBI	67%	87%
EDA	83%	93%

B) Ambulatory Holter versus E4WB wristband HR detection.

B1. Descriptive comparison

A comparison was made to get an understanding of how well the E4WB sensor performs compared to Holter's HR. Therefore, in a representative subject the synchronized signals from the E4WB and Holter were superimposed and compared (Figure 9).

Figure 9. Synchronized signals from the rhythm Holter (R-R, solid line) and from the E4 wristband (IBI/2, dashed line). IBI values were divided by 2 to avoid the graphs superimposition and better graph presentation.



B2. Statistical analysis

The whole session of each subject provided a range of 3625-3830 data points for R-R and IBI estimation (1500-1690 for BL, 1210-1400 for E, 740-920 for R respectively).

The following four Tables (Tables 11, 12, 13, and 14) present Holter R-R, IBI, Holter-IBI and their percentage difference of BaseLine, Exercise, Recovery and together All sessions of the 21 subjects.

Table 11. Holter R-R, IBI, Holter-IBI and their percentage difference of BaseLine session of the 21 subjects.

ID	Holter R-R (ms)	IBI (ms)	Holter-IBI (ms)	(Holter-IBI)/Holter (%)
1	731.933	731.908	0.025	0.003
2	732.533	734.200	-1.667	-0.228
3	732.200	731.700	0.500	0.068
4	731.933	732.117	-0.184	-0.025
5	729.200	732.742	-3.542	-0.486
6	729.867	730.867	-1.000	-0.137
7	729.133	727.117	2.017	0.277
8	727.667	728.783	-1.117	-0.153
9	727.200	727.742	-0.542	-0.074
10	728.133	728.367	-0.233	-0.032
11	727.200	729.825	-2.625	-0.361
12	728.333	728.783	-0.450	-0.062
13	731.333	728.367	2.967	0.406
14	733.733	729.408	4.325	0.589
15	733.600	730.658	2.942	0.401
16	731.200	733.575	-2.375	-0.325
17	730.200	732.117	-1.917	-0.263
18	730.800	729.825	0.975	0.133
19	731.733	727.950	3.783	0.517
20	728.467	730.242	-1.775	-0.244
21	729.733	731.700	-1.967	-0.270

Table 12. Holter R-R, IBI, Holter-IBI and their percentage difference of Exercise session of the 21 subjects.

ID	Holter R-R (ms)	IBI (ms)	Holter-IBI (ms)	(Holter-IBI)/Holter (%)
1	613.548	607.891	5.658	0.922
2	612.742	604.866	7.876	1.285
3	610.081	605.118	4.962	0.813
4	611.048	606.883	4.166	0.682
5	612.016	611.923	0.093	0.015
6	610.645	608.647	1.998	0.327
7	610.484	606.883	3.601	0.590
8	611.048	602.346	8.702	1.424
9	612.823	605.875	6.948	1.134
10	615.323	609.151	6.172	1.003
11	615.806	605.118	10.688	1.736
12	616.935	605.118	11.817	1.915
13	615.645	602.094	13.551	2.201
14	614.516	605.370	9.146	1.488
15	612.419	607.639	4.781	0.781
16	613.226	608.395	4.831	0.788
17	614.274	607.135	7.140	1.162
18	615.565	605.118	10.446	1.697
19	614.919	606.883	8.037	1.307
20	614.032	610.663	3.369	0.549
21	612.419	608.395	4.025	0.657

Table 13. Holter R-R, IBI, Holter-IBI and their percentage difference of Recovery session of the 21 subjects.

ID	Holter R-R (ms)	IBI (ms)	Holter-IBI (ms)	(Holter-IBI)/Holter (%)
1	711.515	711.680	-0.165	-0.023
2	714.394	713.101	1.293	0.181
3	713.182	713.574	-0.393	-0.055
4	712.727	712.154	0.573	0.080
5	713.182	712.154	1.028	0.144
6	716.212	714.521	1.691	0.236
7	715.758	712.627	3.130	0.437
8	719.091	713.101	5.990	0.833
9	718.636	716.889	1.748	0.243
10	718.030	713.101	4.929	0.687
11	718.485	711.680	6.805	0.947
12	720.455	714.521	5.933	0.824
13	720.606	712.627	7.979	1.107
14	722.727	715.942	6.785	0.939
15	722.121	721.624	0.497	0.069
16	720.000	717.362	2.638	0.366
17	718.636	721.150	-2.514	-0.350
18	719.848	721.624	-1.775	-0.247
19	716.818	721.624	-4.806	-0.670
20	718.182	716.415	1.766	0.246
21	717.424	718.783	-1.359	-0.189

Table 14. Holter R-R, IBI, Holter-IBI and their percentage difference of all sessions of the 21 subjects.

ID	Holter R-R (ms)	IBI (ms)	Holter-IBI (ms)	(Holter-IBI)/Holter (%)
1	686.695	684.706	1.988	0.290
2	688.107	686.031	2.077	0.302
3	688.418	683.824	4.594	0.667
4	688.644	684.795	3.849	0.559
5	686.949	686.737	0.212	0.031
6	686.525	684.353	2.172	0.316
7	686.469	682.058	4.411	0.643
8	685.932	682.323	3.609	0.526
9	684.887	681.617	3.270	0.478
10	686.243	683.029	3.214	0.468
11	685.989	686.560	-0.572	-0.083
12	685.282	684.177	1.106	0.161
13	686.215	684.000	2.215	0.323
14	687.712	682.764	4.948	0.719
15	687.740	684.353	3.387	0.492
16	687.740	685.678	2.063	0.300
17	687.232	684.089	3.143	0.457
18	687.684	683.029	4.654	0.677
19	688.249	681.440	6.808	0.989
20	686.299	684.000	2.299	0.335
21	686.102	685.148	0.954	0.139

The plots regarding the correlation between Holter R-R and IBI data for each individual during the entire experiment as well as during each phase are depicted in Figures 10, 11, 12 and 13.

Figure 10. Holter R-R vs IBI in BaseLine

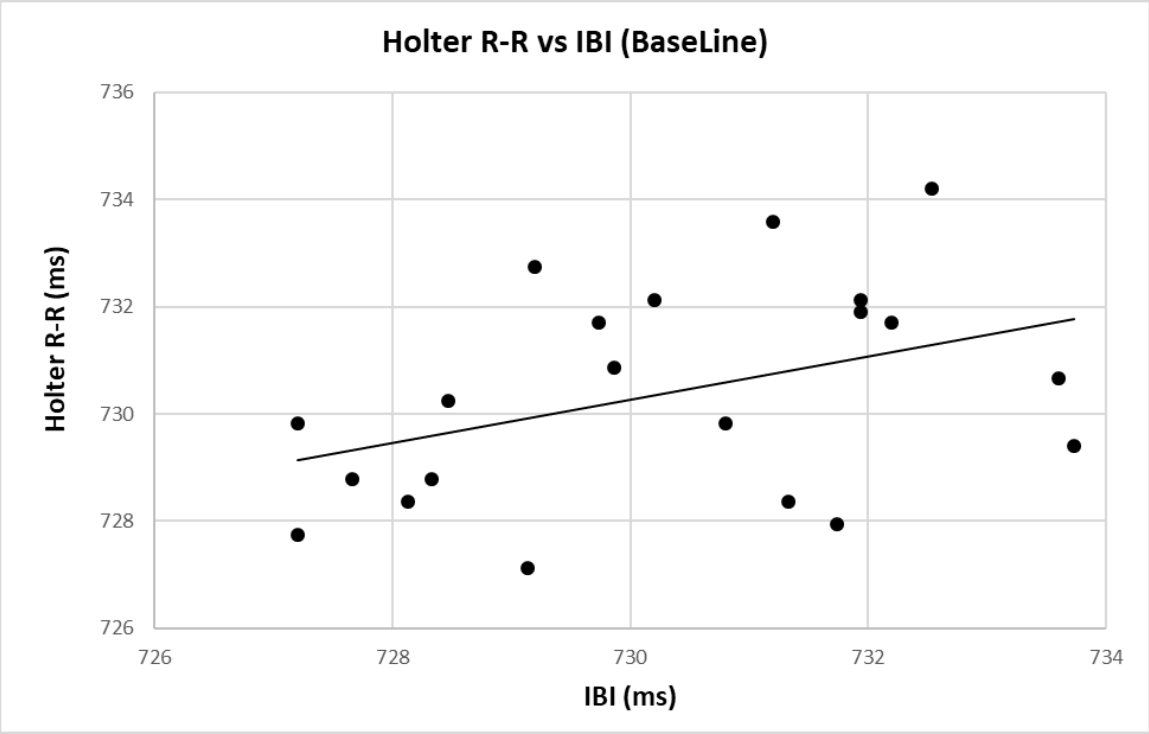


Figure 11. Holter R-R vs IBI during Exercise

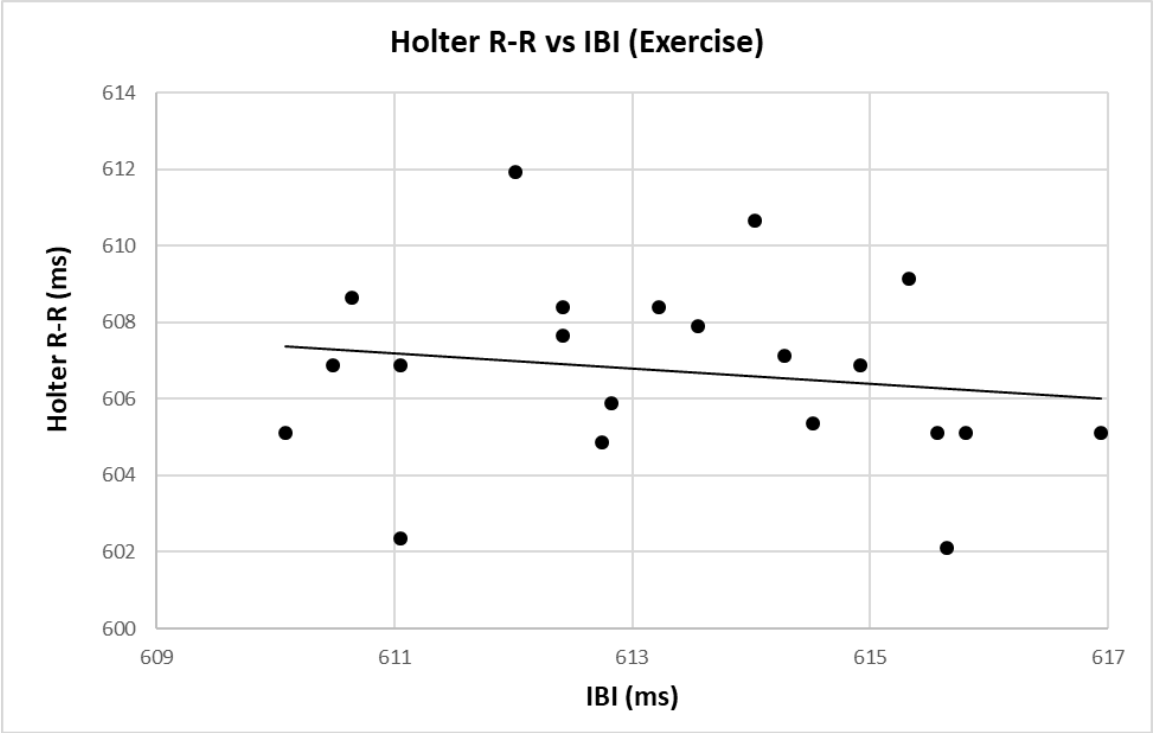


Figure 12. Holter R-R vs IBI during Recovery

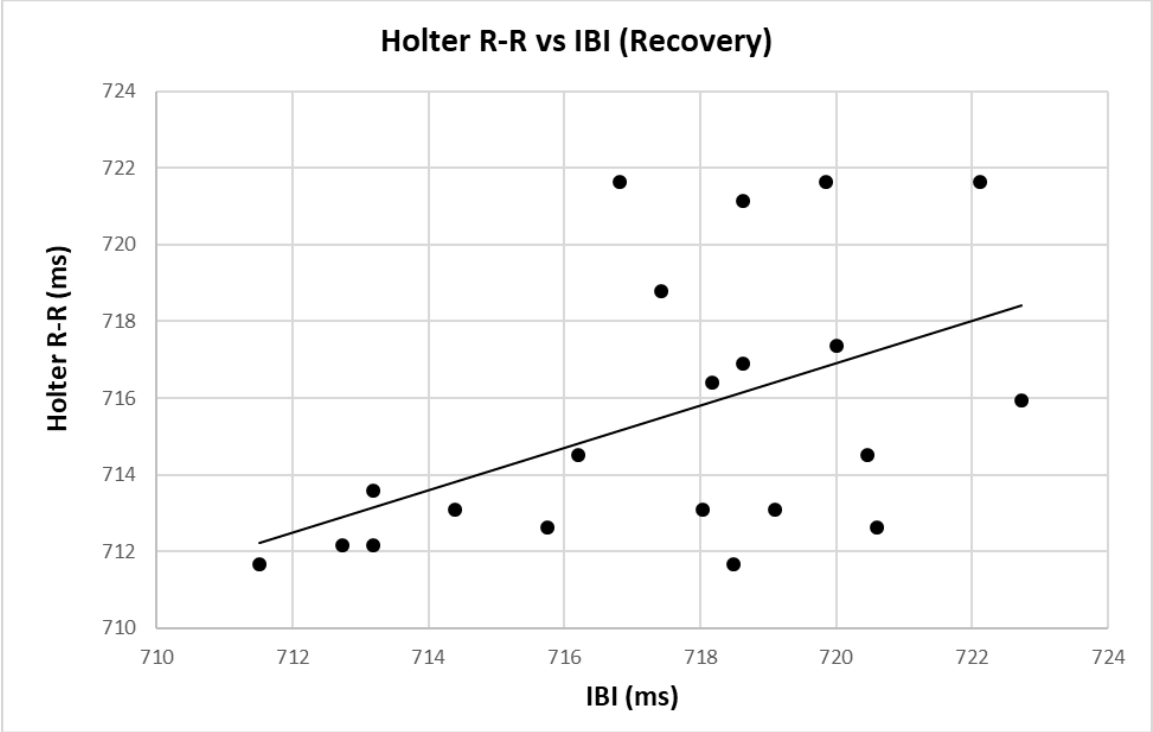
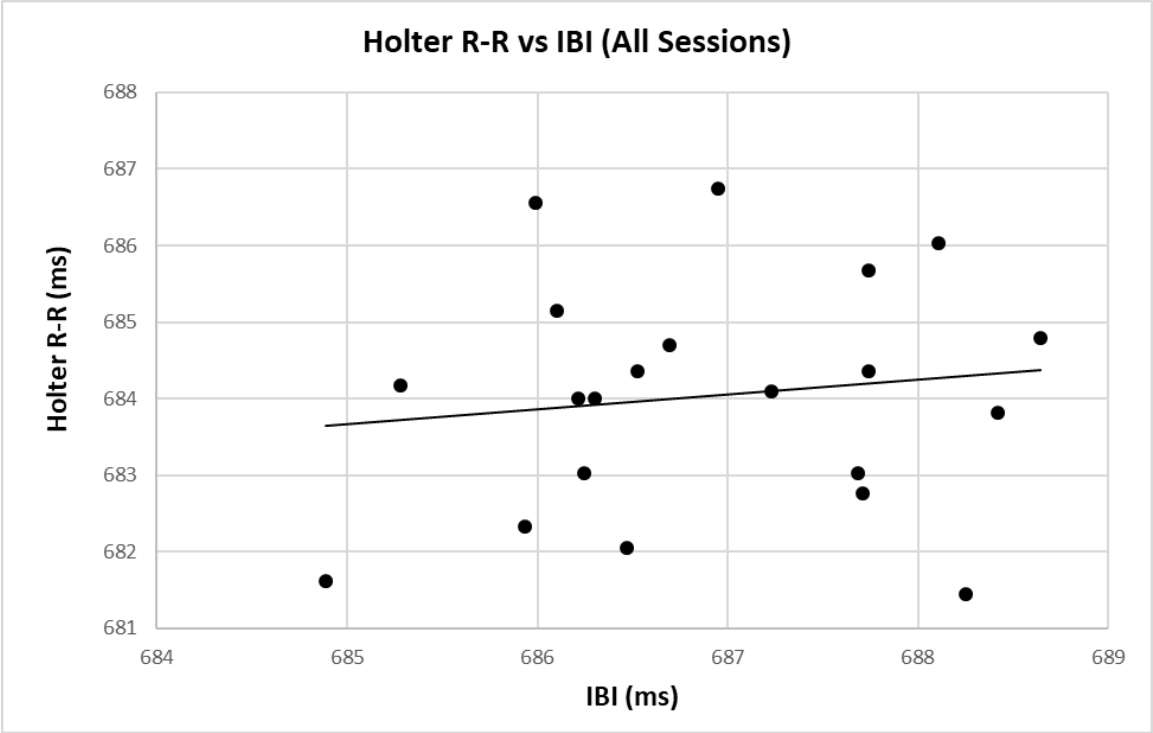


Figure 13. Holter R-R vs IBI during All sessions.



