



# Implementation of Eye-Tracking Technology to Monitor Clinician Fatigue in Routine Clinical Care: A Feasibility Study

Bashar Kadhim<sup>1</sup> Saif Khairat<sup>2,3</sup> Fangyong Li<sup>4</sup> Isabel T. Gross<sup>5</sup> Bidisha Nath<sup>1</sup> Ronald G. Hauser<sup>6,7</sup> Edward R. Melnick<sup>1,4</sup>

<sup>1</sup> Department of Emergency Medicine, Yale School of Medicine, New Haven, Connecticut, United States

<sup>2</sup> Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

<sup>3</sup> School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

<sup>4</sup> Yale School of Public Health, New Haven, Connecticut, United States

<sup>5</sup> Department of Pediatrics (Emergency Medicine), Yale School of Medicine, New Haven, Connecticut, United States

<sup>6</sup> Department of Laboratory Medicine, Yale School of Medicine, New Haven, Connecticut, United States

<sup>7</sup> Department of Pathology and Laboratory Medicine, VA Connecticut Healthcare System, West Haven, Connecticut, United States

**Address for correspondence** Edward R. Melnick, MD, MHS, Department of Emergency Medicine, Yale School of Medicine, 464 Congress Ave, Suite 260, New Haven, CT 06519, United States (e-mail: edward.melnick@yale.edu).

ACI Open 2023;7:e1–e7.

## Abstract

**Background** Physician fatigue increases the likelihood of medical errors. Eye-tracking technology offers an unobtrusive and objective way to measure fatigue but has only been implemented in controlled settings.

**Objective** Our objective was to determine the feasibility of capturing physiological indicators of fatigue using eye-tracking technology in a real-world clinical setting.

**Methods** A mixed-methods feasibility study was performed in a convenience sample of clinicians practicing in an urban, academic emergency department from November 11 to December 15, 2021. Outcomes included fatigue assessed at the beginning and end of each shift via eye-tracking (with low scores indicating greater fatigue) and self-report.

**Results** Among 15 participants, self-reported fatigue and task load increased from the beginning to the end of their shift (fatigue visual analog scale [FVAS] 3.7–4.6,  $p = 0.04$ ; physician task load [PTL] 97.7–154.3,  $p = 0.01$ ). It was feasible to collect eye-tracking data at a fixed computer workstation with twice daily calibration and 61% capture of reliable data when the clinician was working at the study computer. Eye-tracking metrics did not change significantly from the beginning to the end of the shift. Eye metric fatigue score was associated with the change in PTL score ( $r = 0.59$ ,  $p = 0.02$ ) but not FVAS. This association persisted after adjusting for age, gender, and role, with every 10-point increase in PTL, there was a 0.02-point increase in fatigue score ( $p = 0.04$ ).

## Keywords

- ▶ eye tracking
- ▶ fatigue
- ▶ cognitive load
- ▶ usability

received  
May 3, 2022  
accepted after revision  
November 13, 2022

DOI <https://doi.org/10.1055/s-0042-1760267>.  
ISSN 2566-9346.

© 2023. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

**Conclusion** It is unclear whether the inability to detect fatigue via eye-tracking in routine clinical care was due to confounding factors, the technology, study design, sample size, or an absence of physiological fatigue. Further research and advances in functionality are needed to determine the eye-tracking technology's role in measuring clinician fatigue in routine care.

## Introduction

Physician fatigue increases the likelihood of medical errors.<sup>1</sup> Fatigue is a complex phenomenon affected by sleep, health, time-of-day, workload, and multiple lifestyle factors.<sup>2</sup> The electronic health record (EHR) also contributes to physician fatigue and burnout.<sup>3–5</sup> Extraneous load refers to cognitive load or working memory that is necessary to accomplish a task due to disorganized, redundant, or incomplete presentation of information.<sup>6</sup> Currently, pertinent patient information is scattered and hidden throughout clunky EHR interfaces and workflows creating high extraneous load for clinicians.<sup>7</sup>

Current approaches to measure clinician fatigue have all been performed in simulated environments, both by implementing fatigue-related surveys<sup>8</sup> and using biosensors,<sup>9</sup> are cumbersome, interruptive,<sup>10</sup> and subject to response bias. Unobtrusive and objective methods to detect fatigue among clinicians could help to identify opportunities for appropriate interventions to mitigate physician fatigue. Eye tracking uses high-resolution cameras and near-infrared light sources to determine user gaze and pupil diameter which can provide insight into a subject's attention.<sup>11</sup> Pupillary constriction is a validated biomarker for fatigue<sup>12–14</sup> and an indicator of cognitive load.

