UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

Digital biomarkers in Alzheimer's disease

Light reflex pupillometry and actigraphy for early detection and prognosis

PhD Thesis

Mathias Holsey Gramkow

DANISH DEMENTIA RESEARCH CENTRE

REGION H Rigshospitalet

This thesis was submitted to the Faculty of Health and Medical Sciences, University of Copenhagen on 5th of July, 2025

The front cover illustration is adapted from the front page of Nature Reviews Neurology, Volume 20, Issue 12, December 2024, inspired by *The Digitized Memory Clinic* (Gramkow, Waldemar & Frederiksen, **Nature Reviews Neurology**, 2024) with permission from the copyright owner.

PhD Thesis

DIGITAL BIOMARKERS IN ALZHEIMER'S DISEASE: LIGHT REFLEX PUPILLOMETRY AND ACTIGRAPHY FOR EARLY DETECTION AND PROGNOSIS

Mathias Holsey Gramkow, MD
Danish Dementia Research Centre, Department of Neurology,
Copenhagen University Hospital - Rigshospitalet

From the Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital -Rigshospitalet and the Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark:

Principal Supervisor
Associate Professor
Kristian Steen Frederiksen, MD, PhD

Primary co-supervisor
Professor
Steen Gregers Hasselbalch, MD, DMSc

Co-supervisor
Professor
Gunhild Waldemar, MD, DMSc

Assessment committee
Chairperson
Professor
Christina Rostrup Kruuse, MD, PhD, DMSc
Department of Neurology, Copenhagen University Hospital – Herlev and Gentofte
Department of Brain- and Spinal Cord Injury, Copenhagen University Hospital – Rigshospitalet
Department of Clinical Medicine, University of Copenhagen

Assessors
Professor
Jakob Udby Blicher, MD, PhD
Department of Neurology, Aalborg University Hospital
Department of Clinical Medicine, Faculty of Medicine, Aalborg University

Professor
Sebastiaan Engelsborghs, MD, PhD
Reference Center for Biological Markers of Dementia (BIODEM), Laboratory of Neurochemistry and Behavior, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium
Department of Neurology, UZ Brussel

Table of contents

List of manuscripts

This PhD thesis consists of a synopsis and the following four manuscripts (Appendix A):

STUDY I[1] Gramkow, M. H., Clemmensen, F. K., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2024). Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort. *Journal of the Neurological Sciences*, *456*(March), 122856. https://doi.org/10.1016/j.jns.2023.122856

STUDY II[2] Gramkow, M. H., Clemmensen, F. K., Sjælland, N. S., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2024). Diagnostic performance of light reflex pupillometry in Alzheimer's disease. Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring, 16(3), 1–11. https://doi.org/10.1002/dad2.12628

STUDY III[3] Gramkow, M. H., Brink-Kjær, A., Clemmensen, F. K., Sjælland, N. S., Waldemar, G., Jennum, P., Hasselbalch, S. G., & Frederiksen, K. S. (2025). Diagnostic performance of actigraphy in Alzheimer's disease using a machine learning classifier – a cross-sectional memory clinic study. Alzheimer's Research & Therapy 2025 17:1, 17(1), 1–12. https://doi.org/10.1186/S13195-025-01751-5

STUDY IV Gramkow, M. H., Clemmensen, F. K., Lindberg, U., Law, I., Henriksen, O. M., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S.: Prognostic value of light reflex pupillometry in Alzheimer's disease – a longitudinal cohort study (***Accepted*** *June 22, 2025, in Alzheimer's Research and Therapy)*

Preface and acknowledgements

The work included in this PhD thesis was carried out at the Danish Dementia Research Centre (DDRC) with its affiliated memory clinic at Rigshospitalet, Copenhagen. We were grateful to obtain specific grants from the following funding bodies: Alzheimerforeningens forskningsfond, A.P. Møller og Chastine McKinney Møllers Fond til Almene Formaal, Ellen Mørchs Fond, Kong Christian den Tiendes Fond, and Højesteretssagfører L. Zeuthens Mindelegat. The Danish Dementia Research Centre is supported by the Danish Ministry of Health.

There is no 'I' in PhD. As a consequence, several supporting individuals helped these studies come to fruition, and they all deserve mention here.
First and foremost, thanks should be given to my principal supervisor, associate professor Kristian Steen Frederiksen, who stands out for his unwavering support and belief in me as a researcher. Thank you for being an inspiration by keeping a firm focus on the clinically important problems while maintaining a very high scientific standard and originality factor. I am sure we will have many more years of scientific collaboration, which I look forward to.
To professor Steen Gregers Hasselbalch, for your words of wisdom and guidance throughout the project. Your immense knowledge of dementia disorders has not only been integral to the studies but also to my ongoing education as a clinician and dementia researcher. To professor Gunhild Waldemar, deserving of gratitude not only for your guidance on this thesis but also for creating a unique work environment, which I treasure daily.
To Frederikke Kragh Clemmensen, my co-captain on our clinical projects. Without you, I am certain I would not be writing this thesis, as your help in planning and conducting these studies was vital. Being able to share the ups and downs during this demanding process has been a great support throughout. Thank you.
Also, thanks should be given to Nikolai Sulkjær Sjælland for your help with recruiting patients for the IPATCH-AD study.

I would like to thank the collaborators on this journey. Assistant professor Andreas Brink-Kjær, for your help with analyzing actigraphy data, creating cool figures, and helping make sense of our machine learning models. To professor Poul Jennum for your unique input stemming from a deep knowledge in sleep and circadian function, and also for providing the opportunity for an inspiring change of environment. To professor Ian Law and Otto Mølby Henriksen for your collaboration on PET/MR data, both the practical and scientific aspects. To Ulrich Lindberg for your help with neuroimaging data.
To senior lecturer Casper de Boer and the Alzheimer Center in Amsterdam. Thank you for hosting me for a short yet insightful stay at your impressive center.
To all my colleagues at DDRC (especially office 1!): thank you for being the much needed and highly regarded social backdrop to my scientific work. It would not have been the same without you.
To all the doctors and nurses at the memory clinic: thank you for your help in recruiting patients for the studies. Last, but not least, I am also extremely grateful to all the patients and caregivers without whom these studies could not have been conducted.
Also, I would like to thank my friends and family for lending an ear and asking relevant questions, helping me sharpen my scientific communication. To my parents, thank you for always offering me support and cheering me on throughout my educational journey. To my amazing kids, Oskar and Filippa, thank you for just being yourselves and brightening every day. To my wonderful Marie, thank you for your loving support throughout and for taking care of our children (also the one not yet born at the time of writing this preface!), especially when my studies demanded traveling abroad. Our family is the light that illuminates my work desk.

List of abbreviations

[^{18}F]FDG	^{18}F-fluorodeoxyglucose
ACE	Addenbrooke's Cognitive Examination
AD	Alzheimer's disease
ADL	Activities of daily living
AUROC	Area under the receiver operating characteristics curve
DHT	Digital health technology
DLB	Dementia with Lewy bodies
DRN	Dorsal raphe nucleus
CDR	Clinical Dementia Rating
CI	Confidence interval
CSF	Cerebrospinal fluid
CV	Cross-validation
CVD	Cerebrovascular disease
CT	Computed tomography
ELISA	Enzyme-linked immunosorbent assay
EWn	Edinger-Westphal nucleus
EWpg	Pre-ganglionic part of Edinger-Westphal nucleus
HC	Healthy control
ipRGC	Intrinsically photosensitive retinal ganglion cells
ICC	Intra-class correlation coefficients
LMM	Linear mixed model
LC	Locus coeruleus
MRI	Magnetic resonance imaging
MMSE	Mini-Mental State Examination
MCI	Mild cognitive impairment
MRI	Magnetic resonance imaging
ML	Machine learning
mRMR	Minimum redundancy maximum relevance

NIA-AA	National Institute of Aging-Alzheimer's Association
OPN	Olivary pretectal nucleus
PET	Positron emission tomography
qLRP	Quantitative light reflex pupillometry
SoB	Sum-of-Boxes
VCD	Vascular cognitive disorder

English summary

As disease-modifying therapies are becoming available worldwide and the prevalence of Alzheimer's disease (AD) is growing, an imperative is building to provide easily accessible biomarkers that can enable remote monitoring for detection and follow-up, thereby easing diagnosis and prognostication. Digital biomarkers, which are bodily characteristics measured by, e.g., sensors in digital health technologies, could form part of the solution to this challenge. In this PhD thesis, the results of four papers describing the test-retest reliability and diagnostic and prognostic properties of two such promising digital biomarkers, namely light reflex pupillometry and actigraphy, are presented.

In **study I**, we investigated the test-retest reliability and short-term variability of quantitative light reflex pupillometry (qLRP) in a cohort of mixed memory clinic patients. We included 40 patients under suspicion of neurodegenerative disease, regardless of their final diagnosis. We demonstrated excellent reliability of most light reflex parameters and a limited short-term variability, highlighting the applicability of qLRP in a relevant setting, where patients with AD are evaluated. This study paved the way for exploring qLRP further in neurodegenerative diseases, such as AD.

In **study II**, we explored the diagnostic properties of qLRP in early Alzheimer's disease (AD) by investigating the classification accuracy in relevant testing scenarios: against healthy controls, and against the prevalent dementia disorders dementia with Lewy bodies (DLB) and patients with cerebrovascular disease (CVD). We found that qLRP could distinguish between groups with moderate accuracy, proving the concept of qLRP-assisted diagnosis in a real-world memory clinic setting. These results suggested the usefulness of qLRP for differentiating dementia disorders and as a possible tool for screening purposes, given its easy application.

In **study III**, we explored actigraphy for differential diagnosis of early AD, by creating activity profiles from 7-day recordings of real-life behavior outside the clinic setting.

Comparing the activity profiles of patients with AD with relevant diagnostic groups (DLB and CVD) and by applying a machine learning algorithm, we demonstrated a promising utility of actigraphy for differentiating dementia disorders with high accuracy. These results justify a further exploration of actigraphy in a memory clinic setting.

In **study IV**, we leveraged qLRP for prognosis of early AD by assessing baseline and short-term (3-month changes) in pupillary responses and their association with clinical progression in a cohort of early AD patients. We found that short-term decreases in the resting pupillary size were associated with a higher risk of clinical progression and that relative changes in pupil size following a light stimulus were indicative of neurodegenerative progression on [^{18}F]FDG-PET and cognitive decline as measured with the Mini-Mental State Examination. We showed that qLRP, in addition to diagnostic properties, may also reflect disease progression, possibly related to early brainstem changes related to arousal states that are reflected in the pupillary function.

Together, our results showed that qLRP and actigraphy are promising digital biomarkers for assessing early diagnosis and prognosis of AD in an easily applicable manner. Future studies should evaluate the markers in larger cohorts, applying study designs that limit selection bias and make use of external validation.

Dansk resumé | Danish summary

Der er et forventet stigende behov for lettilgængelige biomarkører ved Alzheimers sygdom (AD), efterhånden som målrettet terapi gøres tilgængelig, særligt set i lyset af en stigende prævalens af sygdommen. Digitale biomarkører, som er kropslige karakteristika målt ved hjælp af sensorer fra digital sundhedsteknologi, kan udgøre en del af løsningen på denne udfordring. I denne ph.d.-afhandling præsenteres fire artikler, der beskriver test-gentest-pålidelighed samt diagnostiske og prognostiske egenskaber i AD af to lovende digitale biomarkører, nemlig lys-refleks pupillometri og aktigrafi.

I **studie I** undersøgte vi test-retest pålidelighed og korttidsvariabiliteten af kvantitativ lys-refleks pupillometri (qLRP) i en kohorte af uselekterede patienter fra en hukommelsesklinik. Vi demonstrerede, at qLRP havde en tilfredsstillende pålidelighed for de fleste parametre, hvilket baner vejen for yderligere udforskning af teknikken i neurodegenerative sygdomme, herunder ved AD.

I **studie II** undersøgte vi de diagnostiske egenskaber ved qLRP i tidlig AD ved at vurdere metodens evne til at klassificere sygdomsgrupper i relevante testscenarier: sammenlignet med raske kontrolpersoner samt patienter med de to næst hyppigst forekommende demenssygdomme, nemlig demens med Lewy legemer (DLB) og cerebrovaskulær sygdom (CVD). Vi fandt, at qLRP robust kunne skelne mellem grupperne med moderat præcision, hvilket bekræftede anvendeligheden i en hukommelsesklinik. Resultaterne antydede, at qLRP kan være nyttig til differentiering af demenssygdomme og muligvis som et screeningsværktøj, da metoden er let at anvende.

I **studie III** undersøgte vi aktigrafi til diagnosticering af tidlig AD ved at se på aktivitetsprofiler baseret på 7-dages optagelse af bevægeadfærd i patienternes hjemlige omgivelser. Vi sammenlignede aktivitetsprofilerne for gruppen med AD med patienter med DLB og CVD samt raske kontroller. Ved at anvende maskinlæring til at fremhæve de mest betydende elementer af aktivitetsprofilerne så vi, at aktigrafi fungerede med høj præcision

til differentiering af demenssygdomme (AD, DLB og CVD). Disse resultater understøtter yderligere undersøgelser af aktigrafi i en hukommelsesklinik, f.eks. i opfølgningsstudier.

I **studie IV** undersøgte vi qLRPs prognostiske egenskaber i tidlig AD ved at evaluere statiske mål og kortsigtede (3-måneders) ændringer i pupillens reaktion på lys og så på sammenhængen med klinisk progression i en kohorte af patienter med tidlig AD. Vi fandt her, at en formindskelse af pupillens hvilestørrelse var forbundet med en øget risiko for klinisk progression. Vi tolkede dette som et udtryk for at vågenheden var påvirket tidligt hos de patienter, der oplevede klinisk progression inden for et år. Disse resultater understøtter en videre eksploration af biomarkøren som prognostisk markør for AD.

Resultaterne af disse fire studier understreger, at qLRP og aktigrafi er lovende metoder til vurdering af tidlig diagnose og prognose ved AD på en let anvendelig måde. Selvom de to markører aldrig kan stå alene og bør kombineres med traditionelle metoder, udgør de et potentiel nyttigt bidrag til fremtidens hukommelsesklinik. For at forstå dette bidrag til fulde bør fremtidige studier evaluere biomarkørerne i større kohorter med studiedesigns, der begrænser selektionsbias og søger at validere resultaterne i uafhængige kohorter.

1. Background

1.1 The future of Alzheimer's disease management

The estimated prevalence growth of Alzheimer's disease (AD)[4], a devastating brain disorder and the main cause of dementia[5], is considered a major public health priority[6]. As new disease-modifying therapies are now approved for use[7,8], there is an expected imminent increase in referrals to memory clinics for evaluation. Consequently, this puts pressure on healthcare systems to deliver a timely diagnosis and prognosis[9]. The standard of present-day diagnostics and management of patients with dementia is time- and cost-consuming[10], and given the high price for new disease-modifying treatments[11], this may inadvertently lead to proportionally fewer patients being diagnosed and treated due to high demand[12].

A promising add-on to current diagnostic efforts is digital biomarkers[13] (**Box 1** and **Figure 1**), which are biomarkers obtained through the use of digital health technology, such as sensors embedded in wearable or portable devices[12,14]. Digital biomarkers, by virtue of their accessibility, can be woven seamlessly into the current paradigm using artificial intelligence approaches, which have recently been envisioned in a Perspective piece for a digitized memory clinic[12], which is further elaborated upon below. These digital biomarkers are thought to shed light on hitherto unexplored areas of unobtrusive monitoring and biological evaluation, where clinical examination, brain scans and fluid analysis have been the mainstay of diagnosis for common disorders such as AD[12]. Digital biomarkers offer an easily applied, point-of-care assessment, which seems advantageous given the high associated burden of testing in individuals with a dementia disorder[12,15]. For AD, this seems ever more relevant as the pressure is building on clinics to provide novel treatment at biweekly intervals with regular clinical follow-up during and after treatment[7,11]. Tools that ease these tasks represent an urgent unmet need, which could help break down barriers to delivering timely intervention and management. Consequently, attention has turned to digital biomarkers.

Box 1. The Food and Drug Administration's definition of a digital biomarker. Italics indicate a quote from Vasudevan et al.[16].

Digital biomarker:

"a characteristic or set of characteristics, collected from digital health technologies, that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions."

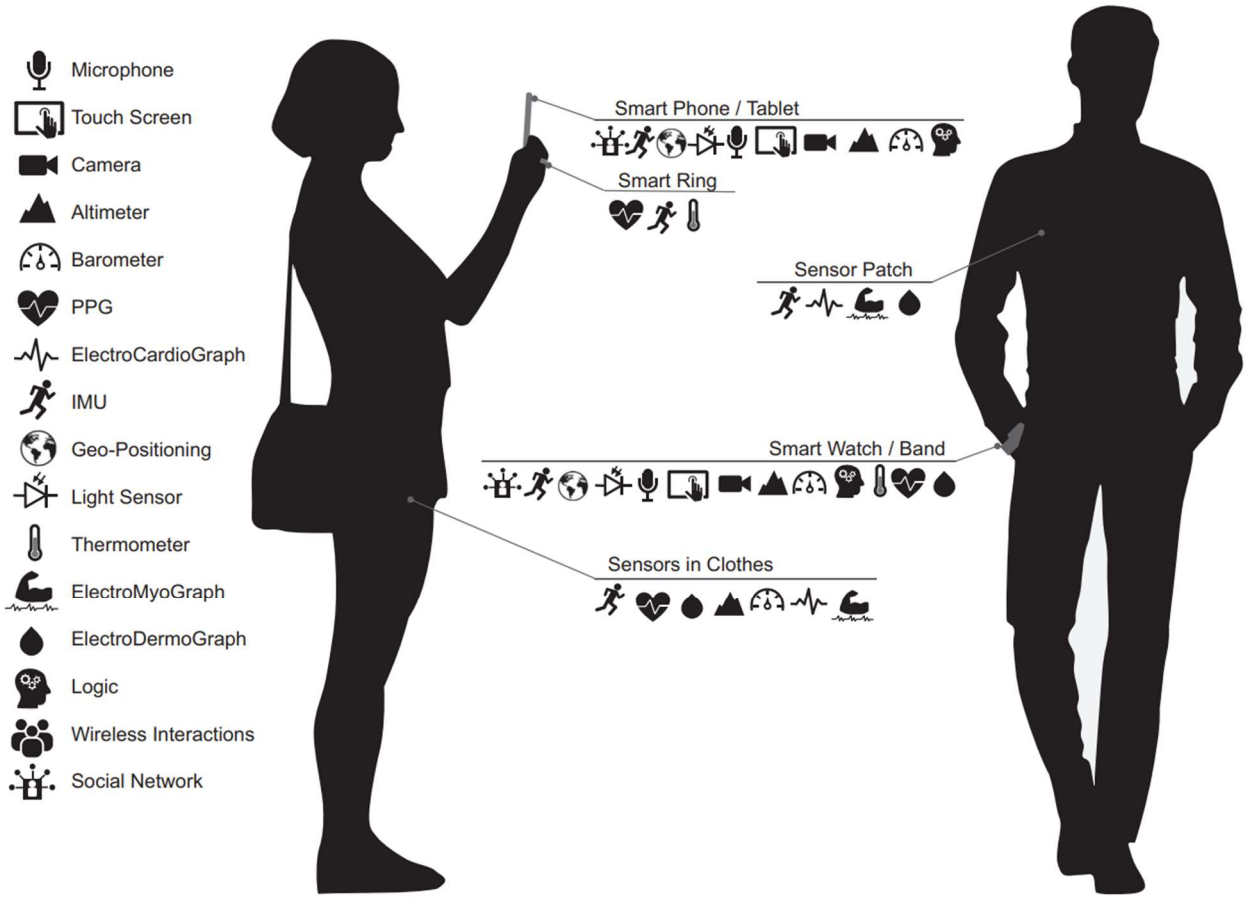
1.2 A digitized memory clinic

As several challenges converge for the future of dementia management, such as an aging population and a decreasing labor force[17], an avenue for rethinking the current practice arises. Incorporating digital health technology (DHT) and derived digital biomarkers in the "brick-and-mortar" memory clinic could alleviate some of the expected diagnostic burden[12]. Furthermore, digital biomarkers that are captured outside the clinic setting could show higher face validity, an example being actigraphy-measured versus proxy-rated questionnaires of daily activity levels, where actigraphy captures more consistent data[18,19]. In addition, they may allow for more dynamic assessment of pathological changes given their capability of collecting data continuously. One caveat to digital biomarkers, however, is the large amount of data, termed *big data*, which requires computational resources. Fortunately, recent technological advances have made it possible to handle *big data* efficiently, using machine learning and artificial intelligence solutions[12,19].

In contrast to the large potential of digital biomarkers, the uptake of digital tools in memory clinics is restricted to highly specialized settings[20], most likely due to a lack of studies that apply the technology in clinically relevant settings, where it may be envisioned to be applied. Further, studies that inform on the diagnostic and prognostic properties of digital biomarkers are essentially missing from the literature[15]. To bridge these gaps in knowledge, we set forth to extensively evaluate two such digital tools that had shown promise in the early AD in a real-world clinical setting.