Correlation coefficient r was used to quantify their mean values and range are presented in Table 15.

Table 15. Mean values and range of the correlation coefficient (r) between R-R and IBI during the whole experiment as well as during each one of the three phases.

	Correlation (r) Mean value	Range
All Sessions	0.938	0.926-0.950
Base Line (BL)	0.947	0.921-0.963
Exercise (E)	0.749	0.671-0.808
Recovery (R)	0.900	0.803-0.955

It can be seen that the mean value of r during all sessions shows a high agreement between the two methods (r=0.938; p<.001), reaching its maximum at BL phase (r=0.947, p<.001)and being minimized at E phase (r=0.749, p<.001). Bland-Altman analysis was also used to compare the two methods. For each subject, percentage differences of the mean values [(Holter-E4WB)/Holter] were calculated for each phase as well as for the whole experiment. The descriptive statistics are presented in Table 16.

Table 16. Descriptive statistics based on the Bland-Altman analysis

	All Sessions	Base Line (BL)	Exercise (E)	Recovery (R)
Sample size	21	21	21	21
Arithmetic mean (%)	0.419	-0.013	1.070	0.276
Standard deviation	0.251	0.301	0.547	0.473
95% Cl for the mean	0.311 - 0.526	-0.141 - 0.116	0.836 - 1.304	-0.074 - 0.479
LoA (lower)	-0.073	-0.603	-0.002	-0.651
LoA (upper)	0.910	0.578	2.143	1.204

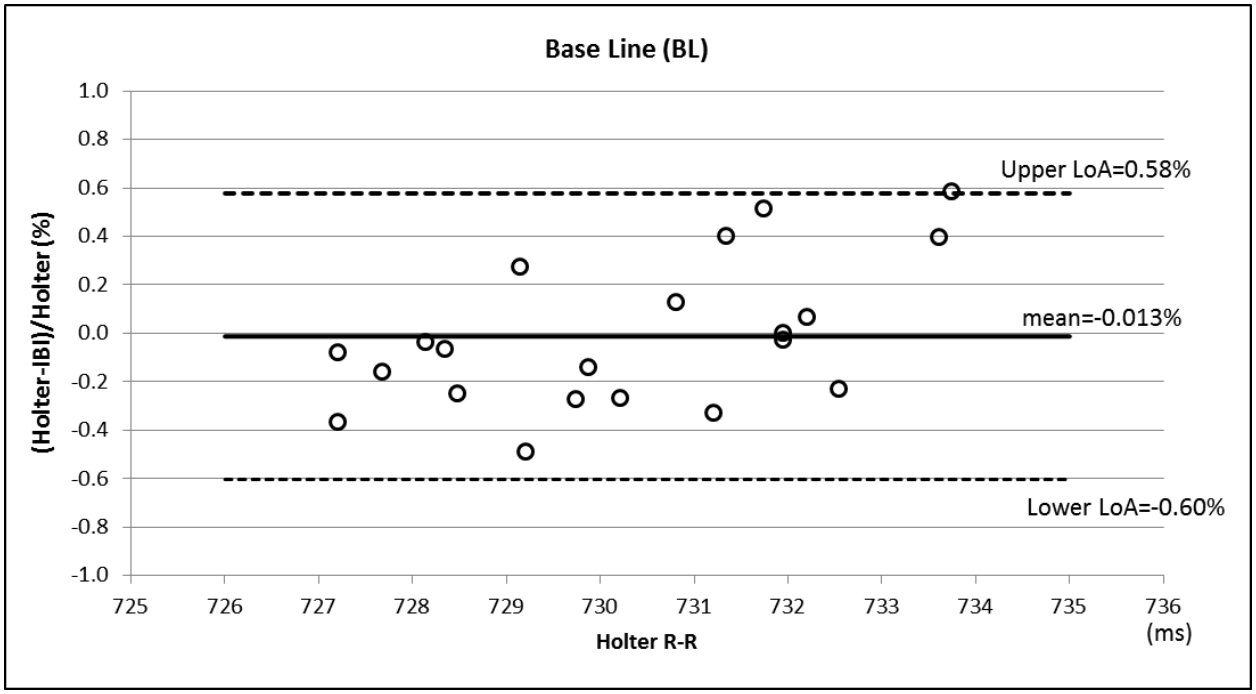
Cl: Confidence level. LoA: Level of Agreement.

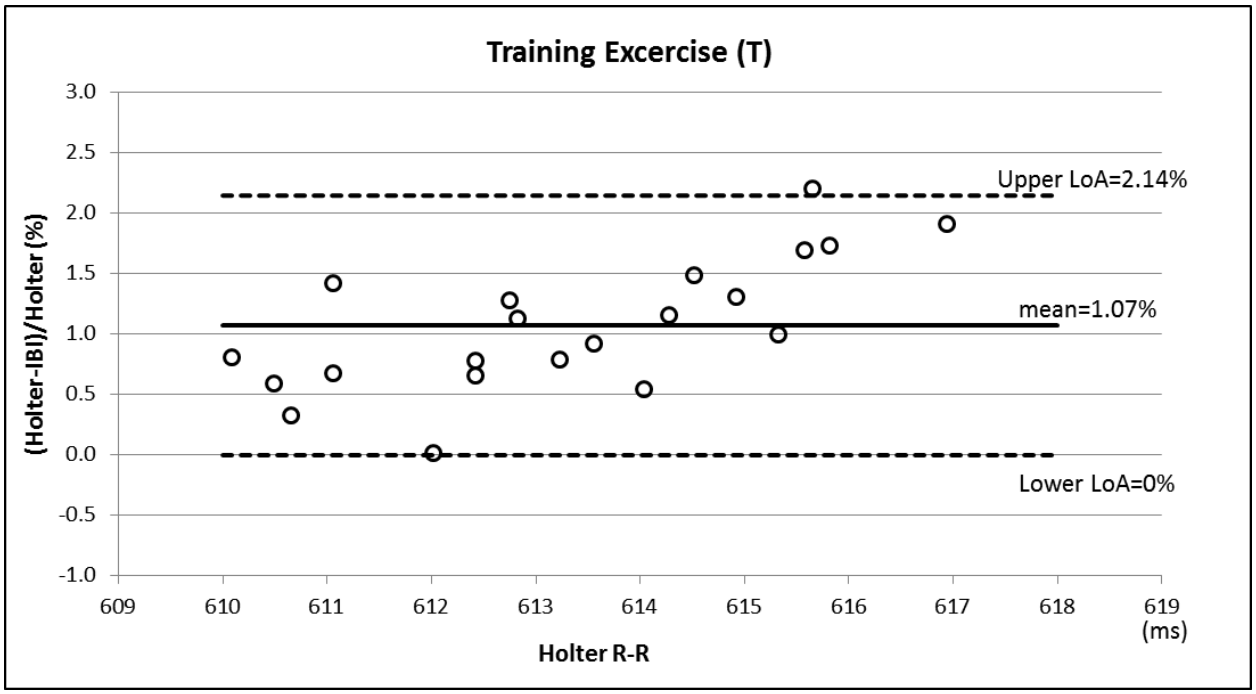
As can be seen from this table, the mean value of the difference is +0.419% during all phases suggesting a close agreement between these two methods. The mean bias is minimum during BL phase (-0.013%) and maximum during the E phase (+1.07%).

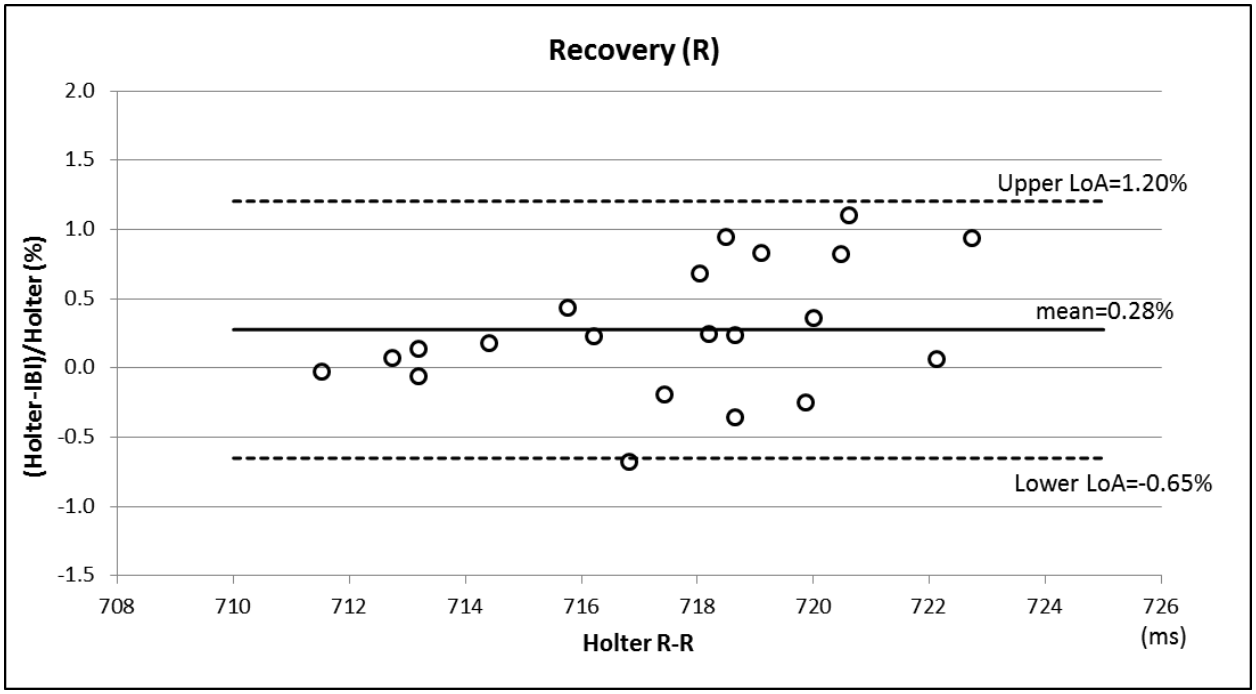
Figure 14 presents the Bland-Altman plots (for each phase and for the whole experiment) of the collected data from the 21 subjects. Y axis represents the percentage difference and X axis represents the value of the HM (as it is the reference method). Moreover, mean values and levels of agreement (LoA)

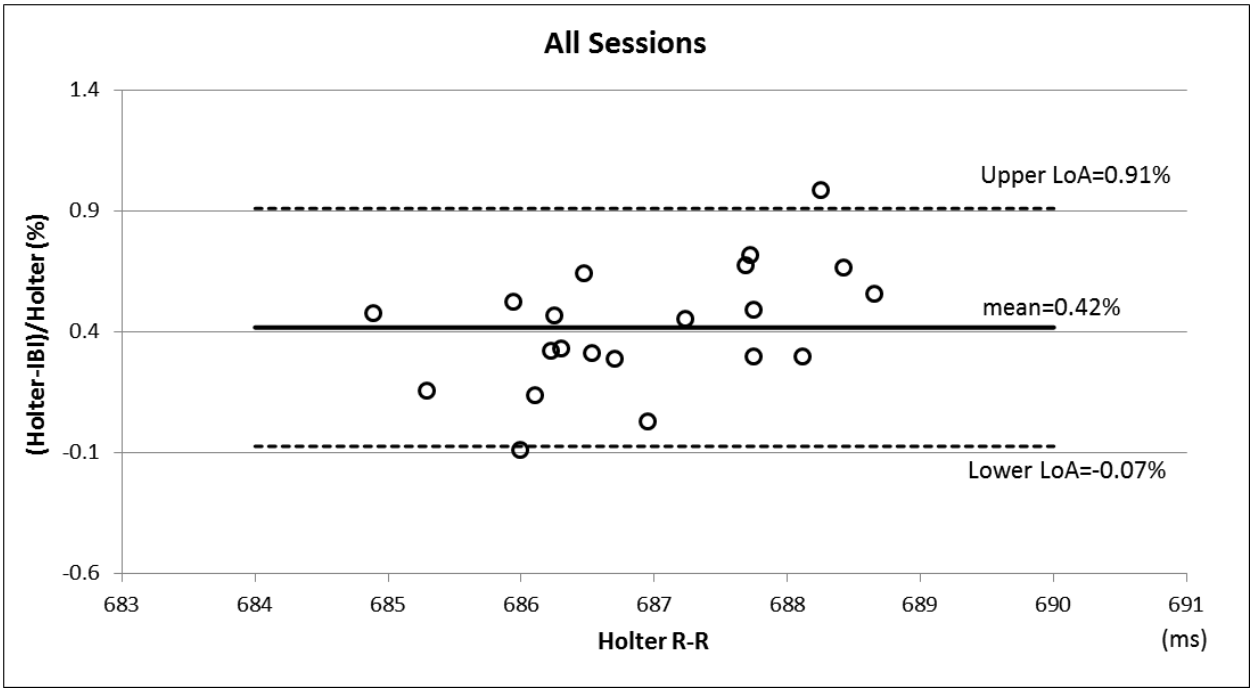
(upper and lower LoA) are presented for each plot. It is obvious that the two methods are in close agreement as the vast majority of the 21 subject points are within the lower and upper LoA.

Figure 14: Bland-Altman plots of the collected data recorded during: a) Base Line phase, b) Training Exercise phase, c) Recovery phase and d) All sessions.




Training Excercise (T)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
(Holter-IBI)/Holter (%)
Upper LoA=2.14%
mean=1.07%
Lower LoA=0%
609
610
611
612
613
614
615
616
617
618
619
(ms)
Holter R-R


Recovery (R)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
(Holter-IBI)/Holter (%)
Upper LoA=1.20%
mean=0.28%
Lower LoA=-0.65%
708
710
712
714
716
718
720
722
724
726
(ms)
Holter R-R



The HRV results from the Kubios derived analysis of the synchronized data extracted from both Holter and E4WB are presented in Table 17.

Table 17. Heart rate variability parameters at rest, exercise and recovery phase by means of ambulatory Holter and E4WB.