Eye-tracking technology has never been implemented in routine clinical care to measure clinician fatigue. Unlike wearable biosensors, new screen-based eye trackers are mounted directly on the computer display, do not require physical contact, and can capture eye metrics with fewer privacy concerns. Since previous studies have validated the ability of eye-tracking technology to capture fatigue in a controlled environment, the goal of this study was to determine the feasibility of capturing physiological indicators of fatigue using eye-tracking technology in a real-world clinical setting. Screen-based eye-tracking technology was implemented in an emergency department (ED). The objectives of this study were to (1) implement eye-tracking technology in a clinical setting and (2) examine the association between eye-tracking metrics and self-reported clinician fatigue over the course of a clinical shift. This study investigates the practicality of using an eye-tracker device in a live clinical setting rather than simulation-based testing, which has been previously explored.<sup>5</sup> Therefore, our analysis focused on fatigue over the course of the shift rather than specific times of interest (TOI). Given the known physiological associations between eye metrics and fatigue as well as empirical eye-tracking fatigue research, we

hypothesized that a decrease in pupil size as well as reduction in peak saccade speed<sup>15,16</sup> and increases in both average duration of whole fixation duration<sup>17</sup> and average saccade amplitude<sup>18,19</sup> would be captured at the end of the shift compared with the beginning of the shift and be associated with self-reported fatigue metrics.

## Methods

### Study Design, Setting, and Participants

This feasibility study was conducted in an urban, academic level I trauma center. All clinicians (attending, residents, physician associates, and advanced nurse practitioners [APRNs]) practicing in the ED were eligible to participate unless they had a documented history of using bifocal eyeglasses, droopy eyelids, loss of one eye/amblyopia, and using certain medications (anticholinergics, antidepressants, benzodiazepines, decongestants, mydriatics, stimulants, and selective serotonin reuptake inhibitors). Participants were recruited by convenience sampling. All participants provided informed consent. The protocol was approved by our institution's IRB (protocol #2000030671).

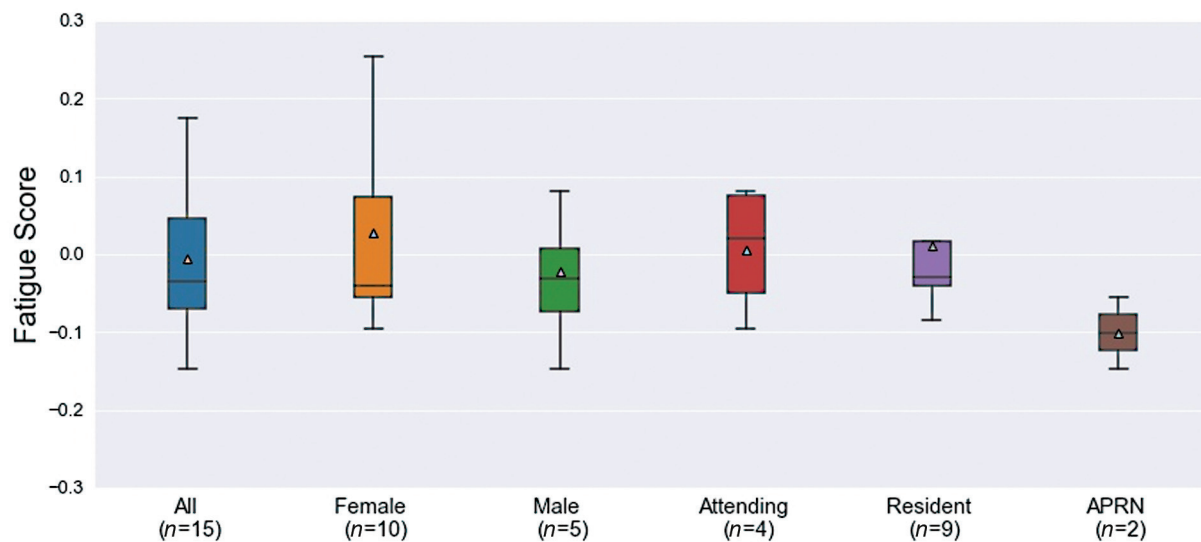
### Survey Measures

Participants completed an initial questionnaire regarding their demographics (age and gender), recent sleep (hours of sleep night before and hours since waking), caffeine, and nicotine use before eye-tracker data collection. To understand their subjective feelings of fatigue, participants also completed a survey at the end of the first 2 hours and the last 2 hours of their shift. Fatigue surveys items included perceived patient load (low, average, and high), perceived shift complexity (low, average, and high), a fatigue visual analog scale (FVAS)<sup>20</sup> physician task load (PTL, a 4-item, modified NASA-TLX),<sup>6</sup> and the Stanford Sleepiness Scale (SSS).<sup>21</sup>