In the following sections, the two digital tools in question, quantitative light reflex pupillometry (qLRP) and actigraphy, are elaborated upon, and the context for their use is given.

Figure 1. An overview of proposed DHTs and digital biomarkers (left panel) for Alzheimer's disease. Reprinted without modification from Kourtis, L. C., Regele, O. B., Wright, J. M., & Jones, G. B. (2019). Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *npj Digital Medicine*, *2*(1), 9. https://doi.org/10.1038/s41746-019-0084-2 © CC BY 4.0., copyright license at https://s100.copyright.com/AppDispatchServlet?title=Digital%20biomarkers%20for%20Alzheimer%E2%80%99s%20disease%3A%20the%20mobile%2Fwearable%20devices%20opportunity&author=Lampros%20C.%20Kourtis%20et%20al&contentID=10.1038%2Fs41746-019-0084-2&copyright=The%20Author%28s%29&publication=2398-6352&publicationDate=2019-02-21&publisherName=SpringerNature&orderBeanReset=true&oa=CC%20BY [14].



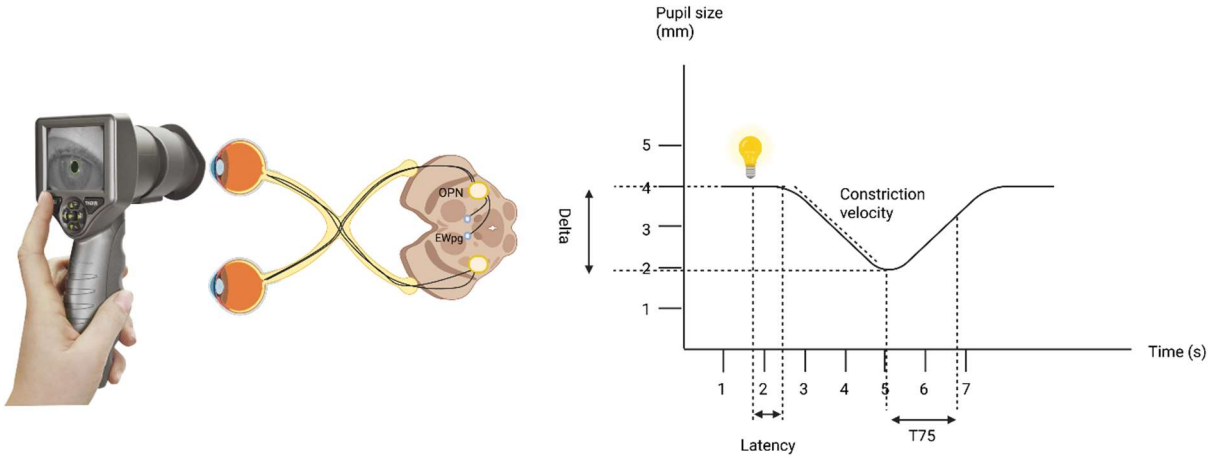
1.3 Quantitative light reflex pupillometry

The study of the pupillary light reflex has a long history within medicine, with the first descriptions written down by the Persian physician Rhazes of Baghdad (850-932 CE)[21]. The actions of the pupil are affected by emotion[22] and are part of wide-reaching physiological indices, such as arousal levels[23,24], while also being a central autonomic reflex[25]. Recent efforts to use pupillary responses as a way of studying the central nervous system have gained traction as a promising marker within the field of neurodegenerative diseases[26,27], in part due to studies describing pathology in the brain stem part of the reflex arc in, e.g., Alzheimer's disease[28–32].

The light reflex is governed mainly by the Edinger-Westphal nucleus (EWn) situated in the mesencephalic portion of the brain stem[33]. It receives a bilateral input from the olivary pretectal nuclei, giving rise to a consensual constriction of the pupil when light is shone only on one eye[33]. The signal to the pretectal nuclei is through the optic nerve and is initiated upon the activation of mainly rod cells in the inner layer of the retina. Recent studies have also highlighted an important role of intrinsically photosensitive retinal ganglion cells (ipRGCs) in initiating and maintaining a tonic response to especially blue light stimuli. Studies have shown that ipRGCs are vulnerable to Alzheimer's pathology, such as tau accumulation[34–38]. The EWn has a parasympathetic, preganglionic section (EWpg) through which efferents are sent to the ciliary ganglion, from where postganglionic neurons finally connect with the constrictor pupillae and the dilatator pupillae muscles. These eye muscles control the size of the pupil, the aperture of the eye, adjusting the amount of light that reaches the retina, thus representing a rudimentary mechanism for adjusting visual acuity under varying illumination conditions[39]. The EWn receives modulatory input from the locus coeruleus, recently identified as a key hub for AD-related neurodegeneration, which projects axons of noradrenergic neurons[40–42]. In addition, the dorsal raphe nucleus, which is involved in arousal processes and has been proposed as a possible early neurodegenerative hub in AD, modulates pupil size through activation of serotonergic neurons[43,44]. Finally, experimental studies have shown a modulatory effect of certain higher cortical areas, such as the frontal eye field situated in the frontal lobe and affected in the visual variant of AD[45] as well as input from the superior colliculus during movement tracking[46].

As a whole, pupillary action is controlled by complex, wide-reaching brain systems, which thereby lend themselves to easy investigation with quantitative pupillometry, where the pupil action is recorded continuously in response to various stimuli such as light (**Figure 2**).

Figure 2. Graphical illustration of the neuroanatomical pathways in the light reflex and the hand-held quantitative light reflex pupillometry technique (PLR-3000, NeurOptics®). OPN = olivary pretectal nucleus, EWpg = postganglionic part of the Edinger Westphal nucleus. Reprinted and modified with permission from the graphical abstract by Gramkow et al.[1]



As several brain areas that are involved in controlling and modulating the pupil, such as the EWn, locus coeruleus and dorsal raphe nucleus, have been shown to undergo neurodegenerative changes in AD[30,31,42,43,47], pupil function could possibly represent a diagnostic marker for AD (see **Table 1**). From a prognostic perspective, brain stem changes early in the neurodegenerative cascade have been described recently[42,47] in areas such as the previously mentioned dorsal raphe nucleus and locus coeruleus. These changes, consisting of tau protein accumulation, precede transentorhinal deposition in AD[43,48], and therefore, early pupillary changes may be prognostic of future cognitive decline.

Table 1. qLRP metrics calculated by the PLR-3000 Pupillometer (NeuroOptics®) and a non-exhaustive list of the physiological interpretation of each marker as well brain areas associated with the marker. The table is adapted from study IV.

Light reflex metric	**Physiological interpretation and associated brain areas**
Resting pupillary diameter	An indicator of autonomic balance[33]. Can measure cognitive strain[40,49] and arousal[44]. Involved brain areas: LC[40], DRn[43], EWn[30,31,50], hypothalamus, insula, frontal eye field, colliculus superior.
Peak constriction diameter	Can be interpreted with the resting pupillary diameter, giving the delta value (see below). Involved brain areas: EWn, pretectal olivary nucleus
Delta, relative pupillary change	Reflects an overall responsiveness of the pupil. Modulation from the frontal eye field (Brodmann area 8). Involved brain areas: EWn, frontal eye field, ciliary ganglion
Latency	Representing the overall nerve conductivity of the reflex arc. Delays could indicate demyelination[51]. Involved brain areas: optic nerve, chiasm and tract, pretectal brain area.
Constriction velocity	An indicator of EWn function, as constriction mainly arises when EWn neurons fire[33]. Involved brain areas: EWn, ciliary ganglion
Maximum constriction velocity	Same as for *constriction velocity*, but could indicate synchronous firing, i.e. robustness, of the EWn.
Average dilation velocity	An indicator of sympathetic recovery dynamics[33,52]. Involved brain areas: Hypothalamus, sympathetic spinal pathways, superior cervical ganglion.
T75	Same as for average dilation velocity, although is less reliably measured according to[1].

Abbreviations: EWn: Edinger-Westphal nucleus, LC: Locus coeruleus, DRn: Dorsal raphe nucleus.

The first studies on the light reflex in AD using quantitative pupillometry were fueled by early pathological changes found in the EWn by Scinto et al.[53,54] and later corroborated by Mavroudis et al.[31]. The detection of distinct neuropathological changes representing hallmarks of AD pathology, namely neurofibrillary tangles and amyloid plaques, was hypothesized to influence the behavior of the light reflex. In the following years, studies emerged that utilized video pupillometry to capture the light reflex in individuals with AD and found convincing differences[55,56], which were later corroborated[29]. Most of these studies applied high-frequency stationary camera systems that had been developed in laboratory settings to capture more minute changes of the pupil[57,58]. As this setup translates poorly to a clinical setting, hand-held pupillometry emerged as a more useful bedside assessment tool[59]. As measurements can be made within a few seconds, with little to no discomfort associated with the procedure[1], the technique could be envisioned to blend seamlessly into a memory clinic[2]. It is also a low-cost application, as a 2020 National Institute for Health and Care Evaluation (NICE) Medtech briefing calculation showed, at a measurement cost of 2.50 £ / day[60]. Also, given the relative ease of obtaining measurements, serial assessment therefore seems highly feasible, whereby dynamic patterns of autonomic and brain stem function may be appreciated.

Previous studies on qLRP have focused on a cross-sectional comparison of healthy individuals and patients with AD[55–57,61–64], with one study comparing patients with AD and Parkinson's disease[65]. What was mainly found was a slowing of the contractile phase upon a light stimulus, termed a "sluggish" response[29]. This was mainly found for patients with clinical AD, but studies applying qLRP in individuals with only neocortical amyloid and no cognitive symptoms, reflecting a high-risk phenotype for AD, have also shown indications of slowing of the pupillary contraction, indicating an early signal[66]. Further, significant correlations with neuropsychological measures such as the Mini Mental-State Examination (MMSE) were demonstrated in a 2015 study, possibly reflecting cortical neurodegeneration as a modulator of the pupil response[61]. However, diagnostic studies that applied a relevant comparator (diseases that could mimic AD) to test the true value in a diverse memory clinic population were missing from the literature. Also, no studies had examined the prognostic value of light reflex pupillometry in AD, which was proposed recently[27]. Together, this makes

the case for exploring qLRP in a real-world clinical setting to evaluate its potential as a digital bedside biomarker for AD.

1.4 Actigraphy

Actigraphy is the automated measurement of activity and circadian rhythm, typically with a device consisting of a chronometer and accelerometer[67]. This enables a continuous measurement of activity across days, as a measure of the circadian rhythm, and actigraphy is the preferred method for investigating circadian function in, e.g., sleep disorders[68]. Actigraphy can be performed using small body-worn devices, such as wristwatches or devices encapsulated in a medical adhesive patch[69]. A plethora of digital measures can be gauged from actigraphic recordings. These are, e.g., the overall activity level, circadian rhythm, sleep metrics[70], and more exotic circadian indices such as fractal regulation[71,72], which appreciates randomness in motor behavior. Together, this allows for a composite measure of activity and circadian rhythm. Actigraphy thus provides easily obtainable biomarkers and represents a scalable, accessible and cost-effective tool[73].

In Alzheimer's disease, early signs of circadian dysfunction and sleep disruption are apparent already at pre-clinical stages[71,74–81]. Distinct neurodegenerative processes in key brain areas, such as the suprachiasmatic nucleus of the hypothalamus, have been linked to the circadian rhythm disruption observed in AD[82]. In addition, midbrain structures such as the locus coeruleus, which are active in maintaining circadian rhythm and influence sleep, show signs of early damage in AD[83,84]. Further, neuropsychiatric symptoms such as apathy, which is highly prevalent in the disease, develop along the AD continuum and are reflected in lower activity levels[85–88].

Some of the earliest studies on actigraphy in Alzheimer's disease were carried out in later stages of the disease, where agitation and restless wandering behavior were recorded using low-frequency sampling of activity[89]. Newer studies have employed technology capable of tracking more minute changes in activity, providing a more exact picture of the motor activity and circadian dysfunction in AD[71]. Also, general activity levels and distinct motor behavior, such as walking, have been linked to key areas of AD pathology, e.g., the hippocampus[90].

Actigraphy, therefore, offers an inexpensive and easily obtained method for assessing changes related to pathology in AD, and the possibility of diagnostic usefulness seems reasonable. While several studies have used actigraphy to describe activity-related changes in the disease[88,91–96], few have investigated the application of actigraphy as a diagnostic marker in the disease[15]. As actigraphy data represent complex time-distributed measures[97], it is prudent to apply techniques for analyzing the data that can handle a large set of variables with inherent collinearity issues. To this end, machine learning techniques that filter relevant features can be applied to reduce the dataset to only include the most valuable predictors from a large initial pool[98]. While actigraphy presents itself as a promising approach for studying behavior and pathology related to AD, there are also a variety of possible confounders to consider when evaluating real-life behavior outside the clinic setting in an unsupervised manner. As many external factors can influence actigraphy, such as medication and comorbidity, it is important to control for confounding and to keep this in mind when interpreting results from actigraphic recordings[99]. Nonetheless, actigraphy remains an interesting and easily applied add-on to the current diagnostic practice, which currently falls short of providing accurate ecological assessment of everyday function and activity, especially in patients without caregivers.

1.5 Digital tools for AD diagnosis and prognosis

Provided the rationale given above for investigating qLRP and actigraphy in AD, we set forth to thoroughly investigate the diagnostic and prognostic usefulness of these markers in clinically relevant settings. Our aim was to assess the individual capability of each marker, as a stepping stone towards clinical implementation. While both markers were promising in the sense that they were deemed clinically applicable and had shown significant associations with AD pathology, we acknowledge some inherent pitfalls of the technologies that we applied. Light reflex pupillometry, while handy to apply, may be influenced by factors in the surroundings of a clinical setting. As such, ambient lighting could affect the results of a sensitive instrument[100], and while more sensitive compared to manual assessment of the light reflex[59], hand-held technology is not capable of the precision and high-frequency sampling rate of laboratory instruments[51]. Actigraphy could likewise be influenced by

external factors such as acute sickness, hospitalization or travelling during recording. These factors are difficult to control for as they would need to be evaluated on a case-by-case basis. Nonetheless, we thought it was necessary to investigate these technologies while keeping these limitations in mind. In addition, we always maintained the perspective that these markers were not meant to substitute the current traditional biomarkers. Rather we Envisioned them as a possible add-on to existing markers to evaluate the AD spectrum to a fuller extent, capturing dynamic features of a complex disorder.

2. Objectives

The overall aim of this PhD thesis was to assess the diagnostic ability of qLRP and actigraphy, as well as the test-retest reliability and prognostic utility of qLRP in early AD.

2.1 Study-specific aims

Study I: We aimed to investigate the test-retest reliability and intra-subject, short-term variability of qLRP in a mixed memory clinic.
Hypothesis: qLRP can be reliably measured and has acceptable short-term variability for clinical application.

Study II: We aimed to investigate the diagnostic accuracy of qLRP in Alzheimer's disease in a real-world diagnostic setting of a memory clinic with relevant diagnostic comparators.
Hypothesis: qLRP has diagnostic value and can differentiate between dementia disorders

Study III: We aimed to evaluate the diagnostic potential of actigraphy in Alzheimer's disease in a real-world memory clinic setting using machine learning with relevant diagnostic comparators.
Hypothesis: Actigraphy has diagnostic value and can differentiate between dementia disorders

Study IV: We aimed to assess the prognostic value of qLRP in early AD.
Hypothesis: Baseline and short-term dynamic changes (3-months) quantitative light reflex pupillometry have prognostic value in the early clinical stages of AD

3. Methodology

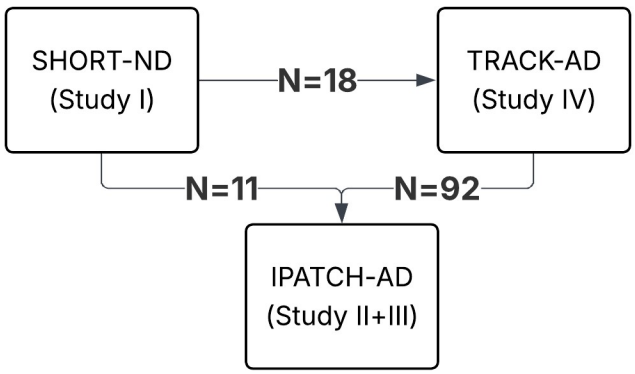
A condensed overview of the methods that were applied in Studies I-IV is provided in the following segment. Certain aspects are overarching across studies and are denoted accordingly.

3.1 Participants

All patients were recruited from a tertiary university memory clinic. The recruitment procedure differed slightly between studies. In study I, a consecutive cohort of patients undergoing lumbar puncture due to suspicion of neurodegenerative disease was included on the day of lumbar puncture, and importantly, before a final diagnosis was made at a consensus conference. The inclusion criteria in study I were: 1) undergoing lumbar puncture under suspicion of neurodegenerative disorder, 2) Mini-Mental State Examination (MMSE) score > 20, 3) ability to cooperate with qLRP procedure. The exclusion criteria for study I were: 1) severe bilateral ophthalmological disorders, 2) a history of stroke within 3 months of lumbar puncture, 3) current or previous major psychiatric conditions, 4) substance or alcohol abuse within the last two years and 5) participation in intervention trials. As such, Study I was agnostic of diagnosis. In Studies II and III, patients were recruited primarily at the time of diagnosis or during clinical follow-up. Patients from study I were subsequently included in study II if they fulfilled the inclusion criteria and none of the exclusion criteria for study II. The inclusion criteria for study II and III were: 1) a diagnosis of mild cognitive impairment (MCI) dementia due to AD[101,102], AD with mixed pathology (cerebrovascular co-pathology)[102], prodromal (MCI) DLB or DLB[103,104], or cognitive dysfunction due to cerebrovascular disease[105], 2) an MMSE score > 15, 3) ability to provide informed consent, 4) ability to cooperate with qLRP or actigraphy procedure, 5) age between 50-90 years. The exclusion criteria for study II and III were: 1) diagnosis of concurrent neurological or psychiatric disorders, 2) substance abuse or excessive alcohol intake (>5 units per day), 3) participation in concurrent intervention trials, 4) other known brain disorders causing cognitive dysfunction, 5) severe ophthalmological disorders (only applied in study II). In study IV, patients were recruited at the time of diagnosis. The

inclusion criteria for study IV were: 1) diagnosis of MCI or dementia due to AD[101,102], 2) an informant caregiver, 3) an MMSE > 19, 4) had a diagnostic hybrid brain [^{18}F]FDG-PET/magnetic resonance imaging (MRI) or [^{18}F]FDG-PET/computed tomography (CT), 5) ability to cooperate to the investigations. The exclusion criteria for study IV were: 1) concurrent alcohol or substance abuse, 2) major neurological or psychiatric disorder other than AD, 3) participation in drug trials. There was an overlap of participants between studies I and IV, such that patients included in study I subsequently diagnosed with AD and fulfilling all inclusion criteria and none of the exclusion criteria could continue in study IV. Patients in study I subsequently diagnosed with AD, DLB and CVD and patients in study IV could also be included in studies II and III following the same criteria. The overlap between study populations is described in each individual report and graphically presented in **Figure 2**. Healthy controls included in studies II and III were recruited from a list of participants in previous studies at the clinic that had consented to be contacted for further research proposals as well as through advertisement in the waiting area of the memory clinic. No compensation was given for participation. The healthy controls fulfilled the following inclusion criteria: 1) ability to cooperate to the investigations, 2) normal cognitive function, evaluated by the investigator, 3) age 50-90 years and did not fulfil any of the exclusion criteria listed for patients in study II and III.

Figure 2. Flow chart of patient overlap between studies. SHORT-ND, TRACK-AD and IPATCH-AD refer to the cohort studies from which patients were included. N refers to participants.



3.1.1 Diagnostic assessment

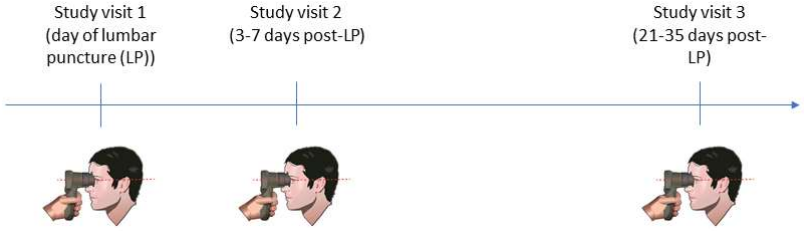
A comprehensive diagnostic evaluation was performed for all included patients, carried out at the same diagnostic center. This evaluation comprised at least a physical examination by a physician, structural brain imaging (CT or MRI), relevant lab work-up (e.g., thyroid hormone levels, B_{12}-vitamin) to rule out reversible conditions, and cognitive screening tests (Addenbrookes' Cognitive Examination[106] and MMSE[107]) and an examination by a nurse. In addition, our center uses advanced radiotracer imaging (e.g., [^{18}F]-fluorodeoxyglucose-positron emission tomography (FDG-PET), dopamine agonist tracers), neuropsychological evaluation, and cerebrospinal fluid sampling for measuring relevant biomarkers, where deemed necessary. Diagnoses were made at a multidisciplinary consensus conference (neurology, geriatrics, psychiatry, neuropsychology, nursing care) according to published NIA-AA criteria for MCI[101] or dementia due to AD[102] (all studies), prodromal (MCI)[104] or Lewy body dementia[103] (study I-III), and vascular cognitive dysfunction[105] (study I-III). In study I, representing a broader cohort, diagnoses of cerebral amyloid angiopathy and frontotemporal dementia were also given according to the most recent published criteria[108,109].