	REST			EXERCISE			RECOVERY		
	Holter Mean ± SD	E4WB Mean ± SD	P Value	Holter Mean ± SD	E4WB Mean ± SD	P Value	Holter Mean ± SD	E4WB Mean ± SD	P Value
HR	83.33 ± 3.33	82.15 ± 4.23	0.061	97.83 ± 2.32	98.88 ±3.80	0.068	83.06 ±0.86	83.33 ±0.83	0.17
LF	24.57 ±23.93	32.97 ± 30.10	0.063	30.86 ± 16.74	48.77 ±22.78	<0.001	23.12 ±8.06	24.57 ±10.73	0.50
HF	67.98 ±38.25	55.95 ± 42.25	0.072	48.36 ± 19.47	41.35 ±24.27	0.081	69.61 ±7.64	67.98 ±11.25	0.46
LF/HF	0.40 ±0.54	0.62 ± 0.84	0.057	0.67 ± 0.51	1.33 ±1.00	<0.001	0.35 ±0.14	0.40 ±0.24	0.25

As can be seen, the agreement between the RR series obtained from both devices is high ($p>.05$) for the most part of the whole experiment as both methods were interchangeable when comparing all parameters studied at BL and R phases. During E phase, while HR and HF still showed close agreement ($p>.05$), LF and consequently LF/HF, did not ($p<.001$) as the E4WB derived LF was almost 1.5 times greater than the one derived from HM.

C) Saliva parameters data analysis and comparison with video score.

Tables 18 & 19 show the three different saliva biochemical stress indices (α-Amylase, Cortisol, and IgA) before ($_{pre}$), after exercise ($_{post}$), as well as the percentage difference between those two values ($_{diff}$).

Table 18. α-Amylase$_{pre}$, α-Amylase$_{post}$ and α-Amylase$_{diff}$ of the 21 subjects.

ID	α-Amylase$_{pre}$(U/mL)	α-Amylase$_{post}$(U/mL)	α-Amylase$_{diff}$
1	37.3	150.6	303.75
2	12.27	14.86	21.11
3	40.68	196.8	383.78
4	75.64	158.05	108.95
5	8.5	46.35	445.29
6	26.26	118.08	349.66
7	45	175.68	290.40
8	5.2	26.24	404.62
9	90	124	37.78
10	35	101.02	188.63
11	12.27	53.5	336.02
12	5.81	15	158.18
13	55.68	223.04	300.57
14	47	170	261.70
15	8.5	38	347.06
16	9	31.36	248.44
17	45.26	131.2	189.88
18	55	190.24	245.89
19	5	10.5	110.00
20	45.5	104.31	129.25
21	16.6	85.94	417.71

$\alpha\text{-Amylase}_{diff}=100\cdot(\alpha\text{-Amylase}_{post}-\alpha\text{-Amylase}_{pre})/\alpha\text{-Amylase}_{pre}$

Table 19. $Cortisol_{pre}$, $Cortisol_{post}$, $Cortisol_{diff}$, IgA_{pre}, IgA_{post} and IgA_{diff} of the 21 subjects.

ID	$Cortisol_{pre}$(µg/dL)	$Cortisol_{post}$ (µg/dL)	$Cortisol_{diff}$	IgA_{pre}(µg/mL)	IgA_{post}(µg/mL)	IgA_{diff}
1	0.4	0.7	75.00	168	804	378.57
2	0.25	0.37	48.00	41	127	209.76
3	0.48	0.8	66.67	35	193	451.43
4	0.72	0.9	25.00	57	335	487.72
5	0.4	0.53	32.50	55	208	278.18
6	0.1	0.21	110.00	140	571	307.86
7	0.3	0.66	120.00	90	340	277.78
8	0.08	0.15	87.50	49	113	130.61
9	0.11	0.12	9.09	40	91	127.50
10	0.59	0.84	42.37	134	368	174.63
11	0.19	0.29	52.63	32	160	400.00
12	0.26	0.29	11.54	47	158	236.17
13	0.11	0.13	18.18	71	193	171.83
14	0.45	0.9	100.00	45	270	500.00
15	0.25	0.44	76.00	48	240	400.00
16	0.1	0.11	10.00	32	197	515.63
17	0.12	0.15	25.00	140	404	188.57
18	0.45	0.5	11.11	49	92	87.76
19	0.18	0.3	66.67	49	206	320.41
20	0.15	0.22	46.67	80	300	275.00
21	0.25	0.4	60.00	40	140	250.00

$Cortisol_{diff}=100\cdot (Cortisol_{post}-Cortisol_{pre})/ Cortisol_{pre}$

$IgA_{diff}=100\cdot (IgA_{post}-IgA_{pre})/ IgA_{pre}$

The correlations of Video score with each one of the saliva derived biochemical markers (namely α-Amylase, Cortisol, and IgA) expressed as their percentage differences ($_{diff}$) is shown at the following figures (Figures 15, 16, and 17).

Figure 15. The correlation of Video score with α-Amylase$_{diff}$

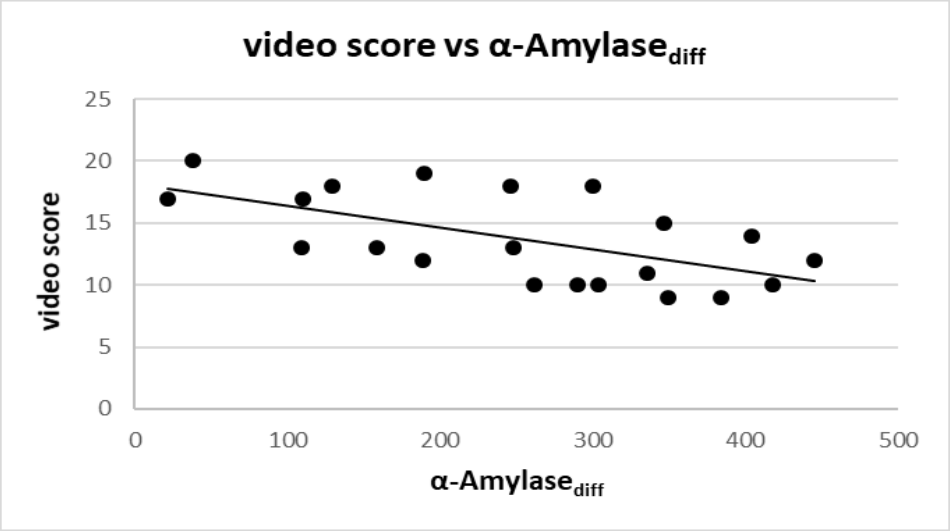


Figure 16. The correlation of Video score with Cortisol$_{diff}$

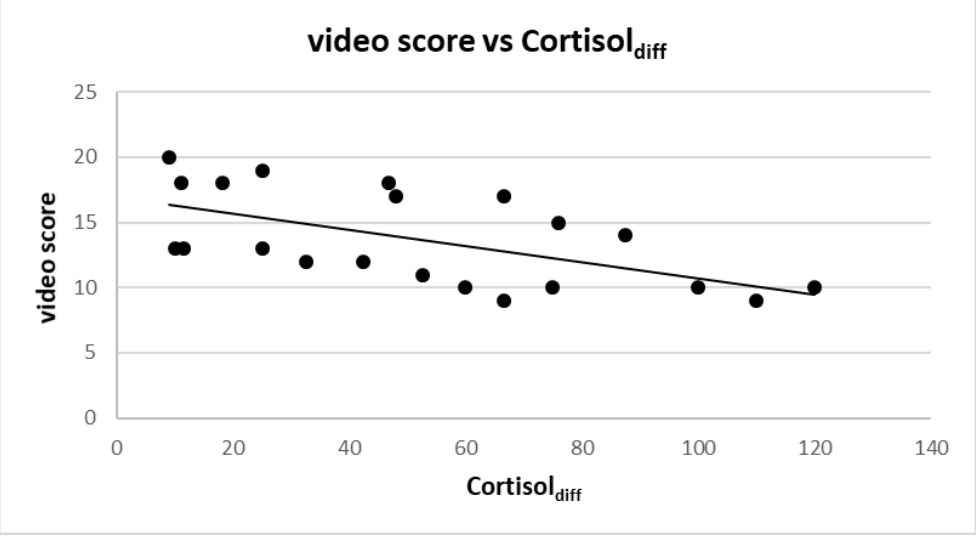
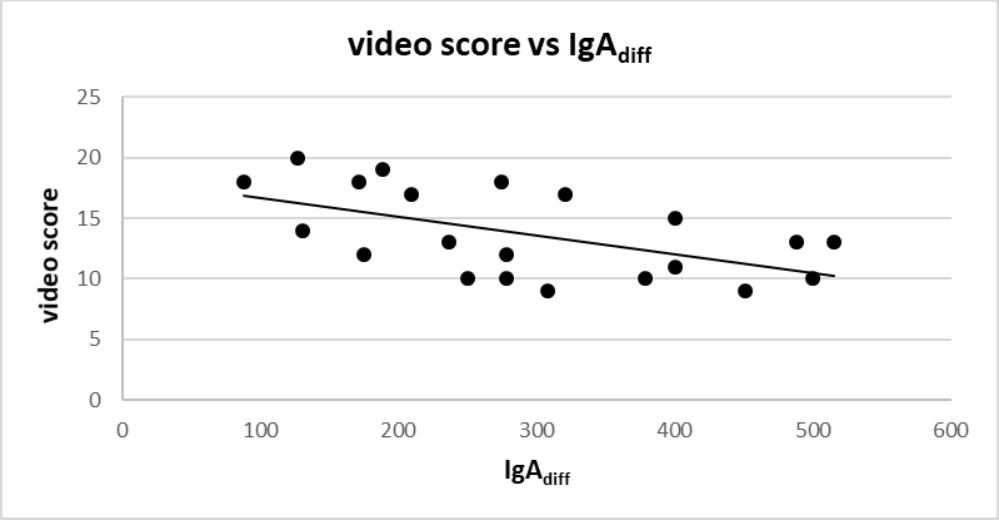


Figure 17. The correlation of Video score with IgA_{diff}



D) Data analysis to define the best stress related parameters

Normality of the collected data was tested using the Kolmogorov-Smirnov test. Pearson correlation coefficient r between each variable and video score was calculated. Multivariate linear regression analysis of all possible combinations of the six predictor variables in order to calculate a predictive model for video score (response variable) was performed. The intercept of each model was set to 21 as this is the maximum score that can be achieved, and therefore the $Value_{diff}$ of each of the six parameters measured would be zero. Since our dataset was small, the selection of the best model, avoiding overfitting but having the minimum expected information loss, was based on the corrected Akaike Information Criterion (AIC_c) [218, 219]. According to this criterion, the model that minimizes the value AIC_c is the preferred model. Additionally, the AIC_c difference $\Delta_i=AIC_{c(min)}-AIC_{c(i)}$ was calculated as well as the Akaike weight w_i:

$$w_i = \frac{\exp(-0.5 \cdot \Delta_i)}{\sum_{n=1}^{N} \exp(-0.5 \cdot \Delta_n)}$$

where N is the number of the candidate models.

The larger the Δ_i is, the smaller the probability of the model i to be the best among the candidate models. Models having $\Delta_i<2$ have substantial evidence of being equal with the model with the minimum AIC_c so they can be used instead of this model [220]. Akaike weight, w_i, can be interpreted as the probability of model i to be the best of the candidate models. Statistical analysis was performed using R software (version 3.5.0) [221].

To each variable, a symbol X_i (i=1,2,…,6) was assigned. At Table 20 the symbol of each variable and the Pearson correlation coefficient r between each variable and video score are presented.

Table 20. The symbol assigned to each variable and its Pearson correlation coefficient (r) with video score.

Variable	Symbol	r
EDA_{diff}	X_1	-0.663
IBI_{diff}	X_2	0.634
HR_{diff}	X_3	-0.624
α-$Amylase_{diff}$	X_4	-0.605
$Cortisol_{diff}$	X_5	-0.587
IgA_{diff}	X_6	-0.554

P was less than 0.01 for all the r values.

Pearson correlation coefficient was also calculated for pairs among the six predictor variables (Table 21). It was found that HR_{diff} had a strong correlation with the IBI_{diff} ($r=-0.988$, $P<0.0001$) and therefore it was omitted from further analysis.

Table 21. Pearson correlation coefficient among the six predictor variables

	EDA_{diff}	HR_{diff}	IBI_{diff}	$Amylase_{diff}$	$Cortisol_{diff}$	IgA_{diff}
EDA_{diff}	1	0.439	-0.475	0.452	0.507	0.261
HR_{diff}	0.439	1	-0.988	0.244	0.351	0.366
IBI_{diff}	-0.475	-0.988	1	-0.263	-0.354	-0.408
$Amylase_{diff}$	0.452	0.244	-0.263	1	0.412	0.194
$Cortisol_{diff}$	0.507	0.351	-0.354	0.412	1	0.266
IgA_{diff}	0.261	0.366	-0.408	0.194	0.266	1

Table 22 presents the predictor variables of each of the 3 best models (the models with $\Delta_i<2$), along with their respective AIC_c, difference (Δ_i) and the Akaike weight (w_i) values.

Table 22. Predictor variables, AIC_c values, differences (Δ_i) and Akaike weights (w_i) of the three best models.

Model	Predictor variables	AIC_c	Δ_i	Akaike weight (w_i)
1	$X_1+X_2+X_4+X_6$	95.81	0.00	0.550
2	$X_2+X_4+X_5+X_6$	97.45	1.64	0.242
3	$X_2+X_4+X_6$	97.76	1.95	0.208

As shown in Table 22, model #1 has the minimum AIC_c value and considers EDA_{diff}, IBI_{diff}, α-$Amylase_{diff}$ and IgA_{diff} as predictor variables. Second best model (model #2) includes the variables IBI_{diff}, α-$Amylase_{diff}$, $Cortisol_{diff}$ and IgA_{diff}, while model #3 uses just three predictor variables i.e. IBI_{diff}, α-$Amylase_{diff}$ and IgA_{diff}.

Looking at Akaike weights, the probability of model #1 to be the best of the candidate models is 55.0% which is more than twice greater than the corresponding probability of model #2 (24.2%) and of model #3 (20.8%). However, as models #2 and #3 have $\Delta_i<2$, they can be also accepted without great loss of information.

Discussion

Despite the fact that surgeon's stress estimation in the OR is of great importance for alleviating and control it, measuring the stress of surgeons during an operation is a very difficult task for several reasons: a) There is no gold standard method to measure stress b) it is rather unfeasible to non-invasively attempt to measure stress parameters in the OR setting. There is a clear need for stress monitoring devices and methods which do not disturb and do not interfere with the surgeon's performance.

From the other hand it is well known that simulation can provide a standardized milieu that allows inter- and intra- individual comparisons among participants [163]. Furthermore, simulation training is a well-established method for improving technical and non-technical skills in a controlled and quantitative manner [188]. It is reported that a high-fidelity simulator can replicate real psychological challenges and stress [123] and yet it can facilitate lower stress loads in real environments [189, 190].

To our knowledge, there are no studies that have used *a wrist wearable device* to assess surgeons' acute stress.

Therefore, the first objective of this study was to assess the clinical validity of a new, easy to wear watch-sized device to a) noninvasively measure stress parameters in novices during a simulation task in a high fidelity simulator as well as to explore the best stress detector among the recorded parameters and b) to compare the cardiac stress parameters such as heart rate and heart rate variability as derived from this device against the most widely method used i.e., the Holter derived HRV in our simulation environment.

The short, six-item STAI scale was chosen as subjective stress indicator as it is well validated and more suitable when time constraints prevent administration of the longer version [22].