### Eye-Tracking Measures

The *main outcome* measure was the eye metric fatigue score adapted from Khairat et al's definition<sup>5</sup> and derived for each participant as the difference between the PM and AM average whole fixation pupil with negative scores. Eye metrics included (1) average duration of whole fixations, (2) average peak velocity of saccades, (3) average amplitude of saccades, and (4) average whole fixation pupil diameter.

Implementation barriers were logged throughout the study to understand limitations in the use of eye tracking in routine clinical care.



**Fig. 1** Eye metric fatigue scores: total and stratified by gender and role. The top and bottom bars indicate the first and third quartile, respectively; the diamond indicates the mean; the horizontal line in the bars indicate the median; and vertical lines indicate minimum and maximum values.

### Data Sources

Eye metric data were collected using the Tobii pro fusion 250 Hz (Tobii Technology, Inc, Reston, VA, version 1.171, screen-based eye-tracking device set to a 250-Hz sampling rate [► Fig. 1]). To standardize measurement and control for potential temporal and environmental confounders, all observations were performed during a 7:00 AM to 3:00 PM shift at a single designated computer workstation during the first and last 2 hours of each shift. The workstation was calibrated before each recording session. The ED environment near the study computer had constant illumination with mean maximum light intensity 447.21 lux (standard deviation [SD]: 33.56) and minimum light intensity 424.25 lux (SD: 18.01). During recording time, clinicians were instructed to go about their shift as usual but, when they needed to use a computer (for EHR or other use), to only use the designated study workstation.

### Statistical Methods

The eye tracker captures data as long as the participant is in front of the system, regardless of whether or not their gaze is fixed on the EHR (e.g., communicating with other staff standing behind the screen). To avoid this as a potential confounder, data were manually reviewed to create multiple TOI representing reliably captured eye metrics; TOI's of less than 30 seconds were discarded. To further guarantee data integrity, the first and the last 5 seconds from each TOI was discarded.

Survey metric and data capture results were summarized using descriptive statistics. Survey and eye metrics were compared from the beginning to end of each shift using paired *t*-tests. Pearson correlation coefficient was calculated to explore bivariate associations between eye metric fatigue score and clinician-perceived fatigue. Based on those findings, a multivariable linear regression model was built to determine if associations persisted after controlling for clinician age, gender, and role. Statistical significance was set at  $p < 0.05$ , two-sided.

### Results

The sample included 15 participants with median age of 32 years (interquartile range 29–36); 5 (33%) were women, 9 (60%) were residents, 4 (27%) were attendings, and 2 (13%) were APRNs. Of the possible 60 hours of eye-tracking recording time, 46.46 hours of data were captured with an average of 3.10 (SD: 0.47) hours for each participant of which 1.88 (SD: 0.16) hours was determined to be valid for analysis.

The majority of participants indicated the following about their sleep, stimulant use, and patient load: 6 to 8 hours of sleep the night before their shift (73%), being awake for 2 hours before their shift started (67%), having a caffeinated beverage before the start of the shift 67%, no additional caffeinated beverages within 4 hours of the end of their shift (54%), no nicotine exposure within 24 hours of their shift (100%), low patient load (66%), and average patient complexity (80%).

### Survey Results from the Beginning to the End of the Shift

Both FVAS and PTL mean scores (SD) increased from the beginning to the end of the shift (FVAS 3.7 [1.7] to 4.6 [2.1],  $p = 0.04$  and PTL 97.6 [68.1] to 154.3 [76.8],  $p = 0.01$ , ► Table 1). Stratified by role, attendings experienced the largest changes in FVAS and PTL 1.25 (1.89), 78.75 (50.23) comparing to residents 0.78 (1.48), 53.33 (85.95), and APRNs 0.50 (0.71), 27.50 (31.82). The SSS score remained low from the beginning to the end of the shift (2.27 [1.22] to 2.4 [1.45],  $p = 0.61$ ).