3.2 Procedure

3.2.1 Follow-up (FU)

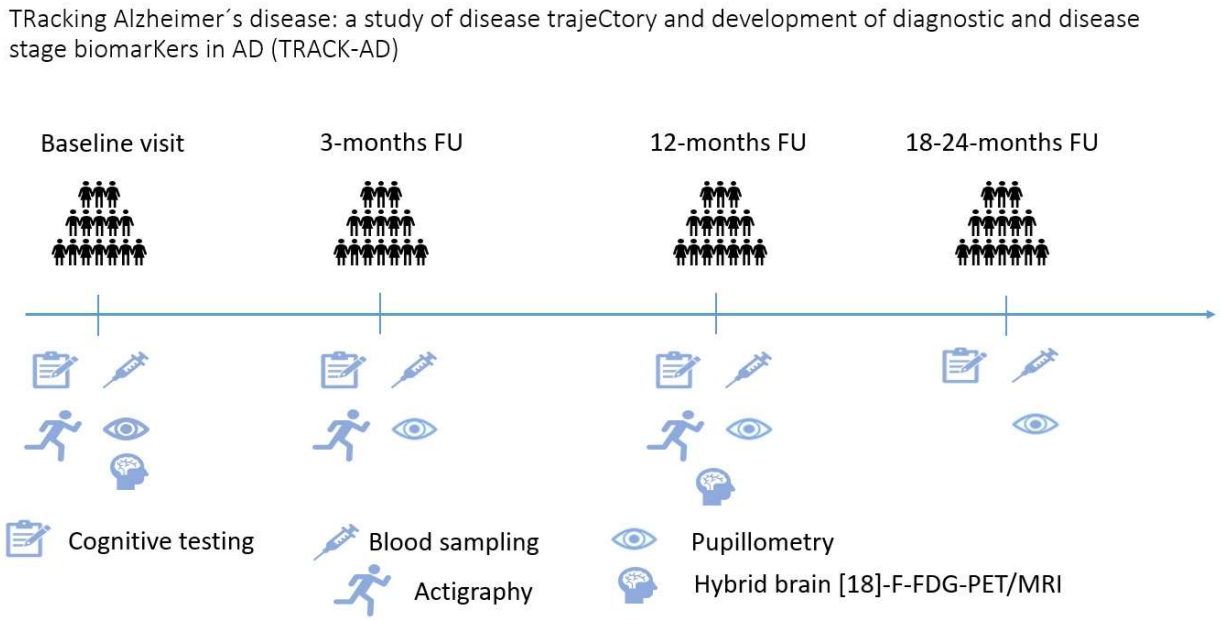
In study I, patients were followed for 21-35 days post lumbar puncture with serial assessment with qLRP (**Figure 3**).

Figure 3. Overview of follow-up procedure in study I.



In study IV, we followed patients for 18-24 months with three scheduled FU visits (see **Figure 4** for study overview).

Figure 4. A schematic overview of the follow-up procedure in study IV. Reprinted from study IV with permission.



3.2.2 Quantitative light reflex pupillometry (qLRP)

In studies I, II and IV, the same protocol for qLRP was used. We used a research-grade hand-held pupillometer (PLR-3000, NeurOptics®), which contains an infrared camera to record the pupil and light-emitting diodes that produce a broad-spectrum white light stimulus. The settings for the pupillometer were: no background lighting, stimulus intensity of 0.81 µW, stimulus onset of 1 s, stimulus length of 0.89 s, and a recording length of 10 s, in line with a recent study[110]. For the recording procedure, participants were seated and kept both eyes open and were told to fixate their gaze on a point behind the examiner with neutral eye position. An examiner recorded the pupillary response, keeping the eyelid of the participant manually open if needed without interfering with the line of sight of the pupillometer. The following metrics were provided by the pupillometer using a proprietary algorithm: resting pupillary diameter (mm), peak constriction diameter (mm), relative

change in pupil size following light stimulus (delta) (percentage change), latency to constriction (s), average and maximum constriction velocity (mm/s), average dilation velocity (mm/s) and time to 75 % re-dilation (T75) (s). A total of three acceptable measurements, without blinking or artifacts, from each eye was obtained. As a screening procedure, ophthalmoscopy was performed prior to qLRP to rule out major obscurities of the lens and vitreous body.

3.2.2.1 Data pre-processing

Following the recording of pupillometry, data were wirelessly transferred to a computer, where recordings flagged as corrupted by blink or eye movement-related artifacts were visually inspected using the *clintools* package in R. Strategic blinks in the contraction phase of the recording were identified and these recordings were subsequently omitted from further analysis. The first three acceptable measurements were used to calculate a mean value, which was the unit of analysis in all three studies (I, II, and IV).

3.2.3 Cerebrospinal fluid

For study I, we used the total tau concentrations as obtained from analysis of cerebrospinal fluid, sampled on the same day as baseline qLRP by lumbar puncture under sterile conditions. The cerebrospinal fluid was handled according to standard operating procedures and total tau was determined with an electrochemiluminescent sandwich enzyme-linked immunosorbent assay (ELISA) in our hospital lab as part of the routine operation. The laboratory equipment used for analysis changed from Innotest, Fujierebio, Ghent to Roche, Cobas 8000 during the study period and there was a change in the normal range cut-off levels. We therefore applied the normal range cut-off values in use at the time of analysis for each participant.

3.2.4 Neuroimaging

In study IV, we used diagnostic [^{18}F]FDG-PET scans obtained at the baseline visit and the 1-year follow-up to assess clinical neuroimaging progression. The scans were obtained on clinical PET/CT or PET/MRI systems in alignment with standard operating procedures[111]. Follow-up scans were always carried out on the same scanner system and an identical system was used in 73% of cases. From these scans, a visual progression rating was

derived in the following manner: two experienced nuclear medicine physicians reviewed each participant's baseline and follow-up statistical surface projections (Siemens syngo.via, MI neurology, Erlangen, Germany) in randomized order to minimize forecast bias. Each scan pair was evaluated as either progression or stable. These reviews were collated and a consensus decision was reached for each scan after the order of the scans was revealed.

3.2.5 Actigraphy

We used the SENS Motion Actigraph System®, a wearable actigraph, that was attached to the participants' sternum and thigh (one sensor in each anatomical location) for a total of seven days of recording. The sensor was embedded in a specially designed medical adhesive patch (3M®). The device records 3-axial acceleration and keeps time with a chronometer. We used the SENS Motion® algorithm, which is proprietary, to derive movement patterns and a 7-day activity profile was constructed with each 5.124 s recording epoch classified as one of eight different activity categories: upright standing, sitting, lying rest, lying movement, moderate intensity (brisk walking), sporadic walking, walking and running.

3.2.5.1 Data pre-processing

Actigrams constructed with the activity profile algorithm were visually inspected for irregularities (**Figure 8**), such as no variation in activity categories or missing data. We used 15-minute epochs for analysis.

3.2.5.2 Circadian indices

The following parameters were calculated for each activity category and for activity counts: L5: the least active five hours + timing, M10: the most active 10 hours + timing, relative amplitude, indicating the robustness of the circadian rhythm ((M10-L5)/(L10+M5)), and the intra-daily variability Indicating the fragmentation of the circadian rhythm.

3.3 Cognitive testing, clinical evaluation and demographic data

In study IV, we performed MMSE at all visits and we carried out a Clinical Dementia Rating (CDR)[112] at baseline and at the 1-year and 18/24-month FU-visit, which provides both a global score and a Sum-of-Boxes (SoB) score. We used the annualized changes in MMSE

and CDR SoB as assessed at 1-year FU as outcomes in study IV. We also evaluated patients according to clinical progression. This was evaluated based on the full spectrum of data, including cognitive tests and patient and informant history relating to the preceding period to each study follow-up visit. This clinically evaluated progression had been explored in a previous study in a mixed memory clinic cohort[113]. Demographic data were extracted from the medical records of the patients, and a history of eye disease was obtained from both the patients and caregivers, while healthy controls filled out a questionnaire describing these items.

3.4 Machine learning (ML)

We applied an ML algorithm to select the most relevant parameters for classifying patients and healthy controls in study III using actigraphy data. We used logistic regression and evaluated feature selection in the inner loop of a nested leave-one-out cross-validation set-up using a minimum redundancy maximum relevance (mRMR) algorithm[114]. We evaluated the performance of our machine learning classifiers by comparing them with the single most important feature (overall activity) according to relevant performance metrics: precision, accuracy, sensitivity, specificity, and the F_1-score, which is the harmonic mean of precision and sensitivity, commonly used to assess ML model performance[115].

3.5 Statistical analysis

We used R (ver. 4.4.2)[116] with the integrated development environment RStudio (ver. 2024.12.1) for statistical analyses, and Python and MATLAB (ver. R2019B, The MathWorks, Natick, Massachusetts, USA) for machine learning modelling.

We reported means and standard deviations or medians and ranges, where appropriate, according to the distribution of the variable.

In study I, we calculated the intraclass correlation coefficient (ICC) from a two-way mixed effects model to evaluate the test-retest reliability of qLRP. Following the McGraw & Wong convention, we used the ICC (A,3)[117] and evaluated the reliability according to intervals provided by Koo & Li[118]. We fitted linear mixed models (LMM) to model variance, assuming an unstructured covariance pattern, with the *lme4* package in R. We entered each qLRP

metric as the outcome with participants as random effects and visits as fixed effects. This allowed us to extract the standard deviation of the residual, indicating the random fluctuation for each participant. In study II, we fitted logistic regression models to evaluate the ability of qLRP to classify participants in the following scenarios: AD vs. healthy, AD vs. CVD (separately for vascular cognitive dysfunction and combined with mixed AD), AD vs. DLB. We had three models for each comparison: all qLRP metrics (pupillometry), only qLRP metrics that were not collinear with adjustment for age and sex (sparse model), and a cubic spline model with 6 degrees of freedom for each qLRP parameter (spline model) using the *splines* package in R. The models were evaluated according to the area under the receiver operating characteristics curve (AUROC) with an associated 95 % confidence interval (CI) computed using the DeLong method. A 1000 times 10-fold cross-validated (cv)AUROC was computed for the sparse model in each discriminatory scenario to evaluate optimism. In study IV, we used logistic regression to fit models with clinical progression at the 1-year visit as the outcome and univariable models with each qLRP metric as measured at baseline included as a predictor. These models were adjusted for age, sex and resting pupillary diameter (except for the model with this variable). We evaluated the odds ratio with an associated 95 % CI. Similar models were evaluated with the short-term (3-month) dynamic changes in qLRP. We used qLRP recordings from the baseline visit and 3-months follow-up for this analysis by calculating the absolute change: $3\ month\ measurement - \ baseline\ measurement$.

These models were further adjusted for the baseline value, thus reflecting estimates for relative changes. Similar models were fitted with [^{18}F]FDG-PET visual progression as the outcome. These logistic regression models were evaluated according to a cvAUROC using 5-fold 10 times cross-validation. We also reported estimates of sensitivity and specificity with associated 95 % CIs obtained with bootstrapping as well as the negative and positive predictive value for the best predictors. We likewise fitted linear models with the MMSE and CDR SoB annualized changes (log-transformed) as separate outcomes with the same predictors.

P-values were considered significant if below the 0.05 level (α). We report p-values for two-tailed tests throughout. No adjustments for multiple comparisons were made, and our results were therefore hypothesis-generating.

3.5.1 Sensitivity and confounder analyses

In studies I and II, we performed sensitivity analyses by stratifying our cohort. In study I, this procedure was done for age, sex, AD diagnosis and CSF-total tau levels. In study II, this was done for disease stage (MCI versus dementia), mild eye disease, and pupil medication. We evaluated the 95 % confidence intervals for the ICCs and the AUROC values for the logistic regression models to check if the stratified analyses produced dissimilar estimates, indicating confounding. In study III, we investigated confounding by examining models with outcome probability as the dependent variable for the machine learning models. Here, we examined the influence of age, sex and medication that could affect the actigraphic variables, such as use of sedative or hypnotic medication. In study IV, we likewise investigated confounding by exploring models that incorporated mild eye disorders and medication that could affect the pupil in separate multivariable adjusted models.

3.6 Ethical considerations and approval

The procedures used in the studies were all deemed safe for the participants. Ethical approval was obtained for all studies (journal numbers: H-21040317, H-21044863, H-21045068) before the commencement of studies and all studies were registered prior to recruitment at clinicaltrials.gov (Clinical Trial Nos.: NCT05175664; NCT05175664; NCT05175664). A written, informed consent was obtained from all participants and we followed the tenets of the 1975 Helsinki Declaration.

4. Summary of results

In the following section, a summary of results from studies I-IV is given.

Study I

In this study, 40 participants provided at least two serial measurements. Participants were followed for a total of 35 days with up to three pupillometry sessions. From these pupillometry recordings we calculated the intraclass correlation coefficients from a two-way mixed effects model to assess the test-retest reliability of hand-held qLRP. **Figure 5** shows a spaghetti plot of individual measurements for the average pupillary constriction velocity. Most measurements were similar between sessions but a few outliers were also identified. When looking broadly at the pupillometry metrics, a general tendency for high test-retest reliability was shown. As a further characterization of the properties of pupillometry in a mixed memory clinic cohort, we also computed the intra-subject visit-to-visit variability from linear mixed models. This provided reference cut-offs for assessing longitudinal changes in pupillometry as listed in **Table 1**.

Figure 5. A spaghetti plot of the constriction velocity with each line representing a single subject (reprinted and modified with permission from Gramkow et al.[1])

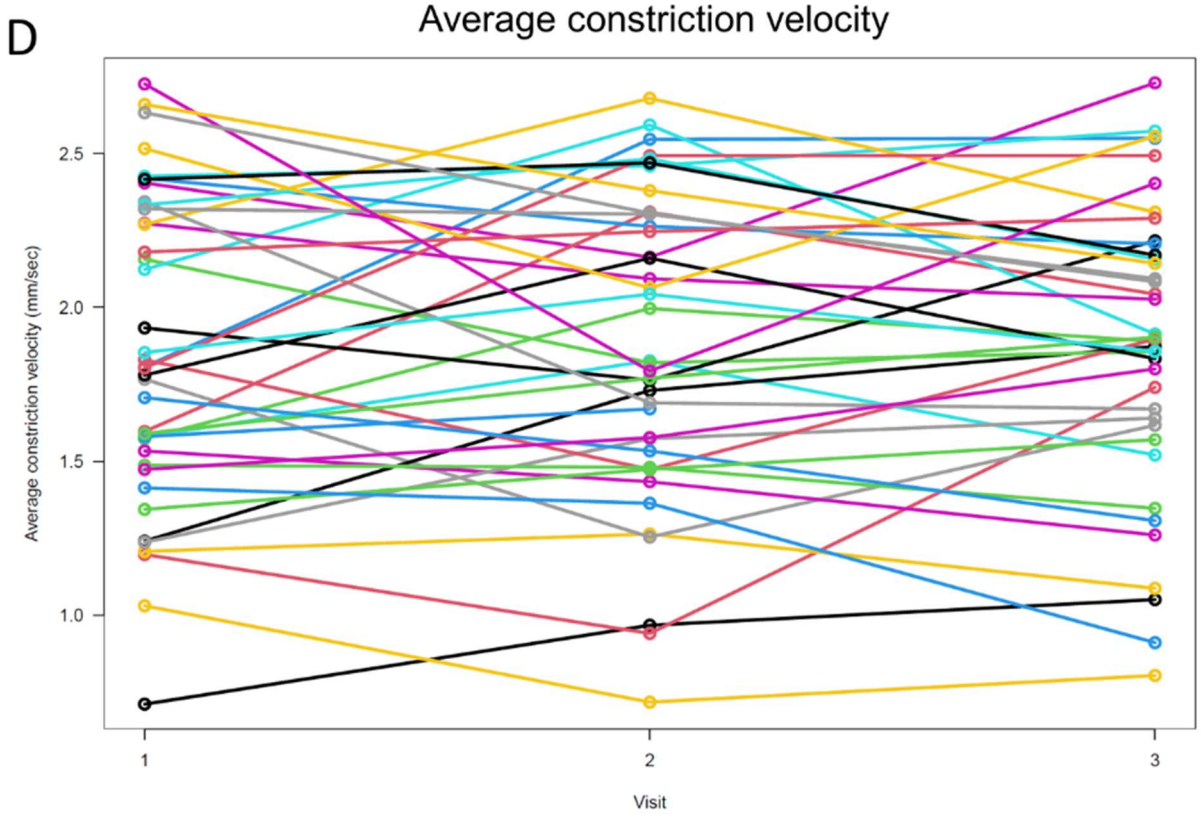


Table 1. Intraclass correlation coefficients with corresponding 95 % CIs indicating test-retest reliability and clinically significant changes derived from the standard deviation of the residuals from LMMs (reproduced with permission from Gramkow et al.[1]).

qLRP parameter	Intraclass correlation coefficient [95 % CI]	Clinically significant change
Resting pupil diameter	0.889 [0.81 - 0.938][a]	0.25 mm
Peak constriction pupil diameter	0.928 [0.87 - 0.961][a]	0.11 mm
Delta change	0.933 [0.885 - 0.962][a]	2.9 %
Latency	0.86 [0.762 - 0.922][a]	0.015 s
Average constriction velocity	0.901 [0.831 - 0.945][a]	0.24 mm/s
Maximum constriction velocity	0.896 [0.823 - 0.942][a]	0.38 mm/s
Average dilation velocity	0.892 [0.816 - 0.94][b]	0.10 mm/s
T75	0.557 [0.235 - 0.758][c]	0.5 s

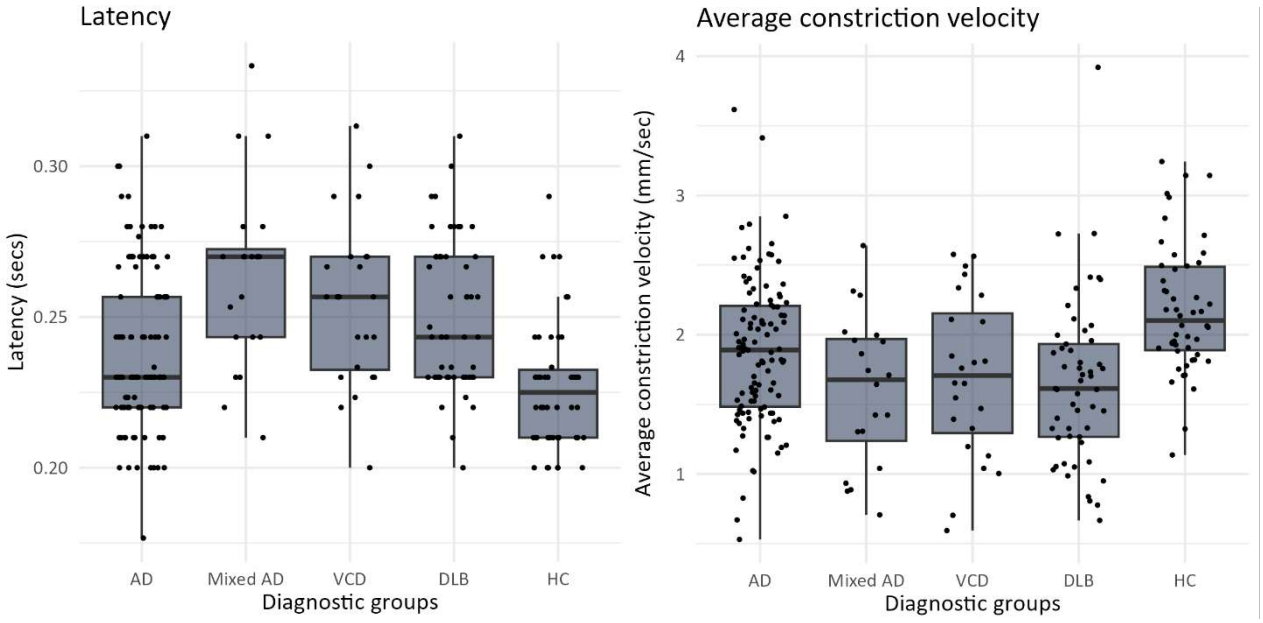
[a]N = 39 patients [b]N = 38 patients, [c]N = 37 patients. T75 = Time to reach 75 % of baseline (s).

We further assessed whether certain demographic variables (i.e., age and sex), a clinical diagnosis of AD, or the presence of neurodegeneration as measured using CSF total-tau impacted the reliability of measurements. We found no evidence of either factor influencing the test-retest reliability of qLRP. There were no adverse events associated with the pupillometry procedure. These results provided justification for the further exploration of qLRP in AD.

Study II

In study II, we explored the diagnostic ability of qLRP in a memory clinic setting, including comparators that are often evaluated alongside patients with AD, namely patients with DLB and CVD. A total of 204 patients and 50 healthy controls were included. We constructed logistic regression classifying models and built a web-based application from these models. We found that all qLRP parameters differed between groups, with most convincing results shown for latency and constriction velocity (**Figure 6**).