We have chosen only male subjects with normal BMI in order to avoid any:

a. Gender related influence [222, 223] and
b. Body size thermoregulatory interference [224] in our results.

In about 1/3 of the subjects the STAI score was less after exercise. This might be due to failure of our exercise to produce subjective stress in some subjects, a finding also observed to a bigger extent in a similar study [159].

Although stress activation has been studied in other occasions using EDA [225, 226], only one study used wrist and/or ankle wearable skin EDA device to measure simulation provoked OR stress [227], while five other simulation studies measured EDA to assess acute stress in surgeons, using electrodes placed on either palmar surface [228, 229, 230], ulnar edge [231] or neck [232], thus obviously reducing clinical application in an OR setting. However, all six aforementioned studies showed that EDA levels are increased in stress.

As can be seen in Figure 3, EDA is the best parameter to detect stress provoking failures during the exercise. As it can be seen in this Figure, EDA started to increase as soon as the E phase starts and remained elevated even during the R phase. EDA was further increased right after each failure by more than 2% ($p<0.001$) per second. Additionally, EDA showed the biggest mean percentage change among the measured parameters, followed by HR while BVP and Acc did not reach statistical significance (Table 6). It can therefore be concluded that EDA is the best parameter to measure failures in our simulation setting followed by HR.

In order to eliminate any individual differences between baseline and post exercise we used the percentage difference between those two values: For

instance in Table 7 we used the STAI value during baseline ($STAI_{pre}$), the STAI value after exercise ($STAI_{post}$), and the percentage difference between the two STAI values ($STAI_{diff}$) in order to compare those values with video score. The same applies for Table 8 where we present the individual values of each of the measured parameters from the E4WB before($_{pre}$), and after exercise ($_{post}$), as well as their percentage differences ($_{diff}$). We found that EDA was the best indicator of stress having better sensitivity than that of the other parameters. Only Holter R-R had specificity equal to that of EDA (Tables 9 and 10). Therefore, we concluded that EDA was the best parameter for detecting stress.

Our results are in line with those of Phitayakorn et al., who used GSR devices in an OR simulation environment. They also observed that during E phase, EDA significantly increased even before failures detection and remained elevated even throughout R phase [228].

Also, in our study, we examined whether the IBI, as derived by a simple to wear, non-invasive wrist device, can substitute the most widely used method of detecting RR intervals through an ambulatory Holter, as a surrogate marker of acute stress. We found excellent correlation coefficient (r) between R-R and IBI during the whole experiment as well as during each one of the three phases (Table 15).

Our study shows that parasympathetic indices of HRV through frequency domain analysis can demonstrate a pronounced reduction in response to acute stress with a concomitant sympathetic augmentation which returns to basic levels in the R phase. This is in total agreement with other studies which showed that upon exercise cessation, HR and HRV demonstrate a time-dependent recovery and eventual return to pre-exercise levels [233].

We found that HR and IBI data derived from PPG signal are highly correlated with HR and R-R intervals from the Holter at BL and R phase. However, during E phase, this correlation is decreasing to moderate levels of agreement. This is in

agreement with other observations that the Holter-derived HRV is in parallel with data obtained from other wearable devices capturing RR intervals either from PPG or from other methods at rest and/or at exercise [206, 234, 235].

The fact that IBI showed a lower specificity to Holter's R-R may be due to the different sampling rate of BVP (64Hz) which is much less than that of ECG (125 to 250 Hz), obviously implying less events detection ability of the PPG devices [194].

Additionally, Bland-Altman analysis was used because the correlation coefficient quantifies only the correlation between two quantities and not their agreement and therefore high correlation does not necessarily mean a close agreement. Instead the Bland–Altman plot compares the agreement among a new measurement technique with the gold standard, as it allows the identification of any systematic difference between the measurements [215].

As is depicted in Figure 14 that the two methods are in close agreement as the vast majority of the 21 subject points are within the lower and upper LoA. Therefore, in our data the Bland-Altman showed that the E4WB derived IBI was in high agreement with Holter RR and therefore E4WB derived heart rate parameters can be used instead of those derived from Holter.

The HRV results from the Kubios derived analysis of the synchronized data extracted from both Holter and E4WB showed an excellent agreement between the RR series obtained from both devices is high ($p>.05$) for the most part of the whole experiment as both methods were interchangeable when comparing all parameters studied at BL and R phases (Table 17). Therefore, it is evident that in our setting we found that BVP was in agreement with HRV not only during the baseline and recovery phases, but also during the exercise phase. According to this finding, E4WB derived BVP can be used as a reliable surrogate of Holter derived HRV.

Except exploiting if the cardiac stress parameters could be accurately be measured through this wrist worn device, we also wanted to explore the value of each one of the suggested stress indices that have been suggested so far. As presented in the introductory section, several biochemical stress markers in saliva have been implied, but most studies have assessed only one of the aforementioned methods, with inconsistent methodology and conflicting results.

Therefore, the second objective of our work was to concomitantly measure the responses of seven noninvasive stress indices and compare them to the video score achieved by novices in a reproducible stressful simulation environment. Namely, we used one subjective method (STAI) plus the three abovementioned ones derived from the wrist-worn device (HR, IBI, EDA) as well as the salivary levels of cortisol, alpha-amylase and IgA. To our knowledge, this study is a pioneer one in assessing many stress-related variables in a simulation environment.

In our simulation setting we used laparoscopic suturing and knot tying which are considered to be the most technically demanding and minimally invasive skills to acquire. Therefore, these two tasks were expected to provoke maximum stress in novice trainees in the simulation environment.

A video review offers unique benefits to measure surgical performance, as the analysis of simulation videos is less susceptible to subjective assessment and observer bias [236]. However, a universally uniform video scoring system that evaluates laparoscopic skills is lacking [237]. Therefore, we used our own video score as an objective method to measure stress and all videos were analyzed accordingly. The range of video score achieved by the 21 subjects was from 9 to 20, (mean= 13.7, SD= ± 3.6) (Table 7).

To include as many possible noninvasive stress parameters in this study, a wrist wearable monitor was concomitantly used to provide cardiac-related stress parameters and EDA.

Although wrist-worn devices usually rely on PPG and tend to be limited in signal accuracy [238], their convenience has led researchers to estimate electrocardiography parameters from PPG signals [239]. Thus far, wearable devices using PPG can only be used as a surrogate for HRV at resting or mild exercise conditions, as their accuracy degrades with increasing exercise load [194]. Hence, in our study, only the HR, IBI, and EDA data, which are relevant for stress estimation, were analyzed.

Additionally, stress biochemical biomarkers in saliva were also assessed. The contemporary use of saliva samples includes a wide range of biomarkers, which, especially in stress research, seems to be an advantageous option because of sample collection simplicity. In our study, 3 stress-related saliva biomarkers were measured, namely, sC, sAA, and sIgA, before and after the simulation exercise aforementioned.

The analysis of all described stress parameters showed that EDA_{diff} had the best correlation with video score ($r=-0.663$), whereas IgA_{diff} showed the worst correlation ($r=-0.554$; see Table 20).

As can be seen from Table 20, the r values of EDA_{diff} and HR_{diff} with VS are negative, whereas that of IBI_{diff} is positive, indicating that the lesser the stress, as measured by small changes between rest and exercise in sweat or heart rhythm values, the better the performance.

Stress response has at least two principal components: one involves a corticotropin-releasing hormone, the activation of the hypothalamic-pituitary-adrenal axis, and the secretion of glucocorticoids (e.g., cortisol) into circulation,

and the other involves activation of the autonomic (sympathetic) nervous system (i.e., catecholamines).

ANS (sympathetic) stimulation increases the sweat production from the eccrine glands of the skin and, therefore, causes fluctuations in electrical conductivity, enabling the measurement of EDA, galvanic skin response, or skin conductance [46]. Although stress activation has been studied using EDA, only one study used a wrist-and/or-ankle-wearable-skin-EDA device to measure simulation-provoked OR stress [227].

However, we could not find any comparison of EDA with other stress parameters. In our study, EDA_{diff} showed the best correlation among all other measured variables with VS. This implies that the provoked mental stress in our study is quickly manifested through an ANS (sympathetic) response.

IBI_{diff} showed the second-best correlation coefficient with VS. This finding is consistent with recent observations indicating that capturing PPG RR intervals can substitute the cumbersome gold-standard Holter-derived HRV at rest and/or at exercise [206, 239].

However, cardiac response to mental stress is not purely mediated from the sympathetic but also has a parasympathetic component [240]. This might explain the inferior correlation of IBI_{diff} vs. EDA_{diff} ($r=-0.444$).

The finding that HR_{diff} strongly correlated with IBI_{diff} was not surprising, because from a physiological viewpoint, both measure an identical parameter. This was an additional reason to omit HR_{diff} from further analysis. Finally, the results of all the parameters measured from the wrist device were readily available, compared with the saliva biomarkers described below.

Regarding the stress saliva biomarkers measured, serum cortisol enters saliva by passive diffusion, maintaining a constant saliva/plasma ratio. Therefore, its saliva concentration is a useful tool for hypothalamic-pituitary-adrenal axis-related stress changes. By contrast, both sIgA and sAA are the products of salivary glands. IgA is produced by local B cells and filtrates actively into saliva as sIgA, whereas the sAA secretion from the salivary glands correlates with sympathetic activity under stressful conditions. However, the production and secretion of both biomarkers from the salivary glands are subject to several other confounding factors, such as intra-individual variations related to local factors. Hence, determining them reliably is difficult [241].

Among the saliva biomarkers measured in our study, α-amylase$_{diff}$ showed the best correlation with VS. This finding is aligned with the previous observation of an ANS-mediated EDA$_{diff}$ rapid increase, when compared with the rest of the other stress variables measured. Salivary α-amylase was found to correlate well and can be used as a surrogate for plasma catecholamine levels, particularly norepinephrine, under a variety of stressful conditions [242].
.
However, a review of the stress literature showed that most sAA studies do not control for the potentially confounding effects of salivary flow rate and do not standardize saliva collection in terms of stimulation or collection duration, although similar observations may apply to other saliva constituents [243]. sAA has been used in only one study that assessed acute mental strain in 51 surgeons, where in sAA and sC samples were obtained during real operations. In that study, higher concentration spikes were observed in the more stressful surgical setting [244].

Because sC is shown to increase in response to acute stressors, it is the most used biomarker in stress studies [163]. sC levels have been used to assess acute stress in simulation settings in conjunction with other subjective (State-Trait

Anxiety Inventory) and/or objective cardiac stress parameters (HR and HRV) [7, 24, 124, 168, 245, 246].

In most studies, increased sC levels were correlated with acute stress. In contrast, Alobid et al. [7] failed to prove that plasma cortisol levels are elevated due to stress, but Wetzel et al. [124] reported lower cortisol levels after stress. As sC is considered as a later stress responder than ANS (acting through the hypothalamus-pituitary-adrenocortical axis), it is possible that a peak of sC levels, which usually occurs 10 minutes after cessation of stress exposure, was missed in our study, although peaks were observed 5 minutes after stress. Hence, a multiple time point sampling is necessary to capture fully the stress-induced cortisol response [13].

Salivary sIgA levels were also measured, as they commonly vary in a complex fashion in response to stress, mood, and emotionality [247].

A recent study analyzed the sC and sIgA levels in saliva and cardiovascular (heart rate and systolic blood pressure) stress responses of oral and maxillofacial surgeons who were engaged in operations with increasing degrees of technical difficulty. It was found that more experienced surgeons have greater stress management capabilities compared with less experienced ones; however, sIgA variations were not significant in any of the groups examined and did not follow the other stress estimation variables that were used [5]. In our study, $sIgA_{diff}$ showed a comparable correlation coefficient with $cortisol_{diff}$.

However, this study sheds light into the noninvasive stress measurement of novice surgeons using an easy-to-wear wrist device and/or by measuring saliva amylase. A large, multicenter, controlled study is necessary to further investigate the utility of multiple stress noninvasive parameters to simulation training.

Finally, in an attempt to find the best model for stress estimation by combining more than one of the parameters studied (Table 22), we found that model #1 has the minimum AIC_c value and considers EDA_{diff}, IBI_{diff}, α-$Amylase_{diff}$ and IgA_{diff} as predictor variables. Second best model (model #2) includes the variables IBI_{diff}, α-$Amylase_{diff}$, $Cortisol_{diff}$ and IgA_{diff}, while model #3 uses just three predictor variables i.e. IBI_{diff}, α-$Amylase_{diff}$ and IgA_{diff}. Looking at Akaike weights, the probability of model #1 to be the best of the candidate models is 55.0% which is more than twice greater than the corresponding probability of model #2 (24.2%) and of model #3 (20.8%). However, as models #2 and #3 have $\Delta_i<2$, they can be also accepted without great loss of information.

It is worth to note that $Cortisol_{diff}$ is not included in model #1 as it increases the value of AIC_c. This means that, in combination with the other parameters, cortisol does not improve the quality of this model as the additional information from cortisol data does not outweigh the cost of adding an extra variable in the model. Model # 2 includes $Cortisol_{diff}$ but does not EDA_{diff} which is more difficult to be applied as it uses three saliva-based parameters which are more time consuming to be measured instead of E4WB-based parameters. Model #3 is easier to be applied than the other two models as it includes one variable less.

Therefore, from a practical point of view, wearing the wrist device for deriving EDA and HR plus measuring a saliva sample of IgA could be suggested as the best non invasive combination for measuring stress at least in our simulation setting.

Limitations

Our study has several limitations. First, it is an observational study that included only novice trainees without using a control group for comparison. Additionally, we may have missed a delayed salivary cortisol response. Furthermore, the number of subjects was rather small and although we collected a large amount of

data it is obvious that more studies with large numbers of subjects is needed. Moreover, the large variations in individual levels of physiological activation observed, suggest that additional research is needed to determine how subjects respond to stressful situations by using these methods. Hence, our findings must be interpreted with caution, before arriving to a clinically relevant conclusion. It is obvious that since our study focused only on a simulated training scenario, our findings cannot be extrapolated into the real OR practice.

Future Directions

As it has been shown that exposure to moderate stress on task trainers may have a beneficial effect and improve performance while excessive stress may have detrimental effects [248, 249], it is evident that further research is required in order to establish a normative threshold beyond which the stress is detrimental and counterproductive.

As our study showed that simple, easy to implement, noninvasive methods of measuring stress in novices in a simulated stressful surgical environment is feasible, it is obvious that further studies with a larger number of both novice as well as of experienced surgeons are needed to exploit the practical implementation and clinical use of our findings.

It is our strong belief that in this way, using each individual’s stress results could be used to effectively formulate individual improvement plans for each practicing surgeon.

Conclusion

The noninvasive techniques used for capturing stress-related data were easy to implement and well-accepted by all participants in the simulation training setting.

Our study demonstrated that a wrist wearable device (E4WB) is an easy to use noninvasive tool capable to noninvasively measure several stress parameters before, during, and after a simulation task in novices. Among them, electrodermal activity was the best stress detector. E4WB derived heart rate and heart rate variability derived from this device highly correlated with the reciprocal values as measured from an ambulatory Holter and therefore they can be used instead of those derived from the Holter in our simulation setting. Therefore, this device may open a practical window to conveniently train surgeons to manage their stress using a surgical simulation training system.

Among the various noninvasive mental stress parameters measured before and immediately after our simulation exercise setting, those under the control of the sympathetic autonomic nervous system (i.e., EDA, HR, IBI, and sAA) showed the best correlation with performance as assessed by video score. EDA had the highest correlation followed by IBI.