### Changes in Eye-Tracking Metrics from the Beginning to the End of the Shift

There were no significant differences in eye metrics from the beginning to the end of the shift. Eye metric fatigue score (displayed stratified by gender and role in ► Fig. 2) was associated with the change in the PTL score ( $r = 0.59$ , 95% confidence interval [CI] [0.85, 0.11],  $p = 0.02$ ) (► Table 2) but

**Table 1** Survey and eye metrics compared from the start and end of the shift

Survey metrics	Start of shift	End of shift	p-Value
	Mean (SD)	Mean (SD)	
FVAS	3.73 (1.67)	4.6 (2.1)	0.04
PTL	97.6 (68.1)	154.3 (76.8)	0.01
SSS	2.27 (1.22)	2.4 (1.45)	0.61
Eye metrics			
Average pupil diameter (fixation filter, mm)	2.65 (0.35)	2.64 (0.37)	0.84
Average duration of whole fixations, ms	260.45 (38.67)	263.67 (42.26)	0.65
Average peak velocity of saccades °/s	163.17 (13.75)	160.36 (17.14)	0.39
Average amplitude of saccades, mm	4.47 (0.42)	4.49 (0.53)	0.86

Abbreviations: FVAS, fatigue visual analog scale; PTL, physician task load; SD, standard deviation; SSS, Stanford Sleepiness Scale.



**Fig. 2** Eye tracker setup.

not FVAS or SSS. This association persisted after adjusting for age, gender, and role in multivariable analysis, with every 10-point increase in PTL there was a 0.02-point increase in pupillometry-based fatigue score (95% CI [0.003, 0.000],  $p = 0.04$ ).

**Discussion**

this feasibility study of eye metric technology in routine emergency care, it was feasible to collect eye metric

data for 15 emergency clinicians in their real-world clinical setting at fixed times and a fixed workstation with even light intensity. Physiological fatigue was not detected with eye-tracking technology. However, all eye metrics did trend in the hypothesized direction: smaller pupil size,<sup>22–24</sup> lower peak saccade speed,<sup>15</sup> and higher whole fixation duration<sup>17</sup> and saccade amplitude<sup>18</sup> at the end of the shift.

It is possible that the participants did not experience physiologic fatigue given that the majority of participants reported being well-rested, having ingested caffeinated



**Table 2** Pearson correlation coefficient (*r*) between eye metric fatigue score and clinician-perceived fatigue survey metrics

Fatigue score compared with	<i>r</i>	95% CI	<i>p</i> -Value
Change in FVAS	−0.04	[−0.54, 0.48]	0.89
Change in PTL	0.59	[0.11, 0.85]	0.02
Change in SSS	0.21	[−0.34, 0.65]	0.45
FVAS at the end of the shift	0.22	[−0.33, 0.66]	0.44
PTL at the end of the shift	0.23	[−0.32, 0.67]	0.41
SSS at the end of the shift	0.07	[−0.46, 0.56]	0.81

Abbreviations: CI, confidence interval; FVAS, fatigue visual analog scale; PTL, physician task load; SSS, Stanford Sleepiness Scale.

**Table 3** Implementation barriers

Challenges	Recommendations
Temporary hardware malfunctions and delays in data transfer within the hardware system	Optimize hardware performance/speed, use gap fill-in algorithm
Noise from minor eye movement (microsaccades and eye tremors are substantial in high-frequency eye tracker [250 Hz])	Use lower frequency eye tracker (60 and 120 Hz) based on the eye metric of interest
Improper fixation filter calibration	Adjust software settings to construct an eye model only when both eyes are detected
Eye metrics captured but subject not engaged in EHR work	Consider different clinical setting
Inconsistent eyeglass wearing gives false reading of eye location	Ensure consistent eyeglasses use or exclude participants who wear glasses
– Obstacle between participant and eye tracker such as paper, EKG, surgical mask, blinks, eye glass fogging. – Suboptimal positioning, such as too close, head tilt, or out of field	Use shorter recording period, gap fill-in algorithm, or discard recording <30 s in duration
Potential bias from manual interval assignment	Need eye tracker software update to include a feature that automates creating TOI intervals if no eye detection for a set period of time