Figure 6. Select qLRP parameters (latency and average constriction velocity) for the included diagnostic groups. The horizontal stroke in box indicates the median, and boxes indicate the 25 % and 75 % interquartile range (IQR). Whiskers are 1.5 x IQR. AD = Alzheimer's disease, Mixed AD = Alzheimer's disease with mixed pathology (AD+vascular pathology), VCD = vascular cognitive dysfunction, DLB = dementia with Lewy bodies, HC = healthy controls. Modified with permission from Gramkow et al. [2]



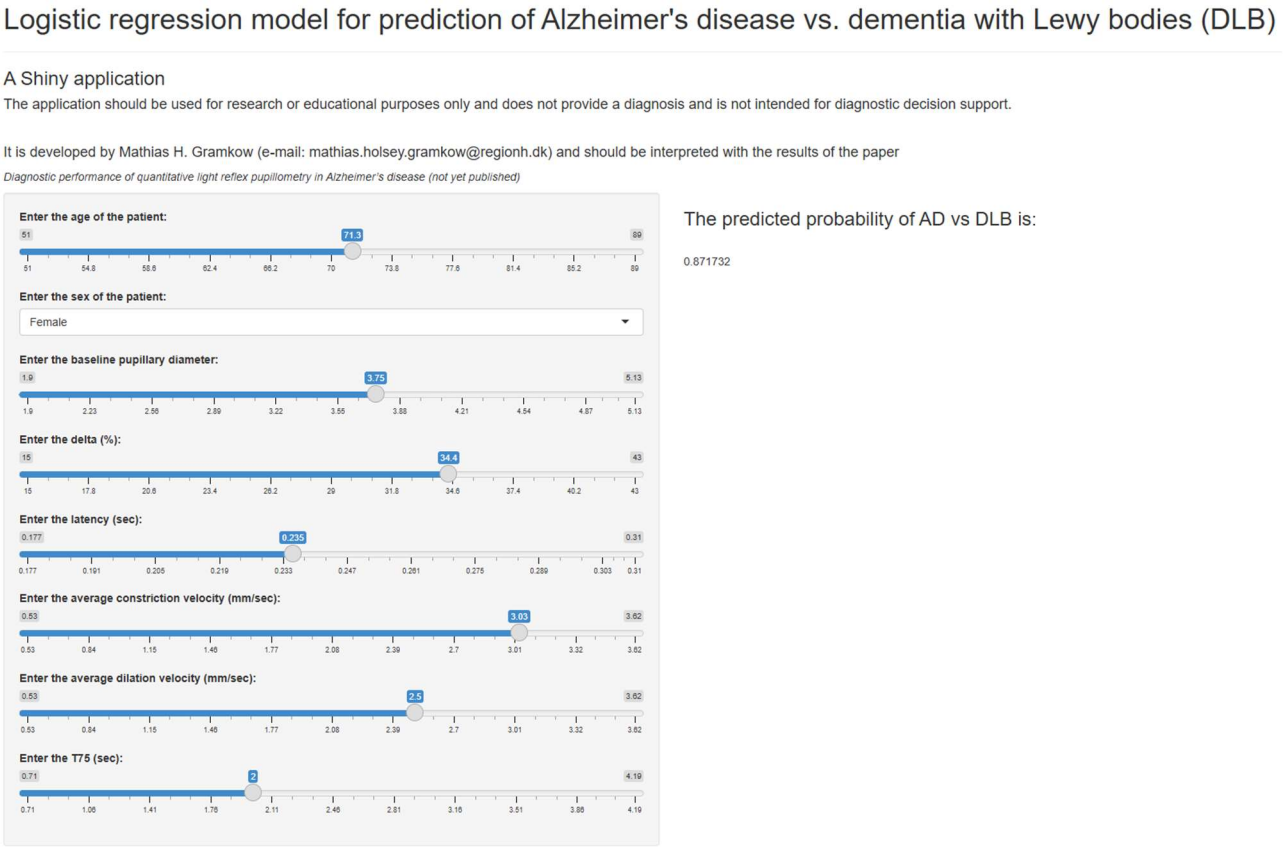
As shown in Table 3, we demonstrated that qLRP combined with age and sex provided above chance AUROC values and that classification accuracy was slightly better for discriminating between AD and cerebrovascular disease than versus healthy and DLB.
In our sensitivity analyses, we explored the influence of age and sex as well as the possible confounders, mild eye disorder and medication that could affect the pupil. We found similar estimates when stratifying for these important possible confounders. The same was true for the classification of individuals with MCI.

Table 3. Diagnostic performance of the classification models as evaluated with both the in-sample and cross-validated area under the receiver operating characteristics curve (AUROC) and a two cut-point procedure for sensitivity and specificity. Modified with permission from [2]. Model 2 refers to a sparse model with age and sex included as predictors.

	AUROC [95 % CI]	Specificity at minimum 85 % sensitivity [95 % CI] (cut-point)	Sensitivity at minimum 85 % specificity [95 % CI] (cut-point)	Cross-validated AUROC
AD vs. HC Model 2	0.74 [0.66-0.83]	40 % [19-68 %] (0.52)	38 % [16-72 %] (0.81)	0.68
AD vs. DLB Model 2	0.75 [0.67-0.83]	47 % [21-67 %] (0.52)	46 % [18-71 %] (0.81)	0.69
AD vs. Mixed AD + VCD Model 2	0.8 [0.72-0.88]	57 % [31-79 %] (0.59)	52 % [36-78 %] (0.82)	0.73
AD vs. VCD Model 2	0.81 [0.72-0.9]	63 % [25-89 %] (0.72)	58 % [33-85 %] (0.86)	0.7

In **Figure 7**, the web application that was developed to explore regression models is shown. Entering the qLRP values and age and sex of an individual a probability can be calculated and evaluated according to the cut offs given in Table 3.

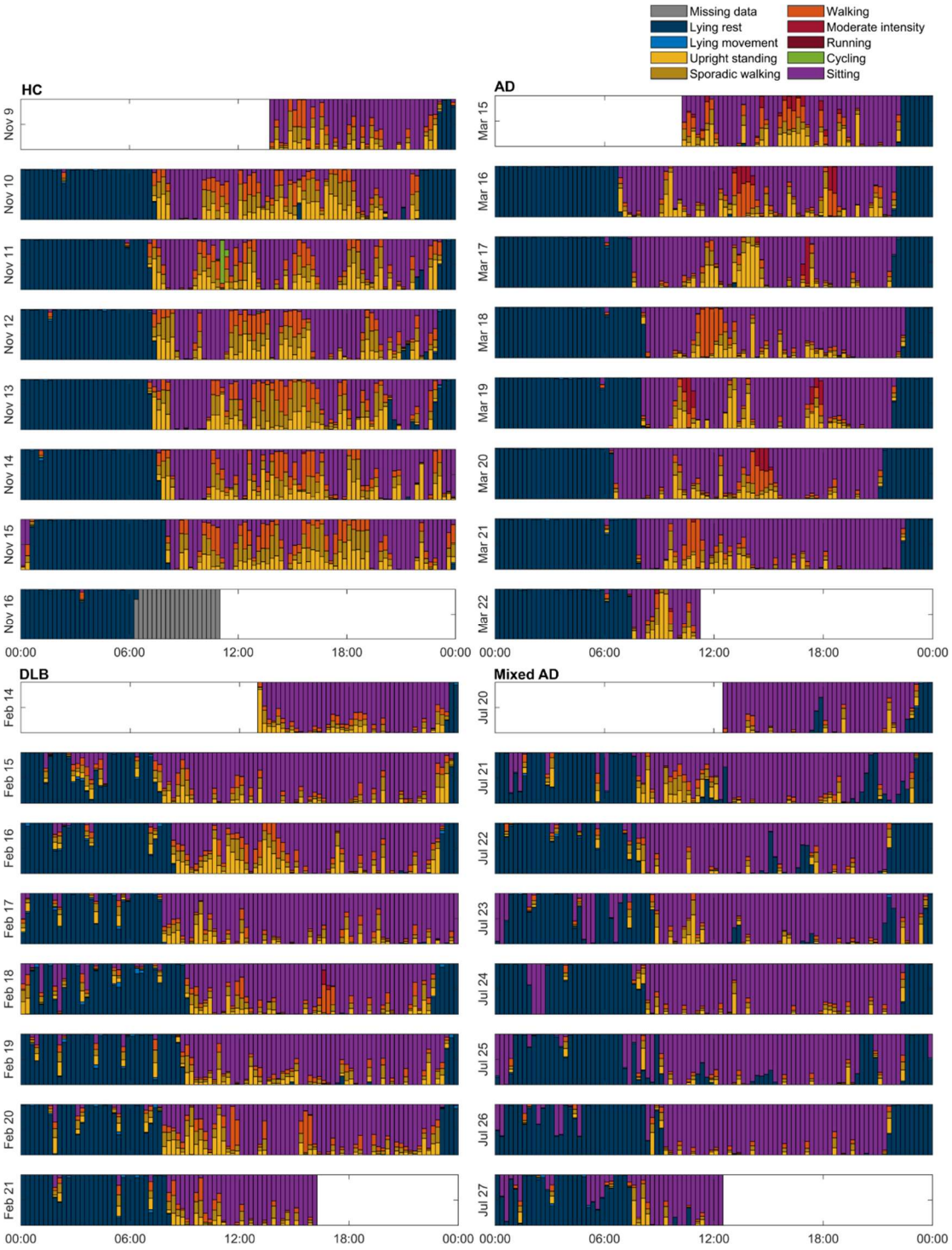
Figure 7. The Shiny web application developed for exploring regression models that were applied in study II. The web application was created by the present author, is referenced in Study II[2] and can be found online at: https://matgram.shinyapps.io/dqLRP-ADDLBDiagnosis/ (accessed 16/5/2025).



Study III

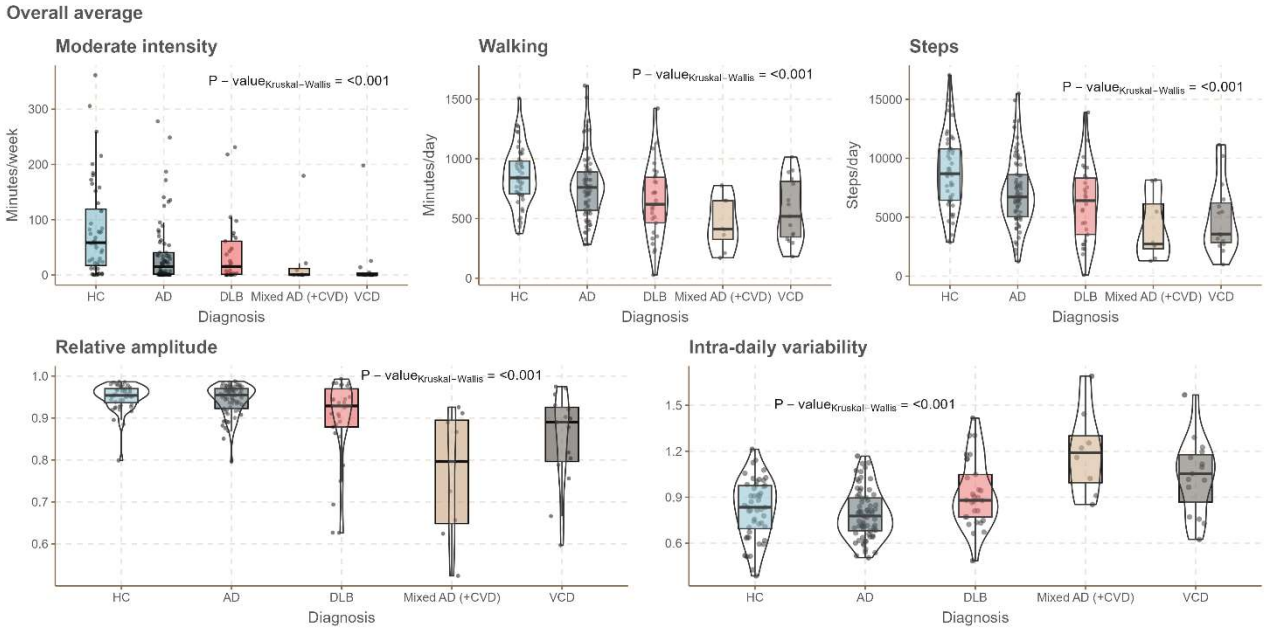
In study III, we used actigraphy to ascertain its diagnostic usefulness when coupled with a machine learning classifier. A total of 122 patients and 48 healthy controls provided complete actigraphy data for analysis. Given the activity profiles created from seven-day recordings (**Figure 8**), we calculated features related to the circadian distribution of the activity categories.

Figure 8. Actigram from one healthy control (HC), and one patient with Alzheimer's disease (AD), dementia with Lewy bodies (DLB), and mixed AD (+ cerebrovascular pathology). The figure shows the distribution of activity patterns during a 7-day recording period. Reprinted with permission from Gramkow et al.[3].



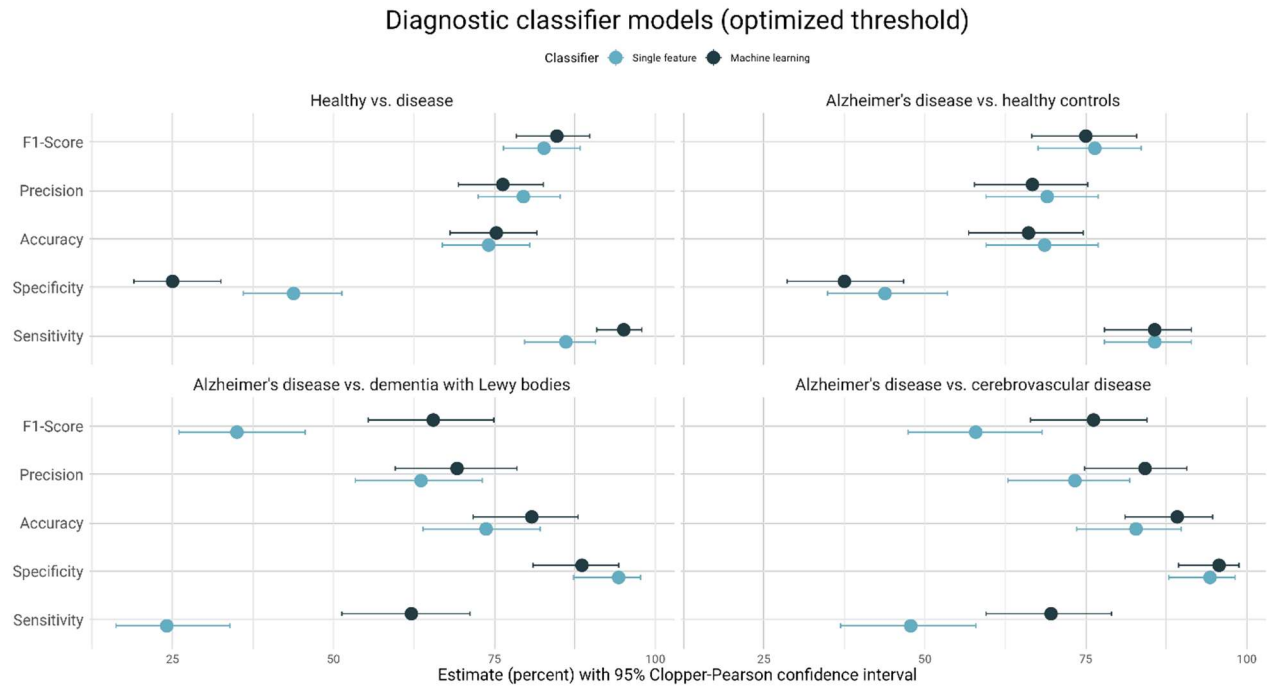
Some of the most discriminative variables are shown in **Figure 9**, namely moderate intensity activity, walking and number of steps. Healthy controls tended to spend more time doing moderate intensity activity, walk more and take more steps. The circadian indices relative amplitude and intra-daily variability were largely similar between the AD group and healthy controls, whereas patients with CVD and DLB exhibited a larger degree of circadian dysfunction, reflected by higher fragmentation (intra-daily variability) and decreased robustness (relative amplitude) of the circadian rhythm.

Figure 9. Discriminatory activity categories and circadian indices for each diagnostic group. Reprinted with permission from Gramkow et al.[3].



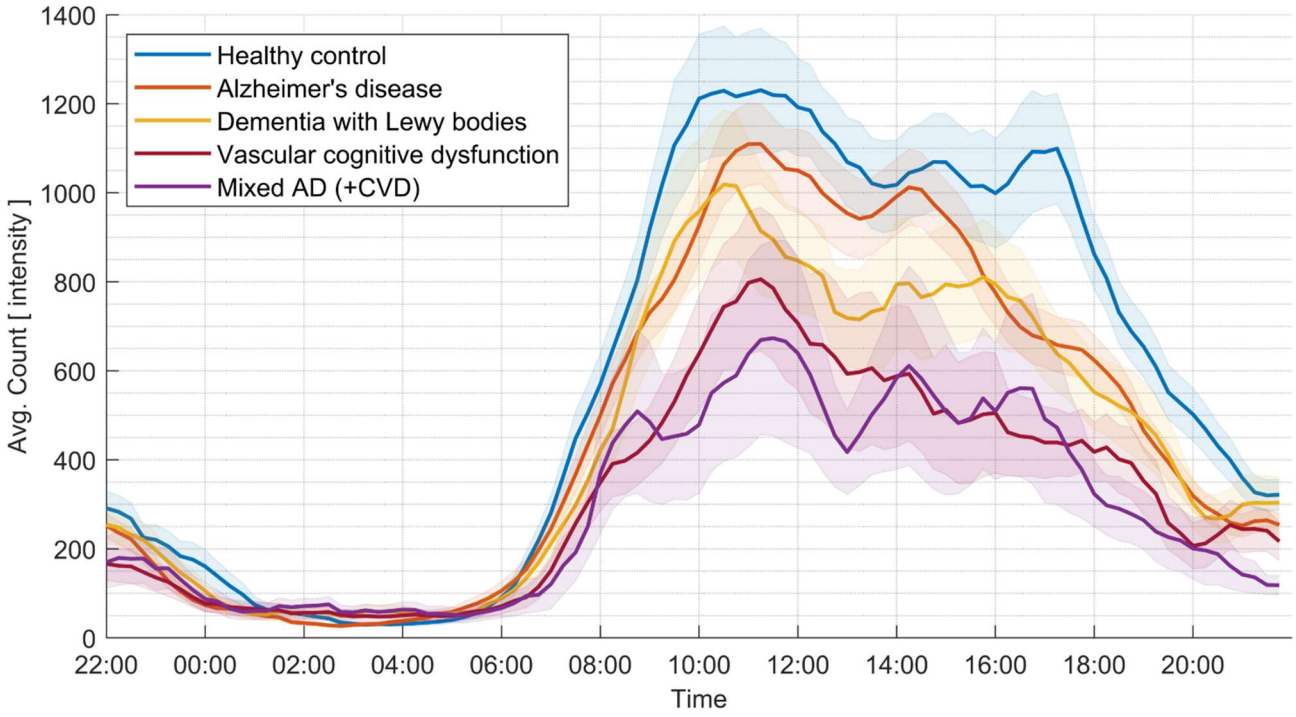
Using the full feature set and applying a machine learning algorithm, we achieved a relatively high accuracy for differentiating dementia disorders (>80%), while the separation of healthy and disease groups was less obvious, with associated accuracies of ~65-75% (**Figure 10**).

Figure 10. Performance estimates with 95 % CIs for machine learning classifier models and the single feature classifier (overall activity). Reprinted with permission from Gramkow et al.[3].



The distribution of overall activity for each diagnostic group over the course of an average day is displayed in **Figure 11**. There was a separation of the groups during daytime activity, where healthy controls maintained a higher level of activity throughout the afternoon. Patients with Alzheimer's disease followed the healthy controls, and patients with CVD had the lowest activity levels. The circadian phase was intact for all groups.

Figure 11. The daily activity levels for each diagnostic group in study III. The curve smoothed to the nearest 15-minute and represent the overall average activity. Reprinted with permission from Gramkow et al.[3].



We investigated possible confounding by examining the output probability for the ML models for the healthy controls and patients. These showed no significant confounders for the healthy group, while age and use of sedatives were associated with a higher output probability in the disease groups, indicating possible confounding of these variables. It cannot be excluded, however, that disease stage was linked to age and therefore could explain this association. Study III demonstrated the diagnostic potential of actigraphy when applied in a real-world clinical setting and highlighted possible areas of caution in interpreting output from ML models.

Study IV

In study IV, we explored qLRP as a prognostic marker in early AD in a prospective, longitudinal, observational cohort study. A total of 86 participants completed the 1-year follow-up or experienced progression during this period and were included in the present study (**Table 4**). Progressors were significantly older than stable patients (p=0.021).

Table 4. Cohort characteristics at the baseline visit with stratification according to the progression status at 1-year follow-up. Reprinted with permission from study IV.