References

1. Phitayakorn R, Minehart RD, Hemingway MW, Pian-Smith MCM, Petrusa E. Relationship between physiologic and psychological measures of autonomic activation in operating room teams during a simulated airway emergency. Am J Surg. 2015; 209:86–92. doi:10.1016/j.amjsurg.2014.08.036. Epub 2014 Oct 22.

2. Pavlidis I, Tsiamyrtzis P, Shastri D, Wesley A, Zhou Y, Lindner P, Buddharaju P, Joseph R, Mandapati A, Dunkin B. Fast by Nature - How Stress Patterns Define Human Experience and Performance in Dexterous Tasks. Sci Rep. 2012; 2:305. doi:10.1038/srep00305.

3. Lazarus R, DeeseE J, Osler SF. The effects of psychological stress upon performance. Psychol Bull. 1952; 49:293–317. PMID: 12983450.

4. Müller MP, Hänsel M, Fichtner A, Hardt F, Weber S, Kirschbaum C, Rüder S, Walcher F, Koch T, Eich C. Excellence in performance and stress reduction during two different full scale simulator training courses: A pilot study. Resuscitation. 2009; 80:919–924. doi:10.1016/j.resuscitation.2009.04.027. Epub 2009 May 21.

5. Marrelli M, Gentile S, Palmieri F, Paduano F, Tatullo M. Correlation between Surgeon's Experience, Surgery Complexity and the Alteration of Stress Related Physiological Parameters. PLoSONE 2014; 9(11):e112444. doi:10.1371/journal.pone.0112444. eCollection 2014.

6. Huerta-Franco M, Vargas-Luna M, Tienda P, Delgadillo-Holtfort I, Balleza-Ordaz M, Flores-Hernandez C. Effects of occupational stress on the gastrointestinal tract. World J GastrointestPathophysiol 2013; 4: 108–118. doi:10.4291/wjgp.v4.i4.108.

7. Alobid I, de Pablo J, Mullol J, Centellas S, Parramon G, Carrasco J, Armario A, Bernal-Sprekelsen M. Increased cardiovascular and anxiety outcomes but not endocrine biomarkers of stress during performance of endoscopic sinus surgery: a pilot study among novice surgeons. Arch Otolaryngol Head Neck Surg 2011; 137: 487–492. doi:10.1001/archoto.2011.60.

8. Chrousos, G. Stress and disorders of the stress system. Nat Rev Endocrinol, 2009; 5: 374–381. doi:10.1038/nrendo.2009.106.

9. Danhof-Pont M, van Veen T, Zitman F. Biomarkers in burnout: a systematic review. J Psychosom Res. 2011; 70: 505–524. doi:10.1016/j.jpsychores. 2010.10.012. Epub 2011 Jan 7.

10. Sominsky L, Spencer S. Eating behavior and stress: a pathway to obesity. Front Psychol 2014; 5: 434. doi:10.3389/fpsyg.2014.00434.

11. Flak J, Myers B, Solomon M, McKlveen J, Krause E, Herman J. Role of paraventricular nucleus-projecting norepinephrine/epinephrine neurons in acute and chronic stress. Eur J Neurosci. 2014; 39:1903–1911. doi:10.1111/ejn.12587. Epub 2014 Apr 26.

12. Kudielka B, Hellhammer D, Kirschbaum C. Ten years of research with the Trier Social Stress Test - Revisited. In: Social Neuroscience: Integrating Biological and Psychological Explanations, by E Harmon-Jones and P Winkielman. New York: Guilford: Guilford Press, 2007.

13. Allen A, Kennedy P, Cryan JF, Dinan TG, Clarke G. Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. Neuroscience and Biobehavioral Reviews. 2014; 38:94–124. doi:10.1016/j.neubiorev.2013.11.005. Epub 2013 Nov 14.

14. Birkett MA. The Trier Social Stress Test Protocol for Inducing Psychological Stress. J Vis Exp. 2011; 56:3238. doi:10.3791/3238.

15. Dickerson S, Kemeny M. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. 2004; 130:355–391. doi:10.1037/0033-2909.130.3.355.

16. von Dawans, B, Kirschbaum C, Heinrichs M. The Trier Social Stress Test for groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. Psychoneuroendocrinology. 2011; 36:514–522. doi:10.1016/j.psyneuen.2010.08.004. Epub 2010 Sep 16.

17. Thoma MV, Kirschbaum C, Wolf JM, Rohleder N. Acute stress responses in salivary alpha-amylase predict increases of plasma norepinephrine. Biol Psychol. 2012; 91:342–348. doi:10.1016/j.biopsycho.2012.07.008. Epub 2012 Aug 28.

18. Nater, U.M., La Marca R, Florin L, Moses A, Langhans W, Koller MM, Ehlert U. Stress-induced changes in human salivary alpha-amylase activity: associations with adrenergic activity. Psychoneuroendocrinology. 2006; 31(1):49-58. doi:10.1016/j.psyneuen.2005.05.010.

19. Rohleder N, Wolf JM, Maldonado EF, Kirschbaum C. The psychosocial stress- induced increase in salivary alpha-amylase is independent of saliva flow rate. Psychophysiology. 2006; 43(6):645–52. doi:10.1111/j.1469-8986.2006.00457.x

20. Lackschewitz H, Hüther G, Kröner-Herwig B. Physiological and psychological stress responses in adults with attention-deficit/hyperactivity disorder (ADHD). Psychoneuroendocrinology. 2008; 33:612–624. doi:10.1016/j.psyneuen.2008.01.016

21. Metzger RL. A reliability and validity study of the State Trait Anxiety Inventory. J Clin Psychol. 1976; 32: 276–8.

22. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). BrJ Clin Psychol. 1992; 31:301–6. PMID: 1393159

23. Arora S, Tierney T, Sevdalis N, Aggarwal R, Nestel D, Woloshynowych M, Darzi A, Kneebone R. The Imperial Stress Assessment Tool (ISAT): a feasible, reliable and valid approach to measuring stress in the operating room. World J Surg. 2010; 34(8):1756-1763. doi:10.1007/s00268-010-0559-4.

24. Jones KI, Amawi F, Bhalla A, Peacock O, Williams JP, Lund JN. Assessing surgeon stress when operating using heart rate variability and the State Trait Anxiety Inventory: Will surgery be the death of us? Color Dis. 2015; 17:335–41. doi:10.1111/codi.12844.

25. Pomeranz, B, Macaulay RJ, Caudill MA, Kutz I, Adam D, Gordon D, Kilborn KM, Barger AC, Shannon DC, Cohen RJ, Benson H. Assessment of autonomic function in humans by heart rate spectral analysis. The American Journal of Physiology. 1985; 248(1): H151–H153. doi:10.1152/ajpheart. 1985.248.1.H151.

26. Park SB, Lee BC, Jeong KS. Standardized tests of heart rate variability for autonomic function tests in healthy Koreans. Int J Neurosci. 2007; 117(12):1707-1717. doi:10.1080/00207450601050097.

27. Yang XL, Liu GZ, Tong YH, Yan H, Xu Z, Chen Q, Liu X, Zhang H, Wang H, Tan SH. The history, hotspots, and trends of electrocardiogram. J Geriatr Cardiol. 2015; 12(4):448–456. doi:10.11909/j.issn.1671-5411.2015.04.018.

28. Pappens M, Schroijen M, Sütterlin S, Smets E, Van den Bergh O, Thayer JF, Van Diest I. Resting heart rate variability predicts safety learning and fear extinction in an interoceptive fear conditioning paradigm. PLoS One. 2014 Sep 2;9(9):e105054. doi:10.1371/journal.pone.0105054. eCollection 2014.

29. Lee JM, Kim HC, Kang JI, Suhet Il. Association between stressful life events and resting heart rate. BMC Psychol. 2014; 2:29. doi.org/10.1186/s40359-014-0029-0.

30. Liang CS, Lee JF, Chen CC, Chang YC. Reactive heart rate variability in male patients with first-episode major depressive disorder. Prog Neuropsychopharmacol Biol Psychiatry.2015; 56:52–57. doi.org/10.1016/j.pnpbp.2014.08.004.

31. Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ, Coumel P, Fallen EL, Kennedy HL, Kleiger RE, Lombardi F, Malliani A, Moss AJ, Rottman JN, Schmidt G, Schwartz PJ, Singer DH. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Circulation, 1996; 93(5):1043–1065. PMID: 8598068. doi:10.1161/01.CIR.93.5.1043.

32. Perini R, Veicsteinas A. Heart rate variability and autonomic activity at rest and during exercise in various physiological conditions. Eur J Appl Physiol. 2003; 90(3-4):317-25. doi:10.1007/s00421-003-0953-9.

33. Lin LY, Wu CC, Liu CC, Ho YB, Liau YL, Lee YT. Derangement of heart rate variability during a catastrophic earthquake: a possible mechanism for increased heart attacks. Pacing Clin Electrophysiol , 2001: 24(11):1596-601. PMID: 11816627.

34. Gil E, Orini M, Bailón R, Vergara JM, Mainardi L, Laguna P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. Physiological Measurement. 2010: 31(9): 1271–1290. doi:10.1088/0967-3334/31/9/015.

35. Shin JH, Hwang SH, Chang MH, Park KS. Heart rate variability analysis using a ballistocardiogram during Valsalva manoeuvre and post exercise. Physiological Measurement. 2011; 32(8):1239–1264. doi:10.1088/0967-3334/32/8/015.

36. Lu G, Yang F, Tian Y, Jing X, Wang J. Contact-free measurement of heart rate variability via a microwave sensor. Sensors. 2009; 9(12):9572–9581. doi.org/10.3390/s91209572.

37. Poh MZ, McDuff DJ, Picard RW. Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Transactions on Biomedical Engineering. 2011; 58(1):7–11. doi:10.1109/TBME.2010.2086456.

38. Sun Y, Hu S, Azorin-Peris V, Kalawsky R, Greenwald S. Noncontact imaging photoplethysmography to effectively access pulse rate variability. Journal of Biomedical Optics. 2013; 18(6):Article ID 61205. doi:10.1117/1.JBO.18.6.061205.

39. Peng RC, Zhou LX, Lin WH, Zhang YT. Extraction of Heart Rate Variability from Smartphone Photoplethysmograms. Computational and Mathematical Methods in Medicine. 2015; 2015: Article ID 516826, 11 pages. http://dx.doi.org/10.1155/2015/516826.

40. Koenig J, Thayer JF. Sex differences in healthy human heart rate variability: A meta-analysis. Neurosci Biobehav Rev. 2016; 64:288-310. doi:10.1016/j.neubiorev.2016.03.007.

41. Böhm B, Rötting N, Schwenk W, Grebe S, Mansmann U. A prospective randomized trial on heart rate variability of the surgical team during laparoscopic and conventional sigmoid resection. Arch Surg. 2001; 136: 305-310. PMID: 11231850.

42. Demirtas Y, Tulmac M, Yavuzer R, Yalcin R, Ayhan S, Latifoglu O, Atabay K. Plastic surgeon's life: marvelous for mind, exhausting for body. Plast Reconstr Surg, 2004: 114: 923–31. PMID: 15468400.

43. Song, MH, Tokuda Y, Nakayama T, Sato M, Hattori K. Intraoperative heart rate variability of a cardiac surgeon himself in a coronary artery bypass grafting surgery. InteractCardiovasc Thorac Surg. 2009; 8:639–41. doi:10.1510/icvts.2008.195941.

44. Prichard RS, O'Neill CJ, Oucharek JJ, Holmes CY, Delbridge LW, Sywak MS. A prospective study of heart rate variability in endocrine surgery: surgical training increases consultant's mental strain. J Surg Educ.2012; 69:453–8. doi:10.1016/j.jsurg.2012.04.002.

45. Thielmann B, Boeckelmann I. Heart Rate Variability as an Indicator of Mental Stress in Surgeons - A Review of the Literature. Zentralbl Chir. 2015; 141(5):577-582. doi:10.1055/s-0034-1396295.

46. Asahina M, Poudel A, Hirano S. Sweating on the palm and sole. Physiological and clinical relevance. Clin Auton Res. 2015; 25:153-9. doi:10.1007/s10286-015-0282-1.

47. Sequeira H, Hot P, Silvert L, Delplanque S. Electrical autonomic correlates of emotion. Int J Psychophysiol. 2009; 71:50–6. doi:10.1007/s10286-015-0282-1.

48. Coli M, Fontanella L, Ippoliti L, Merla A. Multiresolution KLE of psycho-physiological signals. Proceedings of S.Co. 2007, Book of Short Papers. 2007. 116–121.

49. Shastri D, Merla A, Tsiamyrtzis P, Pavlidis I. Imaging facial signs of neurophysiological responses. IEEE Trans Biomed Eng. 2009; 56(2):477–84. doi:10.1109/TBME.2008.2003265

50. Engert V, Merla A, Grant JA, Cardone D, Tusche A, Singer T. Exploring the use of thermal infrared imaging in human stress research. PLoS ONE, 2014; 9(3). doi:10.1371/journal.pone.0090782.e90782.

51. Di Giacinto A, Brunetti M, Sepede G, Ferretti A, Merla A. Thermal signature of fear conditioning in mild post traumatic stress disorder. Neuroscience, 2014; 266:216–223. doi:10.1016/j.neuroscience.2014.02.009.

52. Villanueva I, Valladares M, Goodridge W. Use of Galvanic Skin Responses, Salivary Biomarkers, and Self-reports to Assess Undergraduate Student Performance During a Laboratory Exam Activity. J Vis Exp. 2016; 10:108. doi:10.3791/53255.

53. Bouwknecht JA, Olivier B, Paylor RE. The stress-induced hyperthermia paradigm as a physiological animal model for anxiety: a review of pharmacological and genetic studies in the mouse. Neurosci Biobehav Rev. 2007; 31(1):41–59. doi:10.1016/j.neubiorev.2006.02.002.

54. Vinkers CH, Penning R, Hellhammer J, Verster JC, Klaessens JHGM, Olivier B. The effect of stress on core and peripheral body temperature in humans. Stress. 2013; 16(5):520–530. doi:10.3109/10253890.2013.807243.

55. Ioannou S, Gallese V, Merla A. Thermal infrared imaging in psychophysiology: potentialities and limits: thermal infrared imaging in psychophysiology. Psychophysiology. 2014; 51(10):951–963. doi:10.1111/psyp.12243.

56. Tyler CJ. The effect of skin thermistor fixation method on weighted mean skin temperature. Physiol Meas. 2011; 32:1541–1547. doi:10.1088/0967-3334/32/10/003.

57. Costello J, McInerney C, Bleakley C, Selfe J, Donnelly A. The use of thermal imaging in assessing skin temperature following cryotherapy: a review. J Therm Biol. 2012; 37:103–110. https://doi.org/10.1016/j.jtherbio.2011.11.008

58. Fernandes A , Amorim PR, Brito CJ, de Moura AG, Moreira DG, Costa CM, Sillero-Quintana M, Marins JC. Measuring skin temperature before, during and after exercise: a comparison of thermocouples and infrared thermography. Physiol Meas. 2014; 35: 189–203. doi:10.1088/0967-3334/35/2/189

59. McCafferty DJ. Applications of thermal imaging in avian science. Ibis , 2013; 155(1):4–15. https://doi.org/10.1111/ibi.12010

60. Costello J, Stewart I, Selfe J, Karki A, Donnelly A. Use of thermal imaging in sports medicine research: a short report. Int SportMed J. 2013; 14(2):94–98. https://hdl.handle.net/10520/EJC138120.

61. Ring EFJ. Pioneering progress in infrared imaging in medicine. Quant InfraRed Thermogr J. 2014; 11(1):57–65. https://doi.org/10.1080/17686733.2014.892667.