Abbreviations: EHR, electronic health record; EKG, electrocardiogram; TOI, times of interest.

beverages, low patient load, and average patient complexity, or that physiological fatigue would have been captured with eye-tracking technology in larger sample size, with a more focused recording of fixed tasks, or recording throughout the entire shift.<sup>5</sup>

### Lessons Learned and Recommendations to Overcome Implementation Barriers

Implementation facilitators we encountered that others might leverage in future work include the choice of a screen-based device that is simple to install, does not require batteries, and is less obtrusive than a head-mounted device. Also, the ED setting allowed a single, fixed workstation with stable lighting. Implementation barriers (→Table 3) that posed challenges to data validity and the feasibility of using eye-tracking technology to detect clinician fatigue during routine clinical care were related to inconsistent lighting, temporary hardware malfunctions, measurement noise from minor eye movements (such as microsaccades and eye tremors are factors that contribute to measurement noise in eye tracking<sup>25</sup>), and data loss from participant positioning, mask-wearing, eyeglass-wearing, and gaze.

Data loss can also occur due to other factors, including delays in data transfer, temporary malfunctions of the hardware, time out issues, and reflections from eyeglasses that make it impossible for the device to identify the eyes' geometrical location.<sup>26</sup> In these cases, lost data can affect the reliability of the fixation algorithm, resulting in the misclassification of one fixation as two separate fixations. The application of a gap-fill algorithm can mitigate this issue.<sup>27</sup>

The fixation filter also has a limited ability to detect right and left eye positions. For example, if both eyes are detected during calibration, but only one is detected during recording, the data detected in calibration will be used to calculate the gaze for both eyes. Since the left and right eyes are not identical, this will result in an offset of the constructed eye model.

### Limitations

This study was subject to several limitations. As a pilot study, the sample size may have had too few subjects experiencing physiologic fatigue or to control for other confounding factors. The sample included a younger population. Given physiological differences in eye metrics by age,<sup>28</sup> our findings should be

interpreted with caution in older clinicians. Also, due to the small sample size of this pilot study, differences between survey metrics should be interpreted with caution and warrant further investigation. Unlike a simulated environment, periods of missing data capture were expected as participants were involved in patient care or completing noncomputer work making it difficult to interpret the rate of reliable eye-tracking data capture. Eye-tracking metrics could have been less reliable due to masking and eye protection as part of ongoing COVID-19 pandemic precautions. Inconsistent eyeglasses wearing may have given false readings of eye location. Furthermore, the functionality of the eye-tracking technology itself limited the environment in which the study could be conducted: only permitting minimal variability in lighting, location, and participant positioning. Stimulant use was self-reported without confirmatory drug screening. The presence of the researcher during data collection also may have subjected study data to the Hawthorne effect.<sup>29</sup>

## Conclusion

Eye-tracking metrics did not change significantly from the beginning to the end of an 8-hour ED shift. It is unclear whether this was due to confounding factors, a limitation of the technology, study design, sample size, or an absence of consistent physiological fatigue over the course of an 8-hour ED shift. Further research is needed to determine the role for eye-tracking technology to measure clinician fatigue in routine clinical care.

### Author Contributions

B.K. and E.R.M. conceived of the work. All authors designed the study. B.K. acquired the data. B.K., B.N., and E.R.M. drafted the initial manuscript. All authors analyzed the data and revised the manuscript and approved the final version submitted for publication. B.K. takes responsibility for all aspects of the work.

### Funding

None.

### Conflict of Interest

E.R.M. reports receiving grants from the American Medical Association and National Institute on Drug Abuse that are unrelated to this award. The other authors declare that they have no conflicts of interest in the research.

### Acknowledgments

We would like to thank the attendings, residents, and advanced nurse practitioners who participated in this study as well as Charlotte Newman for providing personnel support.