	Stable (N=39)	Progression (N=47)	Total (N=86)	p-value
Age (years)				**0.021**[1]
Mean (SD)	73.8 (7.4)	77.3 (6.4)	75.7 (7)	
Range	53.3 - 89.9	63.7 - 88.2	53.3 - 89.9	
Sex				0.709[2]
Male	19 (48.7%)	21 (44.7%)	40 (46.5%)	
Female	20 (51.3%)	26 (55.3%)	46 (53.5%)	
Educational level (years)				0.315[3]
Mean (SD)	11.2 (3.4)	12.2 (4)	11.7 (3.7)	
Range	7 - 20	7 - 20	7 - 20	
Disease severity				0.542[2]
MCI	10 (25.6%)	12 (25.5%)	22 (25.6%)	
Mild	28 (71.8%)	35 (74.5%)	63 (73.3%)	
Moderate	1 (2.6%)	0 (0.0%)	1 (1.2%)	
MMSE total score				0.758[1]
Mean (SD)	25.4 (2.6)	25.6 (2.4)	25.5 (2.5)	
Range	20 - 30	21 - 30	20 - 30	

This study showed that short-term dynamic changes in the resting pupillary diameter predicted clinically evaluated progression (**Figures 12 and 13**). This could not be replicated when investigating associations with MMSE or CDR-SoB score. When we adjusted for the presence of mild eye disease, this association with clinical progression was no longer significant. When assessing the cvAUROC this showed a high standard deviation, indicating that the prediction model underperformed in unseen data, and thus was not robust. When predicting neuroimaging visual progression, we found that relative pupillary change (delta) was a predictive marker with a cvAUROC of 0.64, although the

estimate for the odds ratio in the logistic regression model was not significant, indicating low robustness of the finding.

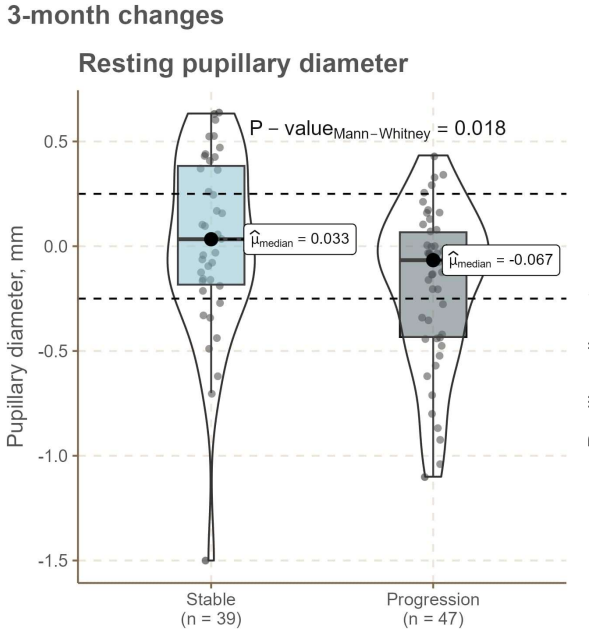
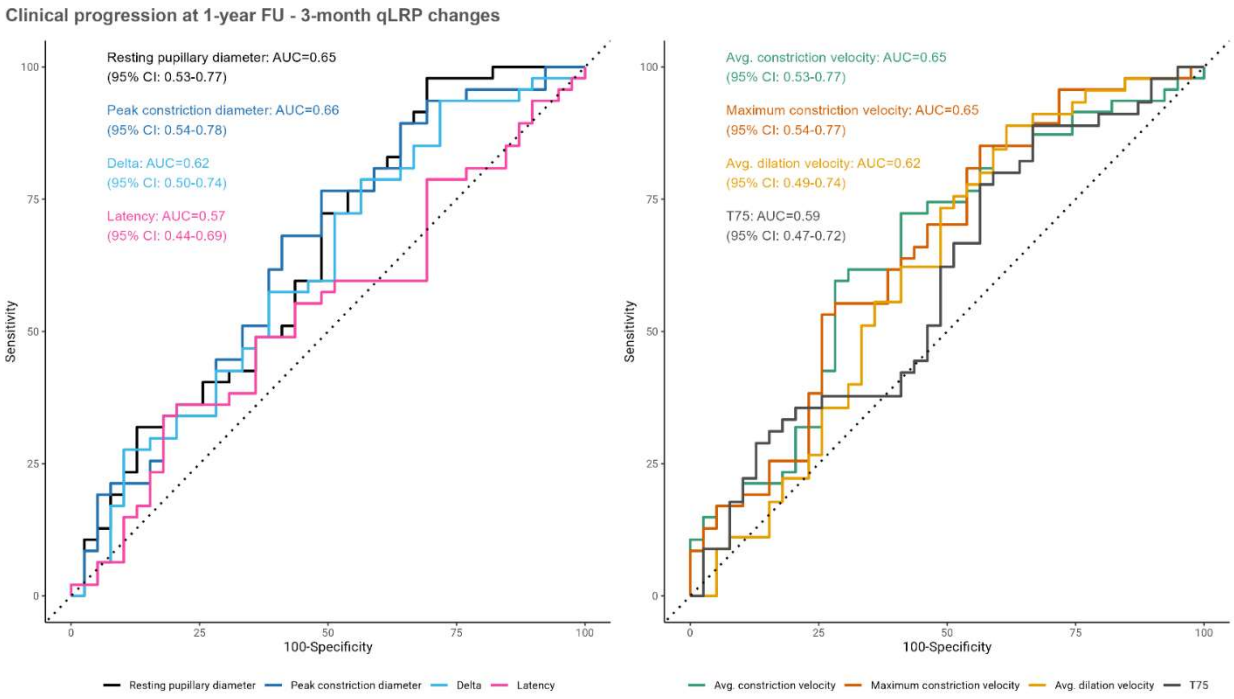


Figure 12. Box-plots with overlayed violin plots showing the group that progressed within one year and the stable group according to pupillometry findings in study IV. The dashed line indicates the intra-individual variability that was determined in study I[1]. Reprinted and modified from study IV.

When MMSE change was used as the outcome, two other qLRP metrics, the peak constriction diameter and the relative pupillary change, were associated with progression.

Figure 13. Receiver operating characteristics for each qLRP metric showing the predictive performance for clinically evaluated progression at the 1-year follow-up. A 95% CI (DeLong) is shown in parentheses. Reprinted with permission from study IV.



From study IV, we could conclude that there are indications that diminishing pupillary response could serve as a biomarker of progression in early AD, although the association was not robust to confounding by medication that could affect the pupil and could not be replicated across progression outcome modalities.

5. Discussion

In the following section, a general discussion of the findings presented in the appended studies I-IV is provided. The discussion is divided into relevant subsections that may span across one or more studies and are denoted accordingly.

qLRP as a diagnostic marker of AD (study I+II)

To establish whether hand-held qLRP could prove valid for clinical application in a memory clinic, we first explored the test-retest reliability of qLRP in this real-world scenario. Using this approach, we provided estimates of the reliability of the marker in the intended use case, namely patients undergoing investigation due to suspicion of a neurodegenerative disorder. The study was the first of its kind, with previous research efforts focused on either healthy individuals[119] or more recently in the acute, non-neurological, setting[59], where the tool has shown excellent reliability. We showed that for most qLRP metrics they were reliable in a test-retest scenario, aligning with results of previous studies, although few metrics, such as the T75 were more prone to be influenced by external factors. We showed that ambient light levels, even when regulated within recommendations for normal working conditions[120], could significantly influence T75 reflecting sub-optimal reliability for clinical application of this particular metric. We further provided estimates of intra-subject short-term variability, which could prove useful for determining changes not due to random fluctuations in longitudinal setting. Determining test-retest validity of digital biomarkers is part of the V3 framework[121] that provides guidance for the development and implementation of DHT, and our study aligns with the sentiments of this framework, namely that digital biomarkers should prove reliable before clinical validation is carried out. Relying on the results of study I we went on to assess the diagnostic performance of qLRP in a memory clinic setting. We showed a tendency towards a slowing of the pupillary constriction as well as a diminished pupil size and increased latency in alignment with previous studies[55–57,63]. One previous study had evaluated individual qLRP metrics as diagnostic markers in AD, demonstrating a near-perfect classification accuracy for some qLRP metrics[57]. We sought to evaluate the diagnostic performance by constructing diagnostic classification models that relied on the full signal of the light reflex. We

evaluated the diagnostic performance both when classifying AD versus healthy as well as differential diagnostic classification accuracy for commonly evaluated disorders such as CVD and DLB. Our results showed a moderate accuracy in differentiating healthy controls and patients with AD, which was in alignment with findings by Frost et al.[63], who showed a modest accuracy in classifying individuals with amyloid pathology, when combining qLRP metrics. Our results may have been driven in part by the more pronounced pupillary changes observed in both the DLB and CVD groups. There are some possible explanations for why this was observed, namely the findings of cholinergic brainstem dysfunction in DLB[122,123] and distinct white matter lesions in CVD[124] that may have impacted the pupil response. Our study was not designed to address these aspects, and future studies should evaluate the pathophysiological underpinnings of these observations. Nonetheless, our results provided proof-of-concept for the use of qLRP in memory clinic setting, with the most promising results for differential diagnostic application of AD. However, external validation of our findings awaits before clinical implementation can be envisioned. In addition, the add-on value of qLRP to traditional markers was not tested in study III and although qLRP represents a low-cost alternative that is not currently available, this aspect should also be the focus of future studies.

qLRP as a prognostic marker for AD (study I+IV)

In study IV, we evaluated the prognostic potential of qLRP in a cohort of patients with early AD. We conducted a thorough exploratory investigation of hand-held qLRP as a predictor of 1-year progression on several indices, such as clinically evaluated progression, progression on the MMSE and CDR-SoB and visual neuroimaging progression. We showed here for the first time that short-term dynamic changes in pupil size, possibly reflecting a decreased arousal level, were indicative of clinically evaluated progression. We further demonstrated that the relative pupillary constriction was associated with progression on the MMSE and could predict visually detectable changes on [^{18}F]FDG-PET. Using the short-term intra-individual clinically significant changes derived in study I, we could evaluate the short-term dynamic changes according to actual changes and random fluctuations. This provided a framework for assessing changes that may be due to neurodegenerative processes.

As this was the first study to evaluate qLRP as a prognostic marker, there was scarce literature to compare findings with, and longitudinal application was only recently suggested[27]. Some indirect evidence of brainstem involvement preceding higher cortical deposition in AD has emerged[42,43,47,48], hinting at an explanation for our observations. These studies mainly highlight a pre-clinical build-up of tau in the brainstem, which seems to plateau after symptoms emerge. Our study was aimed at probing functional aspects of the light reflex, which earlier autopsy studies did not investigate. As for the other observed associations with qLRP metrics, their significance disappeared, when we adjusted for resting pupil size, age and sex. This suggested that prognostic properties of qLRP may solely rely on the resting pupil size exerting a higher order influence on the other qLRP metrics. This was repeatedly shown in the three studies examining the light reflex in this thesis, namely that the resting pupil size was correlated with all other qLRP metrics and could confound associations. This has been shown earlier in both healthy and patients with AD[57] and it highlights the appropriateness of assessing the light reflex as a complex whole when examining its ability as a biomarker. This further supports the method applied in study II, where models investigating all qLRP metrics were evaluated as diagnostic classifiers. Our modest sample size in study IV, however, limited the number of variables we could include in the model without risking overfitting. Additionally, our findings suggest that the light-stimulated pathway was not indicated in clinically evaluated disease progression and the pupil state, governed by higher cortical functions, is more useful for predicting progression.

It should be noted that we observed small changes between stable and progressors, and the prediction model using only short-term dynamic changes in the resting pupil size was not robust in a cross-validation setup. The estimate was also not significant when we adjusted for medication that could affect the pupil as a possible confounder. While our findings did not prove to be robust, the hypothesis could be explored further in a design with a cognitive stressor measured through pupillary responses, such as a mental arithmetic task, which has been shown earlier to associate with cognitive impairment in AD[125]. While we showed that changes in the pupil size were informative for clinically evaluated prognosis, this measure lacked clinical translation to other assessment

modalities such as the MMSE and CDR, restricting its usefulness. We likewise found significant associations between qLRP metrics and MMSE change, which could not be affirmed on the other two progression modalities. These findings highlight the inherent problems of assessing progression in studies on AD, as progression scales capture different aspects of the disease processes[126].

Our choice of clinically evaluated progression as an outcome was due to its ability to easily translate to clinically meaningful change, at least from a clinician's viewpoint. However, this evaluation of progression may rely more heavily on the individual clinician's experience and therefore may not translate well to other settings. Using a visually appreciated neuroimaging progression might mitigate some of the clinically associated confounding, although this measure was not without faults, as it likewise relies on the neuroimaging physician's expertise and may not capture the subtle strategic changes that influence ADL function[127]. Future studies should evaluate qLRP with a quantitative neuroimaging outcome to see whether brainstem changes, as evidenced by pupillometry, can predict AD-related neurodegeneration in key brain hubs.

Actigraphy as a diagnostic marker for AD (study III)

We investigated the diagnostic properties of actigraphy when applied in a real-world memory clinic setting using relevant diagnostic comparators. We found that we could reliably distinguish between dementia disorders, while we achieved a moderate accuracy when differentiating healthy and disease groups. This discrepancy in findings probably reflected the heterogeneity of our disease groups, as the ML models identified different disease-specific features for the dementia group classifications.

Our approach was novel in the sense that we applied actigraphy and ML in a real-world clinical scenario, focusing on the performance in a memory clinic setting, where differentiation of dementia disorders remains a challenge. In our study, the focus was mainly on the aspects of activity and movement patterns and whether these indices could be used for diagnostic purposes, which had not previously been explored. Previous studies have compared healthy groups and patients with AD, and report similar results, with a more sedentary profile for patients with AD[92,95]. A study by Varma & Watts investigated

daily activity patterns and found that patients with AD spent less time doing moderate intensity activity, in line with our results[91].

We did not investigate sleep metrics, as these were not a part of the algorithm that was used to extract activity categories for the construction of activity profiles. Several studies have reported sleep disturbances in AD, possibly reflecting integral neurodegenerative processes[79,80,128,129], and this remains to be explored as a diagnostic marker in AD. While we provide some indication of nighttime rest through our proxy measures, this was probably not sufficient to properly characterize the specific sleep-related disturbances known to be prevalent in AD, such as sleep fragmentation[74,78,130,131], and overall nighttime activity did not differ significantly between groups for most variables.

As evidenced, actigraphy represents an enormous source of data, as various proxy measures can be derived from activity counts using a wide range of algorithms[132]. We are planning to further explore our raw actigraphy data using sleep algorithms for this purpose and it remains to be shown what distinct sleep metrics can provide in this regard. Also, we did not study the add-on value of actigraphy to traditional biomarkers. This was a deliberate choice as we were mainly interested in characterizing the diagnostic properties of actigraphy alone before evaluating the marker in a comparative setting. This choice limited us in drawing firm conclusions on the possible gain by adding actigraphy to a diagnostic workflow. However, the metrics provided by actigraphy are not captured using traditional biomarkers such as cerebrospinal fluid biomarkers, which measure relatively static properties of neurodegenerative processes[133], while remaining uninformative of the phenotypical behaviour linked with the disease in question. To this end, we find that actigraphy provides additional information to the existing battery of examinations, and our results suggest that it may be a disease-specific biomarker that could, for example, help differentiate between disorders such as AD and DLB that share a substantial amount of co-pathology[134].

Machine learning (study III)

In study III, we trained an ML model to filter and select relevant features to classify participants into diagnostic groups. This approach was chosen to accommodate for the inherent complexity of time-distributed actigraphy data[97]. However, interpretability issues

often accompany the application of ML[135]. This could be especially relevant in a diagnostic setting, where results of a diagnostic test need to be communicated to the patient as well as understood by the physician. There is, however, a clear trade-off between explainability and predictive power for machine learning models[135], which suggests that in some cases predictive power should be prioritized. Behind this viewpoint lies the assumption that the model behind the ML algorithm is both rigorously tested and that the appropriate model is applied for the intended use case, which indeed requires human oversight. On the other hand, wanting to understand how a predictive model for medical applications works seems justified given the implications of a positive or negative result, which could inform treatment strategies and impact patient lives. The latter argument was pivotal to our choice of providing feature importance for all models in study III. This allowed us to ascertain which features contributed most to model performance and to understand in more detail the underlying pathology of each disorder.

In the developing nature of machine learning and AI, users of the technology should keep a keen eye on the underlying assumptions as well as the training data used to generate decisions to avoid systematic bias affecting the output[136]. Having explainable models should be preferred, however, letting this be the only criterion for selecting a model seems like an overly simplistic framework and may harm progress in the field.

Digital tools for detection and prognostication of AD

We investigated two low-cost solutions to digital measurements. The tools were specifically chosen for their promising results in AD research as well as their easy application, which is crucial in a population where the disease burden is very high, limiting the amount of diagnostic investigations possible. To this end, applying simple measures seems like a valid approach. However, it is not without disadvantages with regard to the sensitivity and precision of the measures. While hand-held pupillometry does offers an exact evaluation of pupil action, aspects such as the latency of the pupil are more minute than what can be captured with a hand-held unit[51]. This may change in the future if the technology develops, but it may have hindered the detection of subtle differences between groups in our study. On the other hand, even simpler assessments for pupil action are now available through smartphone applications, which account for ambient lighting[137], a

development that ran parallel to our investigations. Controlling for these external factors could prove important for clinical implementation to succeed.

Actigraphy can essentially be boiled down to accelerometrically derived activity counts aligned with time using a chronometer. This type of hardware is also embedded into ubiquitous devices such as smartphones and smartwatches and has been used to assess outcomes after spinal surgery[138]. A similar application within the field of neurodegenerative disorders seems feasible. This development ushers in a new era of medical software and hardware in consumer devices, essentially providing every doctor and patient who owns a smartphone or smartwatch with access to an arsenal of digital biomarkers. Whether this development will lead to better medical care for patients living with a neurodegenerative disease such as AD is yet to be shown.

During the period that the present studies were conducted, new diagnostic guidelines were published by a working group under the Alzheimer's Association[139]. These criteria incorporate the idea of a joint biological and clinical staging of Alzheimer's disease, where the amyloid cascade develops along a diagonal axis of progression representing the accumulation of beta amyloid and subsequent spread of tau pathology with ensuing deterioration of ADL function. Two recent studies[140,141] have sought to validate this model using data from larger cohort studies. The studies together highlight the need for biomarkers that capture aspects of co-pathology to accurately describe clinical phenotypes, as amyloid and tau biomarkers are insufficient to capture the full scope of the disorder. To this end, digital biomarkers as those explored in the present thesis may prove useful, as they allow for more dynamic evaluation of pathological processes and behaviour.

Future studies could incorporate aspects of these new diagnostic guidelines, using the biological framework to assess the impact of new digital biomarkers and what areas of pathology they may highlight in this regard. This could form part of the ongoing effort to validate DHTs in the field of Alzheimer's disease, as well as propel the field for a more elaborate understanding of the biology underlying the clinical phenotypes that we observe in the clinic[121].

Strengths

A significant strength of all included studies was the high degree of real-world application of the used technologies. We included individuals with mild eye disease, a highly prevalent disorder in the oldest age groups, in the studies involving pupillometry, while ensuring that this possible confounder was adjusted for. This ensured that our results had a high generalizability. In addition, for studies II and III we applied relevant diagnostic comparators that could inform on the true diagnostic value in memory clinics, which is one envisioned use case of these digital biomarkers. We evaluated the robustness of our results in studies II, III and IV using approaches for cross-validation, ensuring that our results were robust to unseen data. In all studies, we conducted thorough sensitivity analyses, ensuring that influential confounding variables did not drive the effects we observed.

Limitations

A number of limitations apply to the studies. First of all, there is a risk of selection bias in the studies investigating diagnostic usefulness, as patients were largely included after diagnosis. This could be mitigated in future studies by including patients during diagnostic workup. This seems feasible given the easy application of the investigated markers. In the studies investigating qLRP, we included patients with a history of mild eye disease and allowed for medication that could affect the pupil. This could have impacted the results of our study, as patients may have been more affected by the presence of mild eye disease, due to barriers towards seeking treatment related to their cognitive impairment. We tried to mitigate this by carefully screening for severe eye disease prior to inclusion with ophthalmoscopic examination to rule out larger obscurities. However, we cannot be sure that all relevant eye disorder was described. Applying techniques such as optical coherence tomography may limit the possible confounding by ruling out, for example presence of age-related macular degeneration; however, this technology is not readily available in neurology clinics[142]. To ensure that mild eye disease was not a significant confounder, we carried out sensitivity analyses in studies II and IV.

We also lack external validity of our findings, meaning that we did not include cohorts for replicating our findings and reproducing estimates. This was in part mitigated by our

application of cross-validation in studies II and IV, however, generalizing results to other populations is highly necessary before clinical implementation can be envisioned. Test-retest studies rarely use external validation, although this limitation could also apply to study I.

Our choice of diagnostic comparators was based on the fact that these two disorders, DLB and CVD, represent the second most common neurodegenerative dementia[143] and the second most common cause of dementia[144], respectively. However, other relevant comparators could have been included in the diagnostic study. Frontotemporal dementia is also prevalent, especially in young-onset dementia[145], which were also captured by our inclusion criteria, and could have been included as a comparator. Likewise, the healthy control group were individuals without cognitive complaints. These individuals would rarely be evaluated in memory clinics, whereas mood disorders resembling neurodegenerative disorders, is more often evaluated and confused with AD[146]. These diagnostic groups should be included in future studies to more fully evaluate the diagnostic usefulness in a memory clinic setting.