62. Fernández-Cuevas I a, Joao Carlos BouzasMarins JC, , ArnáizLastras J, Gómez Carmona PM, Cano SP, , García-Concepción MA a, Sillero-Quintana M. Classification of factors influencing the use of infrared thermography in humans: A review. Infrared Physics & Technology. 2015; 71:28–55. https://doi.org/10.1016/j.infrared.2015.02.007.

63. Cardone D, Pinti P, Merla A. Thermal Infrared Imaging-Based Computational Psychophysiology for Psychometrics. Comput Math Methods Med. 2015; 2015: 984353.http://dx.doi.org/10.1155/2015/984353.

64. Ruopsa N, Kujala S, Kaarela O, Ohtonen P, Ryhänen J. Wireless infrared thermometer in the follow-up of finger temperatures. J Hand Surg Eur. 2009; 34:526–529. doi:10.1177/1753193409102456.

65. Kelechi TJ, Good A, Mueller M. Agreement and repeatability of an infrared thermometer. J Nurs Meas. 2011; 19(1): 55–64. doi:10.1891/1061-3749.19.1.55.

66. Herborn KA, Graves JL, Jerem P, Evans NP, Nager R, McCafferty DJ, McKeegan DE. Skin temperature reveals the intensity of acute stress. Physiol Behav. 2015; 152:225–230. doi:10.1016/j.physbeh.2015.09.032.

67. Genno H, Ishikawa K, Kanbara O, Kikumoto M, Fujiwara Y, Suzuki R. Using facial skin temperature to objectively evaluate sensations. Int J Ind Ergon. 1997; 19:161–171.https://doi.org/10.1016/S0169-8141(96)00011-X.

68. Stewart M, Stafford KJ, Dowling SK, Schaefer AL, Webster JR. Eye temperature and heart rate variability of calves disbudded with or without

local anaesthetic. Physiol Behav. 2008; 93(4–5):789–797. doi:10.1016/j.physbeh.2007.11.044.

69. Moe RO, Nordgreen J, Janczak AM, Bakken M, Spruijt BM, Jensen P. Anticipatory and foraging behaviors in response to palatable food reward in chickens: effects of dopamine D2 receptor blockade and domestication. Physiol Behav. 2014; 133:170–177. doi:10.1016/j.physbeh.2014.05.023.

70. Ioannou S, Chotard H, Davila-Ross M. No strings attached: physiological monitoring of rhesus monkeys (Macaca mulatta) with thermal imaging. Front Behav Neurosci. 2015; 9:160. doi:10.3389/fnbeh.2015.00160.

71. Merla A. Thermal expression of intersubjectivity offers new possibilities to human-machine and technologically mediated interactions. Frontiers in Psychology. 2014; 5: article 802. doi:10.3389/fpsyg.2014.00802.

72. Christensen J, Vaeth M, Wenzel A. Thermographic imaging of facial skin-gender differences and temperature changes over time in healthy subjects. Dentomaxillofac. Radiol. 2012; 41:662–667. doi:10.1259/dmfr/55922484.

73. Baum, B, Yates JR, Srivastava S, Wong DT, Melvin JE. Scientific frontiers: emerging technologies for salivary diagnostics. Adv Dent Res. 2011; 23:360-368. doi:10.1177/0022034511420433.

74. Rahim MAA, Rahim ZHA, Ahmad WAW, Hashim OH. Can Saliva Proteins Be Used to Predict the Onset of Acute Myocardial Infarction among High-Risk Patients? Int J Med Sci. 2015; 12(4):329-335. doi:10.7150/ijms.11280.

75. Stokes JR, Davies GA. Viscoelasticity of human whole saliva collected after acid and mechanical stimulation. Biorheology. 2007; 44(3):141-160. PMID: 17851164.

76. Inoue H, Ono K, Masuda W, Inagaki T, Yokota M, Inenaga K. Rheological properties of human saliva and salivary mucins. J Oral Biosci. 2008; 50:134-141.https://doi.org/10.1016/S1349-0079(08)80027-8.

77. Fenoll-Palomares, C, Muñoz Montagud JV, Sanchiz V, Herreros B, Hernández V, Mínguez M, Benages A. Unstimulated salivary flow rate, pH and buffer capacity of saliva in healthy volunteers. Rev Esp Enferm Dig. 2004; 96(11):773-783. doi:10.4321/S1130-01082004001100005.

78. de Almeida Pdel V, Grégio AM, Machado MA, de Lima AA, Azevedo LR. Saliva composition and functions: a comprehensive review. J Contemp Dent Pract. 2008; 9:72-80.PMID: 18335122.

79. Michael DJ, Daugherty S, Santos A, Ruby BC, Kalns JE. Fatigue biomarker index: an objective salivary measure of fatigue level. Accid Anal Prev. 2012; 45S:68-73. doi:10.1016/j.aap.2011.09.029.

80. Azman MY, Aziah D, Norhayati MN, Zahiruddin M. Stress and salivary biomarker among assistant medical officers in Ministry of Health (MOH) hospitals in Kelantan and Terengganu, Malaysia. Int J Collab Res Intern Med Public Health. 2011; 3(8):634-43. http://iomcworld.com/ijcrimph/ijcrimph-v03-n08-04.htm.

81. Kirschbaum C, Hellhammer D. Salivary Cortisol. In G. Fink (Ed.), Encyclopedia of stress. 2000; 3:379-384. San Diego: CA: Academic Press. eBook ISBN: 9780080569772.

82. Malamud D, Rodriguez-Chavez IR. Saliva as a diagnostic fluid. Dent Clin North Am. 2011; 55(1):159–178. doi:10.1016/j.cden.2010.08.004.

83. Davis R, Goddard K. Specimen collection within the CRN: a critical appraisal. THE HMO Cancer Research Network. 2010; 18-20.

84. Trueba AF, Mizrachi D, Auchus RJ, Vogel PD, Ritz T. Effects of psychosocial stress on the pattern of salivary protein release. Physiol Behav. 2012; 105:841-9. doi:10.1016/j.physbeh.2011.10.014.

85. Sadi H, Finkelman M, Rosenberg M. Salivary cortisol, salivary alpha amylase, and the dental anxiety scale. Anesth Prog. 2013; 60:46–53.

86. Obayashi, K. Salivary mental stress proteins. Clin Chim Acta, 2013: 425: 196–201. doi:10.2344/0003-3006-60.2.46.

87. Reshma, AP, Arunachalam R, Pillai JK, Kurra SB, Varkey VK, Prince MJ. Chromogranin A: Novel biomarker between periodontal disease and psychosocial stress. J Indian Soc Periodontol. 2013; 17(2):214–218. doi:10.4103/0972-124X.113076.

88. Yi TC, Moochhala S. Mini-Review Article – Current Opinion on Salivary Biomarkers as a Measurement for Stress and Fatigue. The Open Biomarkers Journal. 2013; 6:9-14. doi:10.2174/1875318301306010009.

89. Bosch J. The Use of Saliva Markers in Psychobiology: Mechanisms and Methods. Monogr Oral Sci. 2014; 24:99-108. doi:10.1159/000358864.

90. Lewis J. Steroid analysis in saliva: an overview. Clin Biochem Rev. 2006; 27:139–146. PMID: 17268582 PMCID: PMC1579286.

91. Bosch J, de Geus EC Veerman E, Hoogstaten J, Nieuw Amerongen AV. Innate secretory immunity in response to laboratory stressors that evoke

distinct patterns of cardiac autonomic activity. Psychosomatic Medicine. 2003; 65(2):245–258. PMID: 12651992.

92. Proctor G, Carpenter G. Regulation of salivary gland function by autonomic nerves. Auton Neurosci. 2007; 133(1):3-18. doi:10.1016/j.autneu.2006.10.006.

93. Schenkels LC, Veerman EC, Nieuw Amerongen AV. Biochemical composition of human saliva in relation to other mucosal fluids. Crit Rev Oral Biol Med. 1995; 6(2):161-75. https://doi.org/10.1177/10454411950060020501.

94. Veerman EC, van den Keybus PA, Vissink A, Nieuw Amerongen AV. Human glandular salivas: their separate collection and analysis. Eur J Oral Sci. 1996; 104:346–352. doi:10.1111/j.1600-0722.1996.tb00090.x.

95. Bosch J, Veerman E, de Geus E, Proctor G. α-Amylase as a reliable and convenient measure of sympathetic activity: don't start salivating just yet. Psychoneuroendocrinology. 2011; 36(4):449-453. doi:10.1016/j.psyneuen.2010.12.019.

96. Berntson GG, Quigley KS, Lozano D. Cardiovascular psychophysiology. In Handbook of psychophysiology (3rd ed.), by JT Cacioppo, LG Tassinary and GG Berntson, 182–210. Cambridge University Press, 2007. https://doi.org/10.1017/9781107415782.009.

97. Mezzacappa ES, Kelsey RM, Katkin ES , Sloan RP. Vagal rebound and recovery from psychological stress. Psychosomatic Medicine. 2001; 63(4):650–657. doi:10.1097/00006842-200107000-00018.

98. Rottenberg J, Wilhelm FH, Gross JJ, Gotlib IH. Vagal rebound during resolution of tearful crying among depressed and nondepressed individuals. Psychophysiology. 2003; 40(1):1–6. doi:10.1111/1469-8986.00001.

99. Brandtzeg P. Do salivary antibodies reliably reflect both mucosal and systemic immunity? Ann N Y Acad Sci. 2007; 1098:288-311. doi:10.1196/annals.1384.012.

100. Crawford JM, Taubman MA, Smith DJ. Minor salivary glands as a major source of secretory immunoglobulin A in the human oral cavity. Science. 1975; 190(4220):1206-9. doi:10.1196/annals.1384.012.

101. Bishop NC, Gleeson M. Acute and chronic effects of exercise on markers of mucosal immunity. Front Biosci. 2009; 1(14):4444-56. doi:10.2741/3540.

102. Campos-Rodriguez, R, Godínez-Victoria M, Abarca-Rojano E, Pacheco-Yépez J, Reyna-Garfias H, Barbosa-Cabrera RE, Drago-Serrano ME. Stress modulates intestinal secretory immunoglobulin A. Front Integr Neurosci. 2013; 7:86. doi:10.3389/fnint.2013.00086.

103. Hucklebridge F, Lambert S, Clow A, Warburton DM, Evans PD, Sherwood N. Modulation of secretory immunoglobulin A in saliva; response to manipulation of mood. Biol Psychol, 2000: 53(1):25-35. doi:10.1016/S0301-0511(00)00040-5.

104. Deinzer R, Kleineidam C, Stiller-Winkler R, Idel H, Bachg D. Prolonged reduction of salivary immunoglobulin A (sIgA) after a major academic exam. Int J Psychophysiol. 2000; 37:219-32. doi:10.1016/S0167-8760(99)00112-9.

105. Saruta J, Tsukinoki K, Sasaguri K, Ishii H, Yasuda M, Osamura YR, Watanabe Y, Sato S.. Expression and localization of chromogranin A gene and protein in human submandibular gland. . Cells Tissues Organs. 2005; 180:237–244. doi:10.1159/000088939.

106. Koh DS-Q, Koh GC-H. The use of salivary biomarkers in occupational and environmental medicine. Occup Environ Med. 2007; 64(3):202-10. doi:10.1136/oem.2006.026567.

107. Helle K. Chromogranins A and B and secretogranin II as prohormones for regulatory peptides from the diffuse neuroendocrine system. Results Probl Cell Differ. 2010; 50:21-44. doi:10.1007/400_2009_26.

108. Garrett JR. The proper role of nerves in salivary secretion: A review. Journal of Dental Research. 1987; 66(2):387–397. doi:10.1177/00220345870660020201.

109. Inder WJ, Dimeski G, Russell A. Measurement of salivary cortisol in 2012 - laboratory techniques and clinical indications. Clin Endocrinol (Oxf). 2012; 77: 645–651. doi:10.1111/j.1365-2265.2012.04508.x.

110. Cabral DM, Antonini SR, Custódio RJ, Martinelli CE Jr, da Silva CA. Measurement of salivary cortisol as a marker of stress in newborns in a neonatal intensive care unit. Horm Res Paediatr, 2013: 79(6):373-378. doi:10.1159/000351942.

111. Bozovic D, Racic M, Ivkovic N. Salivary cortisol levels as a biological marker of stress reaction. Med Arch. 2013; 67:374-377. doi:10.5455/medarh.2013.67.374-377.

112. Desai GS, Mathews ST. Saliva as a non-invasive diagnostic tool for inflammation and insulin-resistance. World J Diabetes. 2014; 5(6):730-738. doi:10.4239/wjd.v5.i6.730.

113. Hellhammer DH, Wust S, Kudielka BM. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology. 2009; 34:163-71. doi:10.1016/j.psyneuen.2008.10.026.

114. Vining, RF, McGinley RA. The measurement of hormones in saliva: Possibilities and pitfalls. J Steroid Biochem. 1987; 27(1-3):81-94. https://doi.org/10.1016/0022-4731(87)90297-4.

115. Lucassen EA, Cizza G. The Hypothalamic-Pituitary-Adrenal Axis, Obesity, and Chronic Stress Exposure: Sleep and the HPA Axis in Obesity. Curr Obes Rep. 2012; 1(4):208-215. doi:10.1007/s13679-012-0028-5.

116. Saiyudthong S, Suwannarat P, Trongwongsa T, Srisurapanon S. Comparison between ECL and ELISA for the detection of salivary cortisol and determination of the relationship between cortisol in saliva and serum measured by ECL. Science Asia. 2010; 36:169-171. doi:10.2306/scienceasia1513-1874.2010.36.169.

117. McGraw LK, Out D, Hammermeister JJ, Ohlson CJ, Pickering MA, Granger DA. Nature, correlates, and consequences of stress-related biological reactivity and regulation in Army nurses during combat casualty simulation. Psychoneuroendocrinology. 2013; 38(1):135-144. doi:10.1016/j.psyneuen.2012.05.009.

118. Pottier P, Hardouin JB, Dejoie T, Bonnaud A, Le Loupp AG, Planchon B, Leblanc V. Stress responses in medical students in ambulatory and in-hospital patient consultations. Med. Educ. 2011; 45(7):678-687. doi:10.1111/j.1365-2923.2011.03935.x.

119. Ebrecht M, Hextall J, Kirtley LG, Taylor A, Dyson M, Weinman J. Perceived stress and cortisol levels predict speed of wound healing in healthy male adults. Psychoneuroendocrinology. 2004; 29(6):798-809. doi:10.1016/S0306-4530(03)00144-6.

120. Takatsuji K, Sugimoto Y, Ishizaki S, Ozaki Y, Matsuyama E, Yamaguchi Y. The effects of examination stress on salivary cortisol, immunoglobulin A, and chromogranin A in nursing students. Biomed Res, 2008: 29(4):221-4. https://doi.org/10.2220/biomedres.29.221.

121. Strahler J, Ziegert T. Psychobiological stress response to a simulated school shooting in police officers. Psychoneuroendocrinology. 2015; 51:80—91. doi:10.1016/j.psyneuen.2014.09.016.

122. Hamer M, Steptoe A. Cortisol responses to mental stress and incident hypertension in healthy men and women. J Clin Endocrinol Metab. 2012; 97(1):E29–34. doi:10.1210/jc.2011-2132.

123. Hassan I, Weyers P, Maschuw K, Dick B, Gerdes B, Rothmund M, Zielke A. Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance. Br J Surg. 2006; 93(12):1554–9. doi:10.1002/bjs.5544.