## References

- 1 Thomas NK. Resident burnout. *JAMA* 2004;292(23):2880–2889
- 2 Caldwell JA, Caldwell JL, Thompson LA, Lieberman HR. Fatigue and its management in the workplace. *Neurosci Biobehav Rev* 2019; 96:272–289
- 3 Melnick ER, Dyrbye LN, Sinsky CA, et al. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clin Proc* 2020;95(03): 476–487
- 4 Melnick ER, Harry E, Sinsky CA, et al. Perceived electronic health record usability as a predictor of task load and burnout among US physicians: mediation analysis. *J Med Internet Res* 2020;22(12): e23382
- 5 Khairat S, Coleman C, Ottmar P, Jayachander DI, Bice T, Carson SS. Association of electronic health record use with physician fatigue and efficiency. *JAMA Netw Open* 2020;3(06):e207385
- 6 Harry E, Sinsky C, Dyrbye LN, et al. Physician task load and the risk of burnout among US physicians in a national survey. *Jt Comm J Qual Patient Saf* 2021;47(02):76–85
- 7 Carayon P, Wetterneck TB, Alyousef B, et al. Impact of electronic health record technology on the work and workflow of physicians in the intensive care unit. *Int J Med Inform* 2015;84(08): 578–594
- 8 Patterson PD, Weaver MD, Fabio A, et al. Reliability and validity of survey instruments to measure work-related fatigue in the emergency medical services setting: a systematic review. *Prehosp Emerg Care* 2018;22(Suppl 1):17–27
- 9 Dawson D, Searle AK, Paterson JL. Look before you (s)leep: evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. *Sleep Med Rev* 2014;18(02):141–152
- 10 O'Reilly-Shah VN. Factors influencing healthcare provider respondent fatigue answering a globally administered in-app survey. *PeerJ* 2017;5:e3785–e3785
- 11 Ashraf H, Sodergren MH, Merali N, Mylonas G, Singh H, Darzi A. Eye-tracking technology in medical education: a systematic review. *Med Teach* 2018;40(01):62–69
- 12 Morad Y, Lemberg H, Yofe N, Dagan Y. Pupillography as an objective indicator of fatigue. *Curr Eye Res* 2000;21(01):535–542
- 13 Szabadi E. Functional neuroanatomy of the central noradrenergic system. *J Psychopharmacol* 2013;27(08):659–693
- 14 van der Wel P, van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: a review. *Psychon Bull Rev* 2018; 25(06):2005–2015
- 15 Prsa M, Dicke PW, Thier P. The absence of eye muscle fatigue indicates that the nervous system compensates for non-motor disturbances of oculomotor function. *J Neurosci* 2010;30(47): 15834–15842
- 16 Di Stasi LL, McCamy MB, Macknik SL, et al. Saccadic eye movement metrics reflect surgical residents' fatigue. *Ann Surg* 2014;259 (04):824–829
- 17 Zargari Marandi R, Madeleine P, Omland Ø, Vuillerme N, Samani A. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Sci Rep* 2018;8(01):13148
- 18 Morris TL, Miller JC. Electrooculographic and performance indices of fatigue during simulated flight. *Biol Psychol* 1996;42(03): 343–360
- 19 Finke C, Pech LM, Sömmmer C, et al. Dynamics of saccade parameters in multiple sclerosis patients with fatigue. *J Neurol* 2012; 259(12):2656–2663
- 20 Khanna D, Pope JE, Khanna PP, et al. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. *J Rheumatol* 2008;35(12):2339–2343
- 21 Herscovitch J, Broughton R. Sensitivity of the Stanford Sleepiness Scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep* 1981;4(01):83–91
- 22 de Rodez Benavent SA, Nygaard GO, Harbo HF, et al. Fatigue and cognition: Pupillary responses to problem-solving in early multiple sclerosis patients. *Brain Behav* 2017;7(07):e00717
- 23 Körber M, Cingel A, Zimmermann M, Bengler K. Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manuf* 2015;3:2403–2409

- 24 Steinhauer SR, Siegle GJ, Condray R, Pless M. Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *Int J Psychophysiol* 2004;52(01):77–86
- 25 Duchowski AT, Duchowski AT. *Eye Tracking Methodology: Theory and Practice*. Springer; 2017
- 26 Munn SM, Stefano L, Pelz JB. Fixation-identification in dynamic scenes: comparing an automated algorithm to manual coding. *Proceedings of the 5th symposium on Applied perception in graphics and visualization*; 2008; Los Angeles, California
- 27 Komogortsev O, Gobert DV, Dai Z. Classification algorithm for saccadic oculomotor behavior. (Report No. TXSTATE-CS-TR-2010-23). Texas State University-San Marcos, Department of Computer Science;2010
- 28 Dowiasch S, Marx S, Einhäuser W, Bremmer F. Effects of aging on eye movements in the real world. *Front Hum Neurosci* 2015;9:46
- 29 McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol* 2014;67(03):267–277