We applied a machine learning model in study III, which takes on inherent interpretability issues. First, model algorithms select relevant variables according to their ability to predict the outcome repeatedly and robustly, and even before this process, calculating features for the machine to sort through may lead to difficulties in interpreting the results. This phenomenon, termed the “black box” problem[147] of machine learning and AI algorithms, poses the question of whether an understanding is needed. In our study, we found that certain walking patterns and the distribution of steps in time had an impact on model performance for differentiating dementia disorders. Whether these variables were truly the best predictors or had a similarity with the overall mean values of the same parameters is difficult to ascertain, given the algorithm we chose for feature selection. It is, however, highly likely that several features exhibited a high degree of collinearity, which is difficult to handle when training an ML model.

We also applied a circadian framework to already complex activity profiles, perhaps further complicating the data than what was necessary, and a sparse approach as the one applied in study II could also have been tried to see if simpler solutions to our prediction

models could be found. We mitigated this aspect by comparing our machine learning classifier with a single-feature classifier, using only the overall activity levels. This approach allowed us to evaluate whether a simpler model using only the strongest predictor in the dataset was better. Following our study, hypothesis-driven approaches may be tried to elucidate the true relationships between circadian activity patterns and dementia disorders.

6. Conclusions and future perspectives

In this thesis, results from four clinical studies on two digital biomarkers for AD diagnosis and prognosis have been motivated, summarized and discussed. In study I, we found that qLRP could be measured reliably for most pupillometric parameters in the intended use population. These results demonstrated a proof-of-concept that qLRP had acceptable reliability for employment in a memory clinic setting for clinical evaluation of patients under suspicion of neurodegenerative disease. We also found low, intra-subject variability, which suggested the possibility of exploring longitudinal dynamic changes using qLRP. This study paved the way for further exploring qLRP as a diagnostic and prognostic marker.
In study II, we likewise provide proof-of-concept that qLRP could inform on diagnosis in a memory clinic setting using valid comparators to ascertain the differential diagnostic value of the tool. While the diagnostic performance of the models was moderate, they could provide a novel biomarker that could supplement traditional biomarkers, although this aspect remains to be shown in studies that evaluate the added value of qLRP to a diagnostic process. Finally, aided by the results obtained in study I, we provided the first indications that qLRP as a marker of arousal states could hold prognostic information in the early stages of AD. The study was exploratory in nature and provided several clues to hypotheses regarding the prognostic properties and what underlying pathology may lie behind the observations. One caveat was the low robustness of these findings and the inability to translate its prognostic value across progression outcome modalities.
Future studies should focus on providing the neuroimaging correlate for the observed prognostic properties, using for example quantitative neuroimaging that captures key aspects of AD pathology. For actigraphy, we demonstrated for the first time its usefulness

as a diagnostic marker of AD, especially for differential diagnostic purposes using a ML classifier. The ability to separate dementia disorders with high accuracy was promising for a future clinical implementation, as we also observed dynamic properties of the markers, capturing diverse aspects of pathology. These findings require further validation in other cohorts, but they represent a first step towards the implementation of a digital biomarker in the memory clinic setting using an ML approach.

Our results point towards the diagnostic and prognostic usefulness of the two digital biomarkers in question, filling an important knowledge gap regarding digital tools in AD[15]. It remains to be shown whether actigraphy holds similar or even better prognostic capability than qLRP in AD, which is the focus of future studies within the same cohort. Being stepping stones towards clinical implementation, our findings should not stand alone, and a large body of work awaits before a digitized memory clinic can be realized utilizing the wealth of DHTs[12]. This work includes rigorous testing in diverse populations, comparison with gold standard diagnostic methods, and other emerging biomarkers, such as blood-based biomarkers, which have gone through a similar development in recent years[148]. However, given the scalability of DHTs this future may not be far from the present.

All together, we present results that support the further exploration of digital biomarkers in AD, representing an exciting field of research that until recently was only in its infancy, but perhaps now is approaching adolescence.

7. References

1. Gramkow, M. H., Clemmensen, F. K., Waldemar, G., Hasselbalch, S. G. & Frederiksen, K. S. Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort. *J Neurol Sci* **456**, 122856 (2024).
2. Gramkow, M. H. *et al.* Diagnostic performance of light reflex pupillometry in Alzheimer's disease. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring* **16**, 1–11 (2024).
3. Gramkow, M. H. *et al.* Diagnostic performance of actigraphy in Alzheimer's disease using a machine learning classifier – a cross-sectional memory clinic study. *Alzheimer's Research & Therapy 2025 17:1* **17**, 1–12 (2025).
4. Gustavsson, A. *et al.* Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's and Dementia* **19**, 658–670 (2023).
5. Blennow, K., de Leon, M. J. & Zetterberg, H. Alzheimer's disease. *The Lancet* **368**, 387–403 (2006).
6. Geneva: World Health & Organization. Global action plan on the public health response to dementia 2017–2025 Licence: CC BY-NC-SA 3.0 IGO. (2017).
7. van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine* **388**, 9–21 (2023).
8. Sims, J. R. *et al.* Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA* **330**, 512–527 (2023).
9. Dubois, B., Padovani, A., Scheltens, P., Rossi, A. & Agnello, G. D. Timely diagnosis for alzheimer's disease: A literature review on benefits and challenges. *Journal of Alzheimer's Disease* **49**, 617–631 (2015).
10. Jönsson, L., Tate, A., Frisell, O. & Wimo, A. The Costs of Dementia in Europe: An Updated Review and Meta-analysis. *PharmacoEconomics 2022 41:1* **41**, 59–75 (2022).
11. Jönsson, L. *et al.* The affordability of lecanemab, an amyloid-targeting therapy for Alzheimer's disease: an EADC-EC viewpoint. *The Lancet Regional Health - Europe* **29**, 100657 (2023).

12. Gramkow, M. H., Waldemar, G. & Frederiksen, K. S. The Digitized Memory Clinic. *Nat Rev Neurol* 1–9 (2024) doi:10.1038/s41582-024-01033-y.
13. Au, R., Kolachalama, V. B. & Paschalidis, I. C. Redefining and Validating Digital Biomarkers as Fluid, Dynamic Multi-Dimensional Digital Signal Patterns. *Front Digit Health* **3**, 751629 (2022).
14. Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *NPJ Digit Med* **2**, 9 (2019).
15. Gramkow, M. H., Gleerup, H. S., Simonsen, A. H., Waldemar, G. & Frederiksen, K. S. Digital biomarkers in early Alzheimer's disease from wearable or portable technology: a scoping review. *medRxiv* 2025.05.02.25326845 (2025) doi:10.1101/2025.05.02.25326845.
16. Vasudevan, S., Saha, A., Tarver, M. E. & Patel, B. Digital biomarkers: Convergence of digital health technologies and biomarkers. *NPJ Digit Med* **5**, 3–5 (2022).
17. Böheim, R., Horvath, T., Leoni, T. & Spielauer, M. The Impact of Health and Education on Labor Force Participation in Aging Societies: Projections for the United States and Germany from Dynamic Microsimulations. *Popul Res Policy Rev* **42**, 39 (2023).
18. Skender, S. *et al.* Accelerometry and physical activity questionnaires - A systematic review. *BMC Public Health* **16**, 1–10 (2016).
19. Qi, W. *et al.* Alzheimer's disease digital biomarkers multidimensional landscape and AI model scoping review. *npj Digital Medicine 2025 8:1* **8**, 1–33 (2025).
20. van Gils, A. M. *et al.* Use of a digital tool to support the diagnostic process in memory clinics-a usability study. *Alzheimers Res Ther* **16**, 75 (2024).
21. ASCHER, K. W. The first pupillary light reflex test ever performed. *Trans Am Ophthalmol Soc* **60**, 53–59 (1962).
22. Oliva, M. & Anikin, A. Pupil dilation reflects the time course of emotion recognition in human vocalizations. *Sci Rep* **8**, 1–10 (2018).
23. Bradley, M. M., Miccoli, L., Escrig, M. A. & Lang, P. J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, 602–607 (2008).

24. Carro-Domínguez, M. *et al.* Pupil size reveals arousal level fluctuations in human sleep. *Nature Communications 2025 16:1* **16**, 1–14 (2025).
25. Papesh, M. H. & Goldinger, S. D. *Modern Pupillometry*. (2024).
26. You, S., Hong, J. H. & Yoo, J. Analysis of pupillometer results according to disease stage in patients with Parkinson's disease. *Sci Rep* **11**, 1–6 (2021).
27. Troiani, V. The future of quantitative pupillometry in health and disease. *Clinical Autonomic Research* **30**, 11–12 (2020).
28. Romagnoli, M. *et al.* Chromatic Pupillometry Findings in Alzheimer's Disease. *Front Neurosci* **14**, 1–10 (2020).
29. Chougule, P. S., Najjar, R. P., Finkelstein, M. T., Kandiah, N. & Milea, D. Light-induced pupillary responses in Alzheimer's disease. *Front Neurol* **10**, 1–12 (2019).
30. Scinto, L. F. M. *et al.* Selective cell loss in Edinger-Westphal in asymptomatic elders and Alzheimer's patients. *Neurobiol Aging* **22**, 729–736 (2001).
31. Mavroudis, I. A. *et al.* Dendritic and spinal alterations of neurons from Edinger-Westphal nucleus in Alzheimer's disease. *Folia Neuropathol* **52**, 197–204 (2014).
32. Scinto, L. F. M. *et al.* Focal pathology in the Edinger-Westphal nucleus explains pupillary hypersensitivity in Alzheimer's disease. *Acta Neuropathol* **97**, 557–564 (1999).
33. McDougal DH & Gamlin PD. Autonomic control of the eye. *Compr Physiol* **5**, 439–73 (2015).
34. La Morgia, C. *et al.* Melanopsin retinal ganglion cell loss in Alzheimer disease. *Ann Neurol* **79**, 90–109 (2016).
35. Mure, L. S. Intrinsically Photosensitive Retinal Ganglion Cells of the Human Retina. *Front Neurol* **12**, 1–10 (2021).
36. Esquiva, G. & Hannibal, J. Melanopsin-expressing retinal ganglion cells in aging and disease. *Histol Histopathol* **34**, 1299–1311 (2019).
37. La Morgia, C. *et al.* Multimodal investigation of melanopsin retinal ganglion cells in Alzheimer's disease. *Ann Clin Transl Neurol* **10**, 918–932 (2023).
38. Davis, M. R. *et al.* Retinal ganglion cell vulnerability to pathogenic tau in Alzheimer's disease. **13**, (2025).

39. Campbell, F. W. & Gregory, A. H. Effect of Size of Pupil on Visual Acuity. *Nature* **187**, 1121–1123 (1960).
40. Huang, R. & Clewett, D. The Locus Coeruleus: Where Cognitive and Emotional Processing Meet the Eye. *Modern Pupillometry* 3–75 (2024) doi:10.1007/978-3-031-54896-3_1.
41. Nobukawa, S. *et al.* Pupillometric Complexity and Symmetricity Follow Inverted-U Curves Against Baseline Diameter Due to Crossed Locus Coeruleus Projections to the Edinger-Westphal Nucleus. *Front Physiol* **12**, 614479 (2021).
42. Theofilas, P. *et al.* Locus coeruleus volume and cell population changes during Alzheimer's disease progression: A stereological study in human postmortem brains with potential implication for early-stage biomarker discovery. *Alzheimer's and Dementia* **13**, 236–246 (2017).
43. Grinberg, L. T. *et al.* The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer̈s disease. A precocious onset? *Neuropathol Appl Neurobiol* **35**, 406–416 (2009).
44. Cazettes, F., Reato, D., Morais, J. P., Renart, A. & Mainen, Z. F. Phasic Activation of Dorsal Raphe Serotonergic Neurons Increases Pupil Size. *Current Biology* **31**, 192-197.e4 (2021).
45. Nestor, P. J., Caine, D., Fryer, T. D., Clarke, J. & Hodges, J. R. The topography of metabolic deficits in posterior cortical atrophy (the visual variant of Alzheimer's disease) with FDG-PET. *J Neurol Neurosurg Psychiatry* **74**, 1521–1529 (2003).
46. Peinkhofer, C., Knudsen, G. M., Moretti, R. & Kondziella, D. Cortical modulation of pupillary function: systematic review. *PeerJ* **7**, e6882 (2019).
47. Grinberg, L. T., Rueb, U. & Heinsen, H. Brainstem: Neglected locus in neurodegenerative diseases. *Front Neurol* **JUL**, 1–9 (2011).
48. Ehrenberg, A. J. *et al.* Quantifying the accretion of hyperphosphorylated tau in the locus coeruleus and dorsal raphe nucleus: the pathological building blocks of early Alzheimer's disease. *Neuropathol Appl Neurobiol* **43**, 393–408 (2017).

49. Steinhauer, S. R., Condray, R. & Kasparek, A. Cognitive modulation of midbrain function: Task-induced reduction of the pupillary light reflex. *International Journal of Psychophysiology* **39**, 21–30 (2000).
50. Scinto, L. F. M. *et al.* Focal pathology in the Edinger-Westphal nucleus explains pupillary hypersensitivity in Alzheimer's disease. *Acta Neuropathol* **97**, 557–564 (1999).
51. Bergamin, O. & Kardon, R. H. Latency of the pupil light reflex: Sample rate, stimulus intensity, and variation in normal subjects. *Invest Ophthalmol Vis Sci* **44**, 1546–1554 (2003).
52. Ferencova, N., Visnovcova, Z., Olexova, L. B. & Tonhajzerova, I. Eye Pupil – A Window into Central Autonomic Regulation via Emotional/Cognitive Processing. *Physiol Res* **70**, S669 (2021).
53. Scinto, L. F. M. F. M. *et al.* Selective cell loss in Edinger-Westphal in asymptomatic elders and Alzheimer's patients. *Neurobiol Aging* **22**, 729–736 (2001).
54. Scinto, L. F. M. *et al.* Focal pathology in the Edinger-Westphal nucleus explains pupillary hypersensitivity in Alzheimer's disease. *Acta Neuropathol* **97**, 557–564 (1999).
55. Prettyman, R., Bitsios, P. & Szabadi, E. Altered pupillary size and darkness and light reflexes in Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **62**, 665–8 (1997).
56. Ferrario, E. *et al.* Is videopupillography useful in the diagnosis of Alzheimer's disease? *Neurology* **50**, 642–644 (1998).
57. Fotiou, D. F. *et al.* Pupil reaction to light in Alzheimer's disease: Evaluation of pupil size changes and mobility. *Aging Clin Exp Res* **19**, 364–371 (2007).
58. Fotiou, D. F. *et al.* Effect of age on pupillary light reflex: Evaluation of pupil mobility for clinical practice and research. *Electromyogr Clin Neurophysiol* **47**, 11–22 (2007).
59. Nyholm, B. *et al.* Superior reproducibility and repeatability in automated quantitative pupillometry compared to standard manual assessment, and quantitative pupillary response parameters present high reliability in critically ill cardiac patients. *PLoS One* **17**, 1–15 (2022).

60. NICE. Medtech briefing on NPi-200 pupillometer. https://www.nice.org.uk/advice/mib235 (2020).
61. Fotiou, D., Kaltsatou, A., Tsiptsios, D. & Nakou, M. Evaluation of the cholinergic hypothesis in Alzheimer's disease with neuropsychological methods. *Aging Clin Exp Res* **27**, 727–733 (2015).
62. Tales, A. *et al.* The pupillary light reflex in aging and alzheimer's disease. *Aging Clin Exp Res* **13**, 473–478 (2001).
63. Frost, S. *et al.* Pupil Response Biomarkers for Early Detection and Monitoring of Alzheimer's Disease. *Curr Alzheimer Res* **10**, 931–939 (2013).
64. Bittner, D. M., Wieseler, I., Wilhelm, H., Riepe, M. W. & Müller, N. G. Repetitive Pupil Light Reflex: Potential Marker in Alzheimer's Disease? *Journal of Alzheimer's Disease* **42**, 1469–1477 (2014).
65. Granholm, E. *et al.* Tropicamide effects on pupil size and pupillary light reflexes in Alzheimer's and Parkinson's disease. *International Journal of Psychophysiology* **47**, 95–115 (2003).
66. Frost, S. *et al.* Pupil Response Biomarkers Distinguish Amyloid Precursor Protein Mutation Carriers from Non-Carriers. *Curr Alzheimer Res* **10**, 790–796 (2013).
67. Martin, J. L. & Hakim, A. D. Wrist actigraphy. *Chest* **139**, 1514–1527 (2011).
68. Smith, M. T. *et al.* Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Clinical Practice Guideline. *Journal of Clinical Sleep Medicine* **14**, 1231–1237 (2018).
69. Staying current with actigraphy devices for sleep-wake monitoring - American Academy of Sleep Medicine – Association for Sleep Clinicians and Researchers. https://aasm.org/staying-current-with-actigraphy-devices-for-sleep-wake-monitoring/.
70. Teicher, M. H. Actigraphy and motion analysis: new tools for psychiatry. *Harv Rev Psychiatry* **3**, 18–35 (1995).
71. Li, P. *et al.* Fractal regulation and incident Alzheimer's disease in elderly individuals. *Alzheimer's and Dementia* **14**, 1114–1125 (2018).

72. Gao, L. *et al.* Fractal motor activity regulation and sex differences in preclinical Alzheimer's disease pathology. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **13**, 1–9 (2021).
73. Filardi, M. *et al.* Sleep and circadian rhythm disruptions in behavioral variant frontotemporal dementia. *Alzheimer's & Dementia* **20**, 1966 (2024).
74. Lim, A. S. P. P., Kowgier, M., Yu, L., Buchman, A. S. & Bennett, D. A. Sleep fragmentation and the risk of incident Alzheimer's disease and cognitive decline in older persons. *Sleep* **36**, 1027–1032 (2013).
75. Buchman, A. S. *et al.* Brain pathology is related to total daily physical activity in older adults. *Neurology* **90**, E1911–E1919 (2018).
76. Wu, R. *et al.* Fragmentation of rest periods, astrocyte activation, and cognitive decline in older adults with and without Alzheimer's disease. *Alzheimer's and Dementia* **19**, 1888–1900 (2023).
77. Li, P. *et al.* Circadian disturbances in Alzheimer's disease progression: a prospective observational cohort study of community-based older adults. *Lancet Healthy Longev* **1**, e96–e105 (2020).
78. Kaneshwaran, K. *et al.* Sleep fragmentation, microglial aging, and cognitive impairment in adults with and without Alzheimer's dementia. *Sci Adv* **5**, eaax7331 (2019).
79. Leng, Y., Musiek, E. S., Hu, K., Cappuccio, F. P. & Yaffe, K. Association between circadian rhythms and neurodegenerative diseases. *Lancet Neurol* **18**, 307–318 (2019).
80. Vanderheyden, W. M., Lim, M. M., Musiek, E. S. & Gerstner, J. R. Alzheimer's disease and sleep–wake disturbances: Amyloid, astrocytes, and animal models. *Journal of Neuroscience* **38**, 2901–2910 (2018).
81. Musiek, E. S. *et al.* Circadian Rest-Activity Pattern Changes in Aging and Preclinical Alzheimer Disease. *JAMA Neurol* **75**, 582–590 (2018).
82. Wang, J. L. *et al.* Suprachiasmatic neuron numbers and rest-activity circadian rhythms in older humans. *Ann Neurol* **78**, 317–322 (2015).

83. Chen, Y., Chen, T. & Hou, R. Locus coeruleus in the pathogenesis of Alzheimer's disease: A systematic review. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **8**, 1–12 (2022).
84. Van Egroo, M., van Someren, E. J. W., Grinberg, L. T., Bennett, D. A. & Jacobs, H. I. L. Associations of 24-Hour Rest-Activity Rhythm Fragmentation, Cognitive Decline, and Postmortem Locus Coeruleus Hypopigmentation in Alzheimer's Disease. *Ann Neurol* **95**, 653–664 (2024).
85. Nobis, L. & Husain, M. Apathy in Alzheimer's disease. *Curr Opin Behav Sci* **22**, 7–13 (2018).
86. Le Heron, C., Apps., M. A. J. & Husain, M. The anatomy of apathy: A neurocognitive framework for amotivated behaviour. *Neuropsychologia* **118**, 54–67 (2018).
87. Moretti, R. & Signori, R. Neural correlates for apathy: Frontal-prefrontal and parietal cortical- subcortical circuits. *Front Aging Neurosci* **8**, (2016).
88. Kuhlmei, A. *et al.* Actigraphic daytime activity is reduced in patients with cognitive impairment and apathy. *European Psychiatry* **28**, 94–97 (2013).
89. Aharon-Peretz, J. *et al.* Sleep-wake cycles in multi-infarct dementia and dementia of the Alzheimer type. *Neurology* **41**, 1616–1619 (1991).
90. Varma, V. R., Chuang, Y. F., Harris, G. C., Tan, E. J. & Carlson, M. C. Low-intensity daily walking activity is associated with hippocampal volume in older adults. *Hippocampus* **25**, 605–615 (2015).
91. Varma, V. R. & Watts, A. Daily Physical Activity Patterns During the Early Stage of Alzheimer's Disease. *Journal of Alzheimer's Disease* **55**, 659–667 (2016).
92. Varma, V. R. *et al.* Continuous gait monitoring discriminates community-dwelling mild Alzheimer's disease from cognitively normal controls. *Alzheimer's and Dementia: Translational Research and Clinical Interventions* **7**, e12131 (2021).
93. Guarnieri, B. *et al.* Multicenter Study on Sleep and Circadian Alterations as Objective Markers of Mild Cognitive Impairment and Alzheimer's Disease Reveals Sex Differences. *J Alzheimers Dis* **78**, 1707–1719 (2020).