124. Wetzel CM, Black SA, Hanna GB, Athanasiou T, Kneebone RL, Nestel D, Wolfe JH, Woloshynowych M. The effects of stress and coping on surgical performance during simulations. Ann Surg. 2010; 25(1):171–6. doi:10.1097/SLA.0b013e3181b3b2be.

125. Keitel A, Ringleb M, Schwartges I, Weik U, Picker O, Stockhorst U, Deinzer R. Endocrine and psychological stress responses in a simulated emergency situation. Psychoneuroendocrinology. 2011; 36(1): 98-108. doi:10.1016/j.psyneuen.2010.06.011.

126. Mandel AL, Peyrot des Gachons C, Plank KL, Alarcon S, Breslin PA. Individual differences in AMY1 gene copy number, salivary alpha-amylase levels, and the perception of oral starch. PLoS One. 2010; 5(10):e13352. doi:10.1371/journal.pone.0013352.

127. van Stegeren A, Rohleder N, Everaerd W, Wolf OT. Salivary alpha amylase as marker for adrenergic activity during stress: Effect of betablockade. Psychoneuroendocrinology. 2006; 31(1):137-141. doi:10.1016/j.psyneuen.2005.05.012.

128. Nater UM, Rohleder N. Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research. Psychoneuroendocrinology, 2009: 34(4): 486-96. doi:10.1016/j.psyneuen.2009.01.014.

129. Rohleder N, Nater U. Determinants of salivary alpha-amylase in humans and methodological considerations. Psychoneuroendocrinology. 2009; 34(4):469–485. doi:10.1016/j.psyneuen.2008.12.004.

130. Takai N, Yamaguchi M, Aragaki T, Eto K, Uchihashi K, Nishikawa Y. Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults. Arch Oral Biol. 2004; 49(12):963–8. doi:10.1016/j.archoralbio.2004.06.007.

131. Capranica L, Lupo C, Cortis C, Chiodo S, Cibelli G, Tessitore A.. Salivary cortisol and alpha-amylase reactivity to taekwondo competition in children.

Eur J Appl Physiol. 2012; 112(2):647–652. doi:10.1007/s00421-011-2023-z.

132. Nagy T, van Lien R, Willemsen G, Proctor G, Efting M, Fülöp M, Bárdos G, Veerman EC, Bosch JA. A fluid response: Alpha-amylase reactions to acute laboratory stress are related to sample timing and saliva flow rate. Biological Psychology. 2015; 109:111–119. doi:10.1016/j.biopsycho.2015.04.012.

133. Yoon SA, Weierich MR. Salivary biomarkers of neural hypervigilance in trauma-exposed women. Psychoneuroendocrinology. 2016; 63:17-25. doi:10.1016/j.psyneuen.2015.09.007.

134. Holmgren J, Czerkinsky C. Mucosal immunity and vaccines. Nature Medicine. 2005; 11(4 Suppl):s45-53. doi:10.1038/nm1213.

135. Li T-L, Gleeson M. The effect of single and repeated bouts of prolonged cycling and circadian variation on saliva flow rate, immunoglobulin A and alpha-amylase responses. J Sports Sci. 2004; 22(11-12):1015-1024. doi:10.1080/02640410410001716733.

136. Tsujita, S, and K Morimoto. Secretory IgA in saliva can be a useful stress marker. Environ. Health Prev. Med, 1999: 4: 1–8.

137. Campisi, J, Y Bravo, J Cole, and K Gobeil. Acute psychosocial stress differentially influences salivary endocrine and immune measures in undergraduate students. Physiology and Behavior., 2012: 107 (212):317-321.

138. Winkler, H, and R Fischer-Colbrie. The chromogranins A and B: The first 25 years and future perspectives. Neuroscience, 1992: 49: 497–528.

139. Smith, WJ, and N Kirshner. A specific soluble protein from the catecholamine storage vesicles of bovine adrenal medulla. I. Purification and chemical characterization. Mol Pharmacol , 1967: 3: 52–62.

140. Taupenot, L, KL Harper, and DT O'Connor. The chromogranin-secretogranin family. N Engl J Med , 2003: 348: 1134–1149.

141. Noto, Y., T. Sato, M. Kudo, K. Kurata, and K. Hirota. The relationship between salivary biomarkers and State-Trait Anxiety Inventory score under mental arithmetic stress: A pilot study. Anesth Analg , 2005: 101: 1873–1876.

142. Obara, S., and H. Iwama. Assessment of psychological tension after premedication by measurement of salivary chromogranin A. J Clin Anesth , 2005: 17: 554–557.

143. Nakane H, Asami O, Yamada Y, Harada T, Matsui N, Kanno T, Yanaihara N. Salivary chromogranin A as an index of psychosomatic stress response. Biomed Res. 1998; 19(6):401–406. https://doi.org/10.2220/biomedres.19.401.

144. Nakane H, Asami O, Yamada Y, Ohira H. Effect of negative air ions on computer operation, anxiety and salivary chromo- granin A-like immunoreactivity. Int J Psychophysiol. 2002; 46:85–89. doi:10.1016/S0167-8760(02)00067-3.

145. Ng V, Koh D, Mok BY, Chia SE, Lim LP. Salivary biomarkers associated with academic assessment stress among dental undergraduates. J Dent Educ. 2003; 67(10):1091-1094.PMID: 14587673.

146. Kanamaru Y, Kikukawa A, Shimamura K. Salivary chromogranin-A as a marker of psychological stress during a cognitive test battery in humans. Stress. 2006; 9(3):127–131. doi:10.1080/14769670600909594.

147. Nomura S, Mizuno T, Nozawa A, Asano H, Ide H. Characteristics of salivary chromogranin A as a short-term mental stress biomarker. Trans Jpn Soc Med Biol Eng. 2010; 48(2):207–212. doi:10.11239/jsmbe.48.207.

148. Nakane H. Salivary chromogranin A as index of psychosomatic stress response. Technical Journal R&D Review of Toyota CRDL. 1999; 34(3):17–22 (in Japanese).

149. Yoto A, Murao S, Nakamura Y, Yokogoshi H. Intake of green tea inhibited increase of salivary chromogranin A after mental task stress loads. Journal of Physiological Anthropology. 2014; 33:20. doi:10.1186/1880-6805-33-20.

150. Dimsdale JE, O'Connor D, Ziegler M, Mills P. Does chromogranin A respond do short-term mild physiologic challenge? Neuropsychopharmacology. 1989; 2(3):237–240. doi:10.1016/0893-133X(89)90027-4.

151. Lee JY, Chung JW, YK Kim, Chung SC, Kho HS. Comparison of the composition of oral mucosal residual saliva with whole saliva. Oral Dis. 2007; 13(6):550-54. doi:10.1111/j.1601-0825.2006.01332.x.

152. Furuholm J, Sorsa T, Qvarnström M, Janket SJ, Tervahartiala T, Nuutinen P, Meurman JH. Salivary matrix metalloproteinase-8 in patients with and without coronary heart disease may indicate an increased susceptibility to periodontal disease may indicate an increased susceptibility to periodontal disease. J Periodontal Res. 2006; 41(5):486-89. doi:10.1111/j.1600-0765.2006.00900.x.

153. Aydin S. A comparison of ghrelin, glucose, alpha-amylase and protein levels in saliva from diabetics. J Biochem Mol Biol. 2007; 40(1):29-35. doi:10.5483/BMBRep.2007.40.1.029.

154. Shi D, Meng H, Xu L, Zhang L, Chen Z, Feng X, Lu R, Sun X, Ren X. Systemic inflammation markers in patients with aggressive periodontitis: A pilot study. J Periodontol. 2008; 79(12):2340-46. doi:10.1902/jop.2008.080192.

155. Sivakumar T, Hand AR, Mednieks M. Secretory proteins in the saliva of children. J Oral Sci. 2009; 51(4):573-80. doi:10.2334/josnusd.51.573.

156. de Oliveira VN, Bessa A, Lamounier RP, de Santana MG, de Mello MT, Espindola FS.. Changes in the salivary biomarkers induced by an effort test. Int J Sports Med. 2010; 31(6):377-81. doi:10.1055/s-0030-1248332.

157. RudneyJD, Smith QT. Relationships between levels of lysozyme, lactoferrin, salivary perioxidase, and secretory immunoglobulin A in stimulated parotid saliva. Infect Immun. 1985; 49(3):469-75. PMID:4030086.

158. Morino M, Masaki C, Seo Y, Mukai C, Mukaibo T, Kondo Y, Shiiba S, Nakamoto T, Hosokawa R. Non- randomized controlled prospective study on perioperative levels of stress and dysautonomia during dental implant surgery. J Prosthodont Res. 2014; 58(3):177–183. doi:10.1016/j.jpor.2014.03.006.

159. Arora S, Sevdalis N, Nestel D, Woloshynowych M, Darzi A, Kneebone R. The impact of stress on surgical performance: A systematic review of the literature. Surgery. 2010;147:318–30. http://dx.doi.org/10.1016/j.surg.2009.10.007.

160. Anton NE, Montero PN, Howley LD, Brown C, Stefanidis D. What stress coping strategies are surgeons relying upon during surgery? Am J Surg. 2015;210:846–51. doi:10.1016/j.amjsurg.2015.04.002.

161. Gurman GM, Klein M, Weksler N. Professional stress in anesthesiology: a review. J Clin MonitComput. 2012;26:329-335. doi:10.1007/s10877-011-9328-7.

162. Wong SW, Smith R, Crowe P. Optimizing the operating theatre environment. ANZ J Surg. 2010;80:917–24. doi:10.1111/j.1445-2197.2010.05526.x.

163. Georgiou K, Larentzakis A, Papavassiliou GA. Surgeons' and surgical trainees' acute stress in real operations or simulation: A systematic review.

The Surgeon. 2017; 15(6):355-365. http://dx.doi.org/10.1016/j.surge.2017.06.003.

164. Andersen LPH, Klein M, Gögenur I, Rosenberg J. Psychological and physical stress among experienced and inexperienced surgeons during laparoscopic cholecystectomy. Surg LaparoscEndoscPercutan Tech. 2012;22(1):73–8. doi:10.1097/SLE.0b013e3182420acf.

165. Porto GG, Carneiro SC, Vasconcelos BC, Nascimento MM, Leal JLF. Burnout syndrome in oral and maxillofacial surgeons: a critical analysis. Int J Oral Maxillofac Surg. 2014;43(7):894–899. doi:10.1016/j.ijom.2013.10.025.

166. Arora S, Sevdalis N, Nestel D, Tierney T, Woloshynowych M, Kneebone R. Managing intraoperative stress: what do surgeons want from a crisis training program? Am J Surg. 2009;197(4):537–43. doi:10.1016/j.amjsurg.2008.02.009.

167. Langelotz C, Scharfenberg M, Haase O, Schwenk W. Stress and heart rate variability in surgeons during a 24-hour shift. Arch Surg. 2008; 143:751–5. doi:10.1001/archsurg.143.8.751.

168. Arora S, Aggarwal R, Moran A, Sirimanna P, Crochet P, Darzi A, Kneebone R, Sevdalis N.. Mental practice: effective stress management training for novice surgeons. J Am Coll Surg. 2011; 212(2):225–233. doi:10.1016/j.jamcollsurg.2010.09.025.

169. Lazarus G, Weis KH. Heart rate of anaesthesiologists under physical and psychological professional stress. Anaesthesist. 1977; 26(8): 375-80. PMID:907071.

170. PayneRL, Rick JT. Heart rate as an indicator of stress in surgeons and anaesthetists. J Psychosom Res. 1986; 30(4):411-420.PMID: 3761226.

171. Kain ZN, Chan KM, Katz JD, Nigam A, Fleisher L, Dolev J, Rosenfeld LE. Anaesthesiologists and acute perioperative stress: a cohort study. Anesth Analg. 2002; 95(1):177-183. doi:10.1097/00000539-200207000-00031.

172. Katz JD. Conflict and its resolution in the operating room. J Clin Anesth. 2007; 19(2):152–8. doi:10.1016/j.jclinane.2006.07.007.

173. D'Addessi A, Bongiovanni L, Volpe A, Pinto F, Bassi P. Human factors in surgery: from Three Mile Island to the operating room. Urol Int. 2009; 83(3):249–57. doi:10.1159/000241662.

174. Driskell, JE, E Salas, and J. Johnston. Does stress lead to a loss of team perspective? Group Dyn. 1999; 3(4):291–302. doi:10.1037/1089-2699.3.4.291.

175. Owen H. Early use of simulation in medical education. Simul Healthc 2012; 7(2):102–16. doi:10.1097/SIH.0b013e3182415a91.

176. Acton R. The Evolving Role of Simulation in Teaching Surgery in Undergraduate Medical Education. Surg Clin N Am. 2015; 95(4):739–750. http://dx.doi.org/10.1016/j.suc.2015.04.001.

177. Sachdeva A. Establishment of American College of Surgeons-accredited Education Institutes: the dawn of a new era in surgical education and training. J Surg Educ. 2010; 67(4):249–250.

178. DuCoin C. Second-place essay-revolution: five-year general surgery residency:reform or revolution? Bull Am Coll Surg. 2014; 99(11):20–2. PMID: 25509224.

179. Pucher PH, Darzi A, Aggarwal R. Development of an evidence-based curriculumfor training of ward-based surgical care. Am J Surg. 2014; 207(2):213–7. doi:10.1016/j.amjsurg.2013.10.006.

180. Buckley CE, Kavanagh DO, Traynor O, Buckley CE, Kavanagh DO, Traynor O, Neary PC. Is the skill set obtained in surgical simulation transferable to the operating theatre? Am J Surg. 2014; 207(1):146–57. doi:10.1016/j.amjsurg.2013.06.017.

181. Scott DJ, Pugh CM, Ritter EM, Jacobs LM, Pellegrini CA, Sachdeva AK. New directions in simulation-based surgical education and training: validation and transfer of surgical skills, use of non surgeons as faculty, use of simulation to screen and select surgery residents, and long-term follow-up of learners. Surgery. 2011; 149(6):735–744. doi:10.1016/j.surg.2010.11.010.

182. American College of Surgeons/Association of Program Directors in Surgery/Association for Surgical Education. ACS/APDS/ASE resident preparatory curriculum.2014. Available at: http://www.facs.org/education/program/resident-prep.

183. Kolozsvari NO, Feldman LS, Vassiliou MC, Demyttenaere S, Hoover ML. Sim one, do one, teach one: considerations in designing training curricula for surgical simulation. J Surg Educ. 2011; 68(5):421–7. doi:10.1016/j.jsurg.2011.03.010.

184. Brown KM, Paige JT. Preface. Surg Clin N Am. 2015; 95(4):xvii–xviii. http://dx.doi.org/10.1016/j.suc.2015.06.002.

185. Kahol K, Ashby A, Smith M, Kahol K, Ashby A, Smith M, Ferrara JJ. Quantitative evaluation of retention of surgical skills learned in simulation. J Surg Educ. 2010; 67(6):421–6. doi:10.1016/j.jsurg.2010.05.005.

186. Schlickum M, Hedman L, Enochsson L, Henningsohn L, Kjellin A, Felländer-Tsai L. Surgical simulation tasks challenge visual working memory and visual-spatial ability differently. World J Surg. 2011; 35(4):710–5. doi:10.1007/s00268-011-0981-2.

187. Cates CU, Gallagher AG. The future of simulation technologies for complex cardiovascular procedures. Eur Heart J. 2012; 33(17):2127-34. doi:10.1093/eurheartj/ehs155.

188. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg. 2002; 236:458-64. doi:10.1097/01.SLA.0000028969.51489.B4.