94. Hatfield, C. F., Herbert, J., Van Someren, E. J. W., Hodges, J. R. & Hastings, M. H. Disrupted daily activity/rest cycles in relation to daily cortisol rhythms of home-dwelling patients with early Alzheimer's dementia. *Brain* **127**, 1061–1074 (2004).
95. Lu, Z., Harris, T. B., Shiroma, E. J., Leung, J. & Kwok, T. Patterns of Physical Activity and Sedentary Behavior for Older Adults with Alzheimer's Disease, Mild Cognitive Impairment, and Cognitively Normal in Hong Kong. *J Alzheimers Dis* **66**, 1453–1462 (2018).
96. Park, J. E. *et al.* Differential associations of age and Alzheimer's disease with sleep and rest-activity rhythms across the adult lifespan. *Neurobiol Aging* **101**, 141–149 (2021).
97. Brink-Kjaer, A. *et al.* Ambulatory Detection of Isolated Rapid-Eye-Movement Sleep Behavior Disorder Combining Actigraphy and Questionnaire. *Movement Disorders* **38**, 82–91 (2023).
98. Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W. & O'Sullivan, J. M. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics* **2**, 927312 (2022).
99. Yousefzadehfard, Y., Wechsler, B. & DeLorenzo, C. Human circadian rhythm studies: Practical guidelines for inclusion/exclusion criteria and protocol. *Neurobiol Sleep Circadian Rhythms* **13**, 100080 (2022).
100. Ong, C., Hutch, M. & Smirnakis, S. The Effect of Ambient Light Conditions on Quantitative Pupillometry. *Neurocrit Care* **30**, 316–321 (2019).
101. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270–9 (2011).
102. McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers & Dementia* **7**, 263–269 (2011).

103. McKeith, I. G. *et al.* Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology* **89**, 88–100 (2017).
104. McKeith, I. G. *et al.* Research criteria for the diagnosis of prodromal dementia with Lewy bodies. *Neurology* **94**, 743–755 (2020).
105. Sachdev, P. *et al.* Diagnostic criteria for vascular cognitive disorders: A VASCOG statement. *Alzheimer Dis Assoc Disord* **28**, 206–218 (2014).
106. Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W. & Hodges, J. R. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* **55**, 1613–1620 (2000).
107. Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–198 (1975).
108. Gorno-Tempini, M. L. *et al.* Classification of primary progressive aphasia and its variants. *Neurology* **76**, 1006–1014 (2011).
109. Greenberg, S. M. & Charidimou, A. Diagnosis of cerebral amyloid angiopathy evolution of the Boston criteria. *Stroke* **49**, 491–497 (2018).
110. Peinkhofer, C. *et al.* Influence of Strategic Cortical Infarctions on Pupillary Function. *Front Neurol* **9**, 1–8 (2018).
111. Guedj, E. *et al.* EANM procedure guidelines for brain PET imaging using [18F]FDG, version 3. *Eur J Nucl Med Mol Imaging* **49**, 632–651 (2022).
112. Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412–2414 (1993).
113. Gramkow, M. H. *et al.* Prognostic value of complementary biomarkers of neurodegeneration in a mixed memory clinic cohort. *PeerJ* **2020**, e9498 (2020).
114. Peng, H., Long, F. & Ding, C. Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans Pattern Anal Mach Intell* **27**, 1226–1238 (2005).
115. Hicks, S. A. *et al.* On evaluation metrics for medical applications of artificial intelligence. *Sci Rep* **12**, 1–9 (2022).

116. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.r-project.org/ (2021).
117. McGraw, K. O. & Wong, S. P. Forming Inferences about Some Intraclass Correlation Coefficients. *Psychol Methods* **1**, 30–46 (1996).
118. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* **15**, 155–163 (2016).
119. Taylor, W. R. *et al.* Quantitative pupillometry, a new technology: Normative data and preliminary observations in patients with acute head injury - Technical note. *J Neurosurg* **98**, 205–213 (2003).
120. Dębska, L. & Białek, A. Lighting conditions as the occupational health related issue – case study. *MATEC Web of Conferences* **354**, 00059 (2022).
121. Goldsack, J. C. *et al.* Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit Med* **3**, (2020).
122. Okkels, N. *et al.* Cholinergic changes in Lewy body disease: implications for presentation, progression and subtypes. *Brain* **147**, 2308–2324 (2024).
123. Mayà, G. *et al.* Post-mortem neuropathology of idiopathic rapid eye movement sleep behaviour disorder: a case series. *Lancet Neurol* **23**, 1238–1251 (2024).
124. Alber, J. *et al.* White matter hyperintensities in vascular contributions to cognitive impairment and dementia (VCID): Knowledge gaps and opportunities. *Alzheimer's and Dementia: Translational Research and Clinical Interventions* **5**, 107–117 (2019).
125. Granholm, E. L. *et al.* Pupillary Responses as a Biomarker of Early Risk for Alzheimer's Disease. *Journal of Alzheimer's Disease* **56**, 1419–1428 (2017).
126. McLaughlin, J., Scotton, W. J., Ryan, N. S., Hardy, J. A. & Shoai, M. Assessing clinical progression measures in Alzheimer's disease trials: A systematic review and meta-analysis. *Alzheimer's & Dementia* **20**, 8673 (2024).
127. Landau, S. M. *et al.* Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging* **32**, 1207–1218 (2011).
128. Musiek, E. S. & Ju, Y.-E. S. Targeting Sleep and Circadian Function in the Prevention of Alzheimer Disease. *JAMA Neurol* **79**, 835–836 (2022).

129. Musiek, E. S., Xiong, D. D. & Holtzman, D. M. Sleep, circadian rhythms, and the pathogenesis of Alzheimer Disease. *Exp Mol Med* **47**, e148 (2015).
130. Van Someren, E. J. W. *et al.* Medial temporal lobe atrophy relates more strongly to sleep-wake rhythm fragmentation than to age or any other known risk. *Neurobiol Learn Mem* **160**, 132–138 (2019).
131. Oosterman, J. M., van Someren, E. J. W., Vogels, R. L. C., Van Harten, B. & Scherder, E. J. A. Fragmentation of the rest-activity rhythm correlates with age-related cognitive deficits. *J Sleep Res* **18**, 129–135 (2009).
132. Patterson, M. R. *et al.* 40 years of actigraphy in sleep medicine and current state of the art algorithms. *NPJ Digit Med* **6**, 1–7 (2023).
133. Blennow, K. A Review of Fluid Biomarkers for Alzheimer's Disease: Moving from CSF to Blood. *Neurol Ther* **6**, 15 (2017).
134. Pichet Binette, A. *et al.* Associations between misfolded alpha-synuclein aggregates and Alzheimer's disease pathology in vivo. *Alzheimer's and Dementia* (2024) doi:10.1002/alz.14225.
135. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **23**, 18 (2020).
136. Belle, V. & Papantonis, I. Principles and Practice of Explainable Machine Learning. *Front Big Data* **4**, 688969 (2021).
137. Bogucki, A. *et al.* Machine learning approach for ambient-light-corrected parameters and the Pupil Reactivity score in smartphone-based pupillometry. *Frontiers in Neurology* **15**, 1363190 (2024).
138. Voglis, S. *et al.* Smartphone-based real-life activity data for physical performance outcome in comparison to conventional subjective and objective outcome measures after degenerative lumbar spine surgery. *Brain and Spine* **2**, (2022).
139. Jack, C. R. *et al.* Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's and Dementia* **20**, 5143–5169 (2024).

140. Mendes, A. J. *et al.* Validating the Amyloid Cascade Through the Revised Criteria of Alzheimer's Association Workgroup 2024 for Alzheimer Disease. *Neurology* **104**, e213675 (2025).
141. Binette, A. P. *et al.* Evaluation of the Revised Criteria for Biological and Clinical Staging of Alzheimer Disease. *JAMA Neurol* (2025) doi:10.1001/JAMANEUROL.2025.1100.
142. Langlo, C. S., Amin, A. & Park, S. S. Optical coherence tomography retinal imaging: narrative review of technological advancements and clinical applications. *Ann Transl Med* **13**, 17 (2025).
143. Palushaj, B., Lewis, S. J. G. & Abdelnour, C. What is the future for dementia with Lewy bodies? *J Neurol* **272**, (2025).
144. Wolters, F. J. & Arfan Ikram, M. Epidemiology of Vascular Dementia: Nosology in a Time of Epiomics. *Arterioscler Thromb Vasc Biol* **39**, 1542–1549 (2019).
145. Onyike, C. U. & Diehl-Schmid, J. The Epidemiology of Frontotemporal Dementia. *Int Rev Psychiatry* **25**, 130 (2013).
146. Marvel, C. L. & Paradiso, S. Cognitive and neurological impairment in mood disorders. *Psychiatr Clin North Am* **27**, 19 (2004).
147. Xu, H. & Shuttleworth, K. M. J. Medical artificial intelligence and the black box problem: a view based on the ethical principle of "do no harm". *Intelligent Medicine* **4**, 52–57 (2024).
148. Teunissen, C. E. *et al.* Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. *Lancet Neurol* **21**, 66–77 (2022).

Appendix A | Journal articles

STUDY I[1] Gramkow, M. H., Clemmensen, F. K., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2024). Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort. *Journal of the Neurological Sciences*, *456*(March), 122856. https://doi.org/10.1016/j.jns.2023.122856

STUDY II[2] Gramkow, M. H., Clemmensen, F. K., Sjælland, N. S., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2024). Diagnostic performance of light reflex pupillometry in Alzheimer's disease. Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring, 16(3), 1–11. https://doi.org/10.1002/dad2.12628

STUDY III[3] Gramkow, M. H., Brink-Kjær, A., Clemmensen, F. K., Sjælland, N. S., Waldemar, G., Jennum, P., Hasselbalch, S. G., & Frederiksen, K. S. (2025). Diagnostic performance of actigraphy in Alzheimer's disease using a machine learning classifier – a cross-sectional memory clinic study. Alzheimer's Research & Therapy 2025 17:1, 17(1), 1–12. https://doi.org/10.1186/S13195-025-01751-5

STUDY IV Gramkow, M. H., Clemmensen, F. K., Lindberg, U., Law, I., Henriksen, O. M., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S.: Prognostic value of light reflex pupillometry in Alzheimer's disease – a longitudinal cohort study (***Accepted*** *June 22, 2025, in Alzheimer's Research and Therapy)*

STUDY I

Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort

Gramkow, M. H., Clemmensen, F. K., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2024). Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort. Journal of the Neurological Sciences, 456(March), 122856. DOI: https://doi.org/10.1016/j.jns.2023.122856

The supplementary material related to this journal article can be accessed at:
https://doi.org/10.1016/j.jns.2023.122856
or by scanning the QR-code below

Journal of the Neurological Sciences 456 (2024) 122856

Contents lists available at ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort

Mathias Holsey Gramkow [a,*], Frederikke Kragh Clemmensen [a], Gunhild Waldemar [a,b], Steen Gregers Hasselbalch [a,b], Kristian Steen Frederiksen [a,b]

[a] Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
[b] Dept. of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

ARTICLE INFO

Keywords:
Quantitative pupillometry
Alzheimer's disease
Memory clinic
Test-retest
Light reflex
Neurodegeneration

ABSTRACT

Background: Quantitative light reflex pupillometry (qLRP) may be a promising digital biomarker in neurodegenerative diseases such as Alzheimer's disease (AD), as neuropathological changes have been found in the midbrain structures governing the light reflex. Studies investigating test-retest reliability and short-term, intra-subject variability of qLRP in these patient groups are missing. Our objective was therefore to investigate the test-retest reliability and short-term, intra-subject variability of qLRP in a memory clinic setting, where patients with neurodegenerative disease are frequently evaluated.

Methods: Test-retest reliability study. We recruited patients from a tertiary memory clinic and qLRP was carried out at a baseline visit and then repeated on day 3–14 and on day 21–35 using a hand-held pupillometer. We evaluated the test-retest reliability of qLRP by calculating intraclass correlation coefficients (ICCs) and intra-subject, short-term variability by fitting linear mixed models. We compared ICCs for subgroups based on age, sex, disease severity (MCI vs. mild dementia), AD diagnosis, and amount of neurodegeneration (cerebrospinal fluid-total tau levels).

Results: In total, 40 patients (mean age 72 years, 15 female, 22 with mild dementia) were included in the study. We found good-excellent reliability (ICC range 0.86–0.93) for most qLRP parameters. qLRP parameters exhibited limited intra-subject variability and we found no large sources of variability when examining subgroups.

Conclusion: qLRP was found to have acceptable test-retest reliability and the study results pave the way for research using longitudinal or cross-sectional measurements to assess the construct in identifying and prognosticating neurodegenerative diseases.

1. Introduction

Quantitative light reflex pupillometry (qLRP) has been proposed as a novel marker in neurodegenerative diseases such as Alzheimer's disease (AD) [1], but studies investigating the test-retest reliability of the method in these patient groups are missing [2,3]. Also, in cross-sectional studies, the aspect of intra-subject, short-term variability may be relevant. Test-retest reliability is defined as the ability of a measurement method to quantify the same property under replicated conditions [4]. The intra-subject variability is defined as the amount that a given measured parameter varies for each repeated measurement in a single individual [5].

In AD in particular, prognostic and disease-tracking markers as well as case-finding tools are urgently needed to accommodate the advent of disease-modifying anti-amyloid therapies, which have been recently approved for use [6]. Due to its low-cost, non-invasive nature, qLRP could be a promising digital disease tracking and prognostic biomarker and aid in case-finding, but the test-retest reliability and the intra-subject, short-term variability of qLRP when used in neurodegenerative diseases are essentially not known.

Light reflex pupillometry is the objective measurement of the pupillary response (contraction of the m. constrictor pupillae) to external light stimuli. It can be carried out bedside using a hand-held point-and-shoot pupillometer [7]. The reflex arc is, with modulation by cortical areas such as the frontal eye field, governed by midbrain structures. Most important for the reflex is the Edinger Westphal nucleus, which is

* Corresponding author at: Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital – Rigshospitalet, Inge Lehmans Vej 8, DK-2100 Copenhagen, Denmark.
E-mail address: mathias.holsey.gramkow@regionh.dk (M.H. Gramkow).

https://doi.org/10.1016/j.jns.2023.122856
Received 29 August 2023; Received in revised form 18 December 2023; Accepted 18 December 2023
Available online 22 December 2023

shown to undergo neurodegeneration in patients with AD [8–10]. qLRP has been investigated in recent studies [11,12] as a marker for AD, describing that several metrics of qLRP are affected in AD, possibly as a result of disrupted autonomic function [1]. In healthy individuals, the test-retest of qLRP has shown excellent reliability [7], and also in critically ill patients [13], but no studies exist on the test-retest reliability and short-term, intra-subject variability in patients suspected of neurodegenerative disease. In addition to this, the underlying factors that might give rise to a posited increased variability are not known.

Patients with neurodegenerative disease exhibit progressive neuronal dysfunction [14] and thus may exhibit higher degrees of variability in qLRP parameters. The specific presence of neurodegeneration as reflected by neurodegenerative markers, such as tau protein measured in cerebrospinal fluid, could for example be related to the intra-subject variability, but this has not been investigated previously. Further, patients with dementia are more often treated with antipsychotic or antidepressant medication, which could influence the pupillary state through modulation of the autonomic nervous system and thereby qLRP parameters [15,16]. Therefore, studies that quantify intra-subject variability in patients undergoing diagnostic evaluation for neurodegenerative disease are needed before the introduction of longitudinal measurements to clinical practice in individual patients.

In the present study, we examined patients in a memory clinic who were undergoing diagnostic evaluation for neurodegenerative disease, at serial visits over a short period to investigate the test-retest reliability and intra-subject variability of qLRP. We hypothesized that qLRP could be reliably measured and that measurements exhibited limited short-term, intra-subject variability.

2. Materials and methods

2.1. Study design

Test-retest reliability study.

2.2. Participants

Patients referred to a tertiary memory clinic for diagnostic evaluation for a possible neurodegenerative disorder were consecutively recruited with the following inclusion criteria: undergoing lumbar puncture due to suspicion of neurodegenerative disease, a Mini-Mental State Examination (MMSE) total score $>$ 20, could cooperate to the qLRP procedure. We applied the following exclusion criteria: no severe bilateral ophthalmological disorders (severe untreated cataract, glaucoma, or similar), no history of stroke within the last 3 months, no previous or current major psychiatric conditions (schizophrenia, bipolar affective disorder, or psychosis), no alcohol or substance abuse within the last two years, and no participation in intervention studies. The inclusion periods were January 2022–May 2023. Patients included in the study had the following diagnoses which were diagnosed using the relevant diagnostic criteria: MCI or dementia due to AD [17,18], dementia with Lewy bodies [19], frontotemporal dementia [20], cerebral amyloid angiopathy [21], and vascular dementia [22]. The criteria for syndromes included mild cognitive impairment (MCI), using the International Working Group criteria [23], or the National Institute of Aging-Alzheimer's Association (NIA-AA) criteria [17] and the NIA-AA dementia criteria [24]. We did not estimate an a priori sample size due to the explorative nature of the study. The study was registered at clinicaltrials.gov (NCT05175664). Permission from the regional ethics committee was acquired before the commencement of the study (H-21044863). We followed the tenets of the 1975 Helsinki Declaration and written informed consent was obtained from all patients.

3. Procedure

3.1. Quantitative light reflex pupillometry

Patients were recruited on the day of lumbar puncture scheduled as part of diagnostic work-up due to suspicion of neurodegenerative disease. After a successful lumbar puncture and subsequently 30 min of rest, the light reflex was measured at this visit (referred to as baseline) in an examination room without disturbances (at hours 09:00–15:30). This procedure was repeated on day 3–14 after lumbar puncture (visit 2) and again at day 21–35 (visit 3) under similar conditions (see study flow chart, Fig. 1), but not always by the same examiner. The examiner was not blinded to previous measurements.

Ophthalmoscopy was performed before the qLRP procedure to check for the absence of obscurity of the lens and vitreous body. Approximately 5 min of adaptation to ambient light was provided afterward. Before the procedure, ambient light intensity was measured using the Light Meter App for iPhone. Direct sunlight was avoided and room lighting was calibrated to normal ambient light conditions at 300–1500 lx [25] simulating real-world conditions.

We used the PLR-3000 (NeurOptics®) Pupillometer system to measure the light reflex. The PLR-3000 Pupillometer is a research-grade, hand-held unit that can wirelessly transfer measurements and raw data in Excel format. It uses a proprietary algorithm to calculate various pupillometric parameters (see Fig. 2 for an explanation). The settings for the pupillometer were: Positive Pulse Stimulus, 180 μW light intensity, 0.84-s stimulus duration, stimulus onset after 1 s, measurement duration of 10 s, and no background light. The pupillometer tracks the pupillary size using an infrared camera recording at a 30 Hz framerate and it is calibrated by the manufacturer. The device was used according to the manufacturer's instructions. In short, the patients were seated and instructed to stare at a fixed point on a wall behind the examiner. If patients had difficulty keeping the examined eye open, the examiner kept the eye open manually during the recording. The right eye was measured first and a total of three acceptable measurements (without excessive blinking or artifacts) were carried out before measuring the contralateral eye in the same manner.