189. Pai DR, Ram S, Madan SS, Soe HHK, Barua A. Causes of stress and their change with repeated sessions as perceived by undergraduate medical students during high-fidelity trauma simulation. Natl Med J India. 2014; 27:192–7. https://www.researchgate.net/publication/272192738.

190. Flanagan B, Nestel D, Joseph M. Making patient safety the focus: crisis resource management in the undergraduate curriculum. Med Educ. 2004; 38(1):56–66. https://doi.org/10.1111/j.1365-2923.2004.01701.x.

191. Issa N, Salud L, Kwan C, Woods K, Pugh C. Validity and reliability of a sensor-enabled intubation trainer: a focus on patient-centered data. J Surg Res. 2012; 177:27–32. doi:10.1016/j.jss.2012.03.047.

192. Rutherford DN, D'Angelo A-LD, Law KE, Pugh CM. Advanced Engineering Technology for Measuring Performance. Surg Clin North Am. 2015; 95:813–26. doi:10.1016/j.suc.2015.04.005.

193. Phitayakorn R, Minehart RD, Pian-Smith M, Hemingway W, Petrusa E. Practicality of using galvanic skin response to measure intraoperative physiologic autonomic activation in operating room team members. Surgery. 2015; 158:1415-20. doi:10.1016/j.surg.2015.04.009.

194. Georgiou K, Larentzakis AV, Khamis N, Alsuhaibani GI, Alaska AA, Giallafos EJ. Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. Folia Medica.2018; 60(1):95-107. doi:10.2478/folmed-2018-0012

195. van Ravenswaaij-Arts CM, Kollée LA, Hopman JC, Stoelinga GB, van Geijn HP. Heart rate variability. Ann Intern Med. 1993; 118:436–47. PMID:8439119.

196. Thayer JF, Friedman BH, Borkovec TD. Autonomic characteristics of generalized anxiety disorder and worry. Biol Psychiatry. 1996; 39:255–66. doi:10.1016/0006-3223(95)00136-0.

197. Vrijkotte TG, van Doornen LJ, de Geus EJ. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. Hypertens. 2000; 35(4):880–6. doi:10.1161/01.HYP.35.4.880.

198. Buccelletti E, Gilardi E, Scaini E, Galiuto L, Persiani R, Biondi A, Basile F, GentilioniSilveri N. Heart rate variability and myocardial infarction: systematic literature review and metanalysis. European Review for Medical and Pharmacological Sciences. 2009; 13(4):299-307. PMID: 19694345.

199. La Rovere MT, Pinna GD, Maestri R, Barlera S, Bernardinangeli M, Veniani M, Nicolosi GL, Marchioli R, Tavazzi L; GISSI-HF Investigators. Autonomic markers and cardiovascular and arrhythmic events in heart failure patients: still a place in prognostication? Data from the GISSI-HF Trial. European Journal of Heart Failure. 2012; 14(12):1410–1419. doi:10.1093/eurjhf/hfs126.

200. von Rosenberg W, Chanwimalueang T, Adjei T, Jaffer U, Goverdovsky V, Mandic DP. Resolving ambiguities in the LF/HF ratio: LF-HF scatter plots for the categorization of mental and physical stress from HRV. Frontiers in Physiology. 2017; 8:360. doi:10.3389/fphys.2017.00360.

201. Pérez-Riera AR, Barbosa-Barros R, Daminello-Raimundo R, de Abreu LC. Main artifacts in electrocardiography. Annals of Noninvasive Electrocardiology. 2018; 23(2): e12494. doi:10.1111/anec.12494.

202. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement. 2007; 28(3):R1–39. doi:10.1088/0967-3334/28/3/R01.

203. Gambarotta N, Aletti F, Baselli G, Ferrario M. A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters. Medical and Biological Engineering and Computing. 2016; 54(7):1025-1035. doi:10.1007/s11517-016-1453-5.

204. Wallén MB, Hasson D, Theorell T, Canlon B, Osika W. Possibilities and limitations of the Polar RS800 in measuring heart rate variability at rest.

European Journal of Applied Physiology. 2012; 112(3):1153–1165. doi:10.1007/s00421-011-2079-9.

205. Weinschenk SW, Beise RD, Lorenz, J. Heart rate variability (HRV) in deep breathing tests and 5-min short-term recordings: agreement of ear photoplethysmography with ecg measurements, in 343 subjects. European Journal of Applied Physiology. 2016; 116(8):1527–1535. doi:10.1007/s00421-016-3401-3.

206. Lee ES, Lee JS, Joo MC, Kim JH, Noh SE. Accuracy of heart rate measurement using smartphones during treadmill exercise in male patients with ischemic heart disease. Annals of Rehabilitation Medicine. 2017; 41(1):129–137. doi:10.5535/arm.2017.41.1.129.

207. Sun Y, Thakor N. Photoplethysmography revisited: from contact to noncontact, from point to imaging. IEEE Transactions on Biomedical Engineering. 2016; 63(3):463–477. doi:10.1109/TBME.2015.2476337.

208. Chen X, Huang YY, Yun F, Chen TJ, Li J2. Effect of changes in sympathovagal balance on the accuracy of heart rate variability obtained from photoplethysmography. Experimental and Therapeutic Medicine. 2015; 10(6):2311–2318. doi:10.3892/etm.2015.2784.

209. Weinschenk SW, Beise RD, Lorenz, J. Heart rate variability (HRV) in deep breathing tests and 5-min short-term recordings: agreement of ear photoplethysmography with ecg measurements, in 343 subjects. European Journal of Applied Physiology. 2016; 116(8):1527–1535. doi:10.1007/s00421-016-3401-3.

210. Shin H. Ambient temperature effect on pulse rate variability as an alternative to heart rate variability in young adult. Journal of Clinical Monitoring and Computing. 2016; 30(6):939–948. doi:10.1007/s10877-015-9798-0.

211. Flatt A, Esco MR. Validity of the Ithlete™ smart phone application for determining ultra-short-term heart rate variability. Journal of Human Kinetics. 2013; 39:85–92. doi:10.2478/hukin-2013-0071.

212. Esco MR, Flatt AA, Nakamura FY. Agreement Between a Smartphone Pulse Sensor Application and Electrocardiography for Determining ln RMSSD. Journal of Strength and Conditioning Research. 2017; 31(2):380-385. doi:10.1519/JSC.0000000000001519.

213. Heathers JA. Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in

psychophysiological research. International Journal of Psychophysiology. 2013; 89(3):297–304. doi:10.1016/j.ijpsycho.2013.05.017.

214. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV–heart rate variability analysis software. Computer Methods and Programs in Biomedicine. 2014; 113(1),210–220. doi:10.1016/j.cmpb.2013.07.024.

215. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310.

216. Schäfer A, Vagedes J. How Accurate Is Pulse Rate Variability as an Estimate of Heart Rate Variability? A Review on Studies Comparing Photoplethysmographic Technology with an Electrocardiogram. Int J Cardiol. 2013; 166(1):15–29. doi:10.1016/j.ijcard.2012.03.119.

217. Nater UM, Rohleder N, Schlotz W, Ehlert U, Kirschbaum C: Determinants of the diurnal course of salivary alpha-amylase. Psychoneuroendocrinology 2007;32(4):392–401. doi:10.1016/j.psyneuen.2007.02.007.

218. Sugiura N: Further analysis of the data by Akaike's information criterion and the finite corrections. Commun Stat Theory Methods. 1978;7(1):13-26. doi:10.1080/03610927808827599.

219. Hurvich C and Tsai C: Regression and Time Series Model Selection in Small Samples. Biometrika. 1989;76(2): 297-307. doi:10.2307/2336663

220. Burnham KP, Anderson DR: Multimodel interference: understanding AIC and BIC in model selection. Sociol Methods Res. 2004;33(2):261-304. doi:10.1177/0049124104268644.

221. R Core Team (2018). R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org. Accessed June 10, 2018.

222. Ali A, Subhi Y, Ringsted C, Konge L. Gender differences in the acquisition of surgical skills: a systematic review. Surg Endosc. 2015;29(11):3065-73. doi:10.1007/s00464-015-4092-2

223. Flyckt RL, White EE, Goodman LR, Mohr C, Dutta S, and Zanotti KM. The Use of Laparoscopy Simulation to Explore Gender Differences in Resident Surgical Confidence. ObstetGynecol Int. 2017;2017:1945801. doi:10.1155/2017/1945801

224. Jay O, Cramer MN. A new approach for comparing thermoregulatory responses of subjects with different body sizes. Temperature (Austin). 2015;2(1):42-3. doi:10.1080/23328940.2014.996093

225. Kreibig SD. Autonomic nervous system activity in emotion: a review. Biol Psychol. 2010;84(3):394-421. doi:10.1016/j.biopsycho.2010.03.010

226. Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems 2005;6(2):156-166. doi:10.1109/TITS.2005.848368

227. Phitayakorn R, Minehart RD, Pian-Smith MCM, Hemingway MW, Petrusa ER. Practicality of using galvanic skin response to measure intraoperative physiologic autonomic activation in operating room team members. Surgery 2015;158(5):1415-20. doi:10.1016/j.surg.2015.04.009

228. Smith WD, Chung YH, Berguer R. A Virtual Instrument Ergonomics Workstation for Measuring the Mental Workload of Performing Video-Endoscopic Surgery. Studies in Health Technology and Informatics. 2000;70 (Medicine Meets Virtual Reality): 309-315. doi:10.3233/978-1-60750-914-1-309

229. Warren D Smith WD, Berguer R, Rosser JC. Wireless virtual instrument measurement of surgeons' physical and mental workloads for robotic versus manual minimally invasive surgery. Studies in health technology and informatics 2003;94 (Medicine Meets Virtual Reality 11):318-24. doi:10.3233/978-1-60750-938-7-318

230. Berguer R, Smith W. An Ergonomic Comparison of Robotic and Laparoscopic Technique: The Influence of Surgeon Experience and Task Complexity. Journal of Surgical Research 2006;134(1):87–92. doi:10.1016/j.jss.2005.10.0031

231. Berguer R, Smith WD, Chung YH. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. Surg Endosc. 2001;15:1204–1207. doi:10.1007/s004640080030

232. Schuetz M, Gockel I, Beardi J, Hakman P, Dunschede F, Moenk S, Heinrichs W, Junginger T. Three different types of surgeon-specific stress reactions identified by laparoscopic simulation in a virtual scenario. Surg Endosc. 2008;22:1263–1267. doi:10.1007/s00464-007-9605-1

233. Hernando D, Garatachea N, Almeida R, Casajús JA, Bailón R. Validation of heart rate monitor Polar RS800 for heart rate variability analysis during

exercise. J Strength Cond Res 2018 Mar;32(3):716-725. doi:10.1519/JSC.0000000000001662

234. Hong S, Yang Y, Kim S, Shin S, Lee I, Jang Y, Kim K, Yi H. Performance study of the wearable one-lead wireless electrocardiographic monitoring system. Telemed J E Health 2009;15(2):166-75. doi:10.1089/tmj.2008.0071

235. Heathers JAJ. Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in psychophysiological research. Int J Psychophysiol. 2013;89(3):297-304. doi:10.1016/j.ijpsycho.2013.05.017

236. Mellinger JD, Williams RG, Sanfey H, Fryer JP, DaRosa D, George BC, Bohnen JD, Schuller MC, Sandhu G, Minter RM, Gardner AK, Scott DJ. Teaching and assessing operative skills: from theory to practice. Curr Probl Surg 2017;54:44-81. doi:10.1067/j.cpsurg.2016.11.007

237. Sleiman Z, Tanos V, Van Belle Y, Carvalho JL, Campo R. The European Academy laparoscopic “Suturing Training and Testing” (SUTT) significantly improves surgeons’ performance. Facts Views Vis Obgyn 2015;7:153-160. PMCID: PMC4788330.

238. Wang, R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, Gillinov M. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017;2:104-106. doi:10.1001/jamacardio.2016.3340.

239. Lee SP, Ha G, Wright DE, Ma Y, Sen-Gupta E, Haubrich NR, Branche PC, Li W, Huppert GL, Johnson M, Mutlu HB, Li K, Sheth N, Wright JA Jr., Huang Y, Mansour M, Rogers JA, Ghaffari R. Highly flexible, wearable, and disposable cardiac biosensors for remote and ambulatory monitoring. NPJ Digit Med 2018;1:2. doi:10.1038/s41746-017-0009-x.

240. Yamanouchi K, Hayashida N, Kuba S. Increase in operator’s sympathetic nerve activity during a complicated hepatobiliary surgery: evidence for surgeons’ mental stress. Tohoku J Exp Med 2015;237:157-162. doi:10.1620/tjem.237.157.

241. Bosch JA. The use of saliva markers in psychobiology: mechanisms and methods. Monogr Oral Sci 2014;24:99-108. doi:10.1159/000358864.

242. Nater UM, Rohlederb N. Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: current state of research. Psychoneuroendocrinology 2009;34:486-496. doi:10.1016/j.psyneuen.2009.01.014.

243. Bosch J, Veerman E, de Geus E, Proctor GB. α-Amylase as a reliable and convenient measure of sympathetic activity: don't start salivating just yet! Psychoneuroendocrinology 2011;36:449-453. doi:10.1016/j.psyneuen.2010.12.019.

244. Engelmann C, Schneider M, Kirschbaum C, Grote G, Dingemann J, Schoof S, Ure BM. Effects of intraoperative breaks on mental and somatic operator fatigue: a randomized clinical trial. Surg Endosc 2011;25:1245-1250. doi:10.1007/s00464-010-1350-1.

245. Wetzel CM, George A, Hanna GB, Athanasiou T, Black SA, Kneebone RL, Nestel D, Woloshynowych M. Stress management training for surgeons: a randomized, controlled, intervention study. Ann Surg 2011;253:488-494. doi:10.1097/SLA.0b013e318209a594.

246. Ghazali DA, Faure JP, Breque C, Oriot D. Evaluation of stress patterns during simulated laparoscopy in residency. Minerva Chir 2016;71:252-261. PMID: 27077271.

247. Bosch JA, Ring C, de Geus EJC, Veerman EC, Amerongen AV. Stress and secretory immunity. Int Rev Neurobiol 2002;52:213-253. doi:10.1016/S0074-7742(02)52011-0.

248. LeBlanc V, Woodrow SI, Sidhu R, Dubrowski A. Examination stress leads to improvements on fundamental technical skills for surgery. Am J Surg. 2008;196(1):114-9. doi:10.1016/j.amjsurg.2007.08.059

249. Angeli A, Minetto M, Dovio A, Paccotti P. The overtraining syndrome in athletes: A stress-related disorder. J Endocrinol Invest. 2004;27(6):603–612. doi:10.1007/BF03347487

Appendix

Abbreviations

AIC_c, Akaike Information Criterion

ANS, autonomic nervous system

ANTS, Anaesthetist's Non-Technical Skills

BVP, blood volume pressure

BL, baseline phase

CgA, Chromogranin A

E, exercise phase

ECL, electrochemiluminescence

EDA, electrodermal activity

ELISA, enzyme-linked immunosorbent assay

GSR, galvanic skin response

HF, high frequency

HPA, hypothalamus-pituitary-adrenocortical axis

HR, heart rate

HRV, heart rate variability IBI, interbeat interval

IR, infrared

IRT, infrared thermography

LF, low frequency

OR, operating room

PPG, photoplethysmography

sAA, saliva α-amylase

sC, saliva cortisol

sIgA, saliva secretory immunoglobulin A

SAM, sympathetic-adrenomedullary system

ST, skin temperature

STAI, State Trait Anxiety Inventory

TSST, Trier Social Stress Test

VAS, visual analogue scale

VLF, very low frequency

VS, video score