3.2. Pre-processing of data

The summarized and raw pupillometric data were transferred in Excel format to a computer and imported into *R Studio*, where low-quality recordings (flagged by the pupillometer) were visually inspected using the *clintools* package function *dilations*, which was adapted for this purpose. If blinks were present in the pre-stimulus and contraction phases, the recording was discarded. The pupillometer does not calculate dilation velocity and T75 when blinking occurs during the recording, and as such some patients in which blinking could not be avoided had measurements missing for these variables ($N = 2$ observations). The T75 parameter is dependent upon the pupil returning to 75% of the baseline value, and in cases where this did not occur the parameter cannot be calculated ($N = 3$ observations). We did not interpolate pupillary tracings for blinks or artifacts and did not calculate derivatives of qLRP parameters, and thus only pupillometric parameters provided by the PLR-3000 Pupillometers proprietary algorithm were used for analysis: baseline pre-stimulus pupil diameter (mm), peak constriction pupil diameter (mm), delta change between baseline and peak pupil diameter (%), latency (m*sec*), average constriction velocity (mm/s), maximum constriction velocity (mm/s), average dilation velocity (mm/s), and time to reach 75% of the baseline value after peak constriction (T75) (sec). A mean of up to three acceptable measurements were used for analysis. The right eye measurements were used for analysis as a significant effect of the number of trials was present and some qLRP parameters were shown to vary between eyes (see Supplementary material). In cases where significant monocular disease was present (one patient had congenital unilateral blindness and one patient

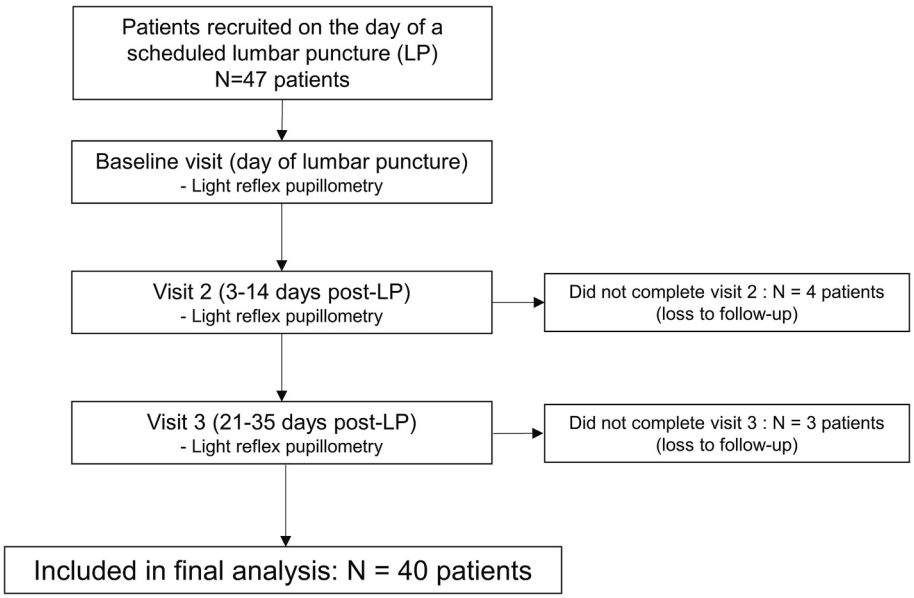


Fig. 1. Study flow chart.

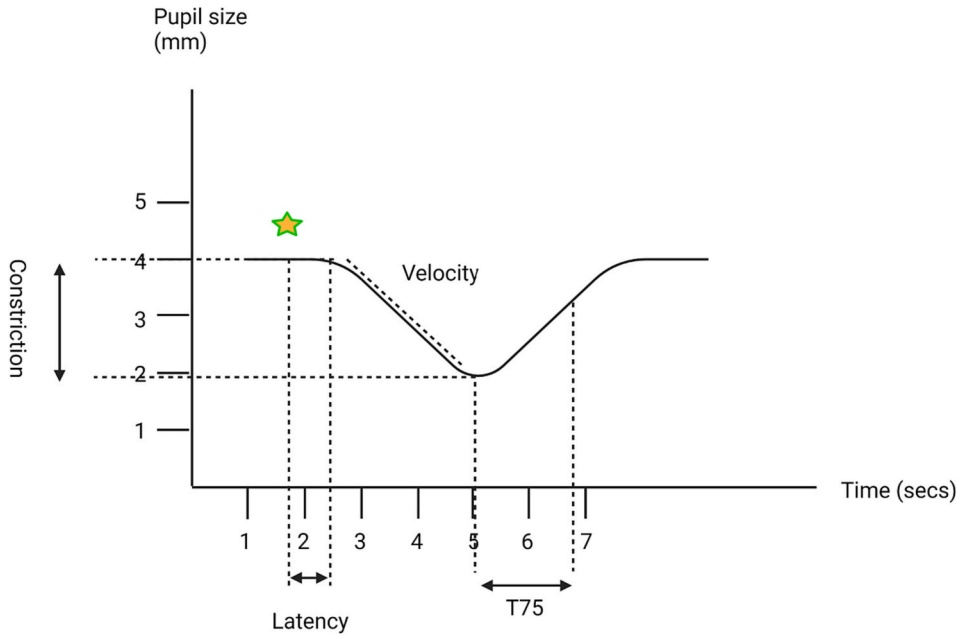


Fig. 2. A plot of the pupillary constriction as a function of time. The various metrics that are extracted by the PLR-3000 pupillometer are shown. The star marks the beginning of the light stimulus.

had unilateral central vein thrombosis), the left eye measurements were used.

3.3. Cognitive testing and demographic data

The Mini-Mental State Examination [26] assesses global cognitive function and was administered by a trained nurse at an initial evaluation visit to the memory clinic. A history of ophthalmological conditions, subjective visual acuity, medications, and eye surgery was recorded. Visual acuity was not measured objectively. Demographic data, including age, sex, educational level, information on medication use, and eye surgery was extracted from medical files and entered into an electronic database (REDCap, Vancouver).

3.4. Cerebrospinal fluid

We measured total-tau in the cerebrospinal fluid as a marker of neurodegeneration to investigate whether elevated levels were associated with poorer test-retest reliability of qLRP. CSF was obtained by lumbar puncture, which was performed due to suspicion of a neurodegenerative disorder and CSF was handled according to standard operating procedures. Total tau was measured in cerebrospinal fluid (CSF) using a commercially available enzyme-linked immunosorbent assay (ELISA) (Innotest, Fujirebio, Ghent) as part of routine diagnostic work-up at the hospital lab. The assay used by our hospital laboratory changed during the study period (after 1/7–2022) to a sandwich electrochemiluminescence-immunoassay (Roche, Cobas 8000). We used the reference cutoffs of our hospital lab before and after the

measurement method changed (before 1/7–2022: 400 ng/L; after 1/7–2022: 250 ng/L) to dichotomize groups with normal and high total tau.

3.5. Statistical analysis

We present summary statistics of variables according to their distribution (normal or non-normal) and present them as means with corresponding standard deviations and medians and ranges where appropriate. We calculated Spearman's *rho* for correlations between all pupillometric and demographic variables and constructed a correlogram using the function *corrplot* in the *ggcorrplot* package in R. We fitted linear mixed models (LMMs) using the *lme4* package and the function *lme*, assuming an unstructured covariance pattern (only one missing measurement). To test for the effect of the number of trials, patients were included as a random effect and trial as a fixed effect and separate univariate models including each pupillometric variable measured at the baseline visit for up to three acceptable trials as the outcome. We extracted fixed effect estimates from the models and calculated

95% confidence intervals (95% CIs) using the *confint* function. To investigate the inter-visit, intra-subject variability, we fitted LMMs with patients and visits as random effects and visits as a fixed effect. The residuals for the random effects from separate LMMs including each pupillometric parameter were used to derive a proposed clinically significant change representing the standard deviation of the residuals, reflecting the intra-subject, short-term variability. Further, these same LMMs were fitted with adjustment for baseline pupil diameter included as a log-transformed fixed effect, as this measure is tightly correlated with the other variables, and had been shown to exhibit a logarithmic relationship with constriction velocity [7,27,28]. Model assumptions were not checked for individual LMMs, as it was assumed to be robust to the non-normality of variables due to few missing data [29]. To investigate the test-retest reliability we calculated estimates of intraclass correlation coefficients (ICCs) from a two-way agreement model (McGraw and Wong convention ICC (A,3)) [30] considering complete cases only (acceptable measurements for all three visits) for all pupillometric parameters with corresponding 95% CIs, using the *irr* package and the function *icc* for average measurements, as we used the mean of three measurements at each visit. For ICCs calculated based on single measurements the ICC (A,1) model was used. To investigate for possible sources of variability, ICCs were also calculated for subgroups dichotomized according to the median of age ($>$ and $<$ 72 years), sex (male vs. female), disease severity (MCI vs. mild dementia), total tau concentration in CSF (above 400 ng/L before 1/7–2022 and $>$ 250 ng/L after this date), diagnosis of AD (vs. non-AD), and antidepressant medication use (yes vs. no) and the 95% CIs for the ICC estimate were compared, with an overlap indicating no difference between groups. The ICC estimates were evaluated according to the intervals provided by Koo & Li [4]. We fitted univariate linear regressions for the relationship between ambient light intensity measured in lux and all pupillometric parameters measured at baseline separately. Linear regression model assumptions were checked by inspection of Q-Q and scale-location plots. If assumptions were violated, transformation of the data was tried and subsequently a Spearman's rank correlation coefficient was calculated instead if no non-linear transformation fitted the data better.

R (ver. 4.2.2) was used for all analyses [31]. A p-value<0.05 was considered significant and only two-tailed tests were used. In the supplementary material, a full read-out of the results from *R* with the source code used to produce the output is provided.

4. Results

A total of 47 patients were recruited of which 40 completed all visits, comprising the analyzed cohort (see Fig. 1 for a study flow chart).

The mean age of patients was 71.6 years (range 60.8–81.7) and there was a slight overrepresentation of males (62.5%) (see Table 1 for baseline characteristics).

Table 1
Study cohort characteristics.

	Overall ($N = 40$)
Age (years)	
Mean (SD)	71.6 (5.9)
Range	60.8–81.7
Sex	
Female	15 (37.5%)
Male	25 (62.5%)
MMSE total score	
Median	27
Range	21–30
Educational level (years)	
Mean (SD)	11.5 (3.3)
Range	6–17
Antidepressant medicine	
No	34 (85.0%)
Yes	6 (15.0%)
Disease severity	
MCI	22 (55.0%)
Mild dementia	18 (45.0%)
CSF-Total tau	
Elevated total tau	19 (47.5%)
Normal total tau	21 (52.5%)
Etiological diagnosis	
Affective	1 (2.5%)
Alzheimer's disease	19 (47.5%)
Alzheimer's disease (Posterior cortical variant)	2 (5.0%)
Dementia with Lewy bodies	4 (10.0%)
Frontotemporal dementia	1 (2.5%)
Frontotemporal dementia - Progressive non-fluent aphasia	1 (2.5%)
Traumatic brain injury	1 (2.5%)
Vascular + motor neuron disease	1 (2.5%)
Vascular dementia (Cerebral amyloid angiopathy)	1 (2.5%)
Neurodegenerative – not further classifiable	9 (22.5%)

The median MMSE score was 27. The patients had a wide range of diagnoses, and >50 % of the subjects were diagnosed with AD. Most baseline pupillometric variables, except latency, were positively correlated with moderate-strong correlations (see Fig. 2 and Fig. 3), while educational level was weakly positively correlated with average constriction velocity, although this correlation was not found when we examined the full dataset (i.e., with all available measurements, see Supplementary Figs. 1 and 2). No further correlations were found between pupillometric parameters and demographic variables.

4.1. Test-retest reliability and short-term variability of qLRP

We examined the test-retest reliability of measurements across visits by calculating ICCs based on the means of measurements from each visit. These results are shown in Table 2.

We generally found good-excellent reliability of measurements (ICC range 0.86–93) for most pupillometric variables, although T75 showed moderate reliability (ICC 0.56). ICCs based on the first measurements from each visit showed moderate-good reliability (see Supplementary Table 2). We further explored the inter-visit variability by fitting LMMs for all variables with visits and patients as random effects. The residuals of these models indicate the amount of variation that could occur randomly, and we have defined a clinically significant change for each variable based on the standard deviation of these residuals reflecting the intra-subject, short-term variability (Table 2). This value can be used to indicate whether an individual change (either due to pharmacological influence or neurodegenerative disease processes) could be ascribed to the actual intervention or disease process and not the intra-subject, short-term variability. In Table 3 we also present mean values for all pupillometric measurements at each visit as well as spaghetti plots of all measurements for all patients in a subset of pupillometric variables in Fig. 4 (see Supplementary Fig. 3 for the spaghetti plots of the remaining variables).

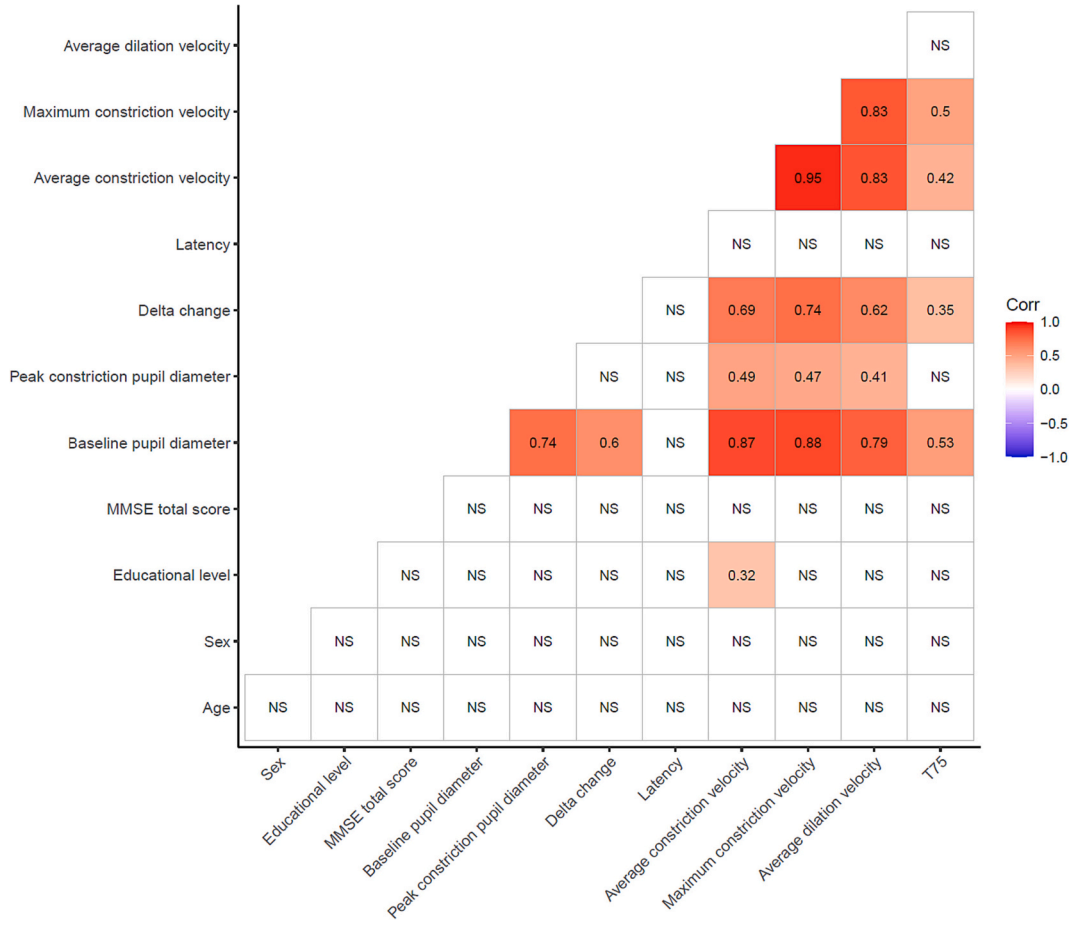


Fig. 3. A correlogram showing correlations between baseline pupillary measurements (mean of the first three acceptable measurements at the baseline visit) and demographic variables. Spearman's *rho* is given, and only significant correlations are shown (NS = not significant). For the pupillometric measurements, only complete cases are considered (conditioned on having at least one acceptable T75 measurement) and one patient was excluded due to not having a baseline T75 value ($N = 39$ patients).

Table 2
Intraclass correlation coefficients and proposed clinically significant change.

qLRP parameter	Intraclass correlation coefficient [95% CI]	Clinically significant change
Baseline pupil diameter	0.889 [0.81–0.938][a]	0.25 mm
Peak constriction pupil diameter	0.928 [0.87–0.961][a]	0.11 mm
Delta change	0.933 [0.885–0.962][a]	2.9%
Latency	0.86 [0.762–0.922][a]	0.015 s
Average constriction velocity	0.901 [0.831–0.945][a]	0.24 mm/s
Maximum constriction velocity	0.896 [0.823–0.942][a]	0.38 mm/s
Average dilation velocity	0.892 [0.816–0.94][b]	0.10 mm/s
T75	0.557 [0.235–0.758][c]	0.5 s

T75 = Time to reach 75% of baseline (secs).
Note: The proposed clinically significant change is derived from residuals of linear mixed-effects models. Patients and visits are included as random effects. The ICCs are calculated from the means of the first three acceptable qLRP measurements on complete cases (acceptable qLRP measurements at all three visits).
[a] N = 39 patients.
[b] N = 38 patients.
[c] N = 37 patients.

As can be seen from the spaghetti plots in Fig. 4, some individual observations for T75 were far from the prior or subsequent observations.

4.2. Possible sources of variability

To further explore whether the reliability of qLRP measurements could be influenced by test-day conditions such as ambient lighting and the number of trials, we investigated the association between these variables and qLRP parameters. We found a significant association between the number of trials and baseline pupil diameter, peak constriction diameter, latency, and average constriction velocity, although the effects were small (see Supplementary Table 3 and Supplementary Figs. 4 and 5). We found an association between ambient light and T75 ($P = 0.01$), and a trending association between ambient light intensity and relative change in pupillary size ($P = 0.06$) and dilation velocity ($P = 0.07$), although assumptions for the linear regression model for T75 were violated due to non-linearity and heteroscedasticity (see Supplementary Table 4). Log transformation of ambient light and T75 did not improve the model fit (data not shown). A statistically significant negative correlation (Spearman's *rho* $= -0.36$, $p = 0.03$) between ambient light and T75 was found.

The results of models with adjustment for baseline pupil diameter are shown in Supplementary Table 5. The adjustment for baseline pupil diameter reduced the residuals of some of the models, indicating that the parameter could explain part of the inter-visit variability, mainly for the qLRP parameters peak constriction velocity, delta, and constriction velocity.

Table 3
Means of the first three acceptable qLRP measurements at the three study visits.

	Baseline (N = 40)	Visit 2 (N = 40)	Visit 3 (N = 39)
Baseline pupil diameter (mm)			
Mean (SD)	3.05 (0.49)	3.20 (0.49)	3.18 (0.48)
Range	2.1–4	2.1–4.07	2.23–4.43
Peak constriction pupil diameter (mm)			
Mean (SD)	2.07 (0.27)	2.15 (0.28)	2.15 (0.26)
Range	1.47–2.8	1.6–2.83	1.73–2.77
Delta change (percent)			
Mean (SD)	31.5 (6.4)	32.1 (7.1)	31.6 (6.7)
Range	17.7–44.3	17.3–46	18.3–43
Latency (secs)			
Mean (SD)	0.24 (0.03)	0.25 (0.03)	0.25 (0.03)
Range	0.197–0.297	0.210–0.290	0.200–0.300
Average constriction velocity (mm/s)			
Mean (SD)	1.86 (0.5)	1.88 (0.5)	1.88 (0.5)
Range	0.71–2.73	0.72–2.68	0.8–2.73
Maximum constriction velocity (mm/s)			
Mean (SD)	2.85 (0.79)	2.91 (0.75)	2.89 (0.72)
Range	1.21–4.49	1.14–4.04	1.34–4.48
Average dilation velocity (mm/s)			
N missing	1	0	1
Mean (SD)	0.90 (0.25)	0.94 (0.23)	0.94 (0.22)
Range	0.31–1.38	0.5–1.35	0.39–1.32
Time to reach 75% of baseline (secs)			
N missing	1	1	1
Mean (SD)	1.64 (0.5)	1.74 (0.82)	1.8 (0.65)
Range	0.79–2.58	0.93–4.44	0.75–3.83

4.3. Sub-group analyses

We investigated whether certain groups were more prone to exhibit poorer test-retest reliability for qLRP measurements, possibly reflecting higher intra-subject, short-term variability. We investigated sex (male vs. female), disease severity (MCI vs. mild dementia), age groups (< and > 72 years), normal vs. high CSF-total tau, AD vs. non-AD, as well as antidepressant medicine use (yes vs. no). Neither of these subgroups had significantly different ICC values for any qLRP parameter (see Supplementary Material).

4.4. Feasibility and data quality

The procedure was generally well tolerated with no adverse events. One patient could not cooperate due to discomfort associated with qLRP. For one patient, no acceptable measurements could be obtained at the last visit due to excessive blinking.

5. Discussion

In this test-retest study, we investigated the reliability and intra-subject, short-term variability of quantitative light reflex pupillometry in a memory clinic setting. We generally found good-excellent reliability and limited intra-subject, short-term variability for most qLRP parameters. We found that ambient light possibly influenced T75, but this could not fully explain the poor test-retest reliability of this particular qLRP parameter. Also, a small effect of the number of trials was found for some pupillometric variables, which could account for some intra-subject variability. Otherwise, we found no inherent sources of intra-subject, short-term variability pertaining to certain sub-groups of patients with regard to sex, age, disease severity, or antidepressant medication use.

qLRP has previously shown a high degree of reliability in both healthy individuals [7] and critically ill patients [13]. However, this study is, to our knowledge, the first to investigate the test-retest reliability and intra-subject, short-term variability of qLRP in a memory clinic setting. Our results corroborate these findings and extend the generalizability to include cohorts of patients with cognitive dysfunction and varying degrees of neurodegeneration.

We found that patients generally tolerated the procedure well with no adverse events registered and only one patient could not cooperate due to discomfort. The data loss was limited to a single visit for one patient for a few qLRP parameters (dilation velocity and T75). This indicates that the feasibility of applying this procedure in a memory clinic is high, although the study was not designed to examine this aspect.

We did not identify serious sources of patient or condition-related variability, such as sex, disease severity, and ambient light, although our sample size may not have been large enough to identify small sources of variability. We found that ambient light intensity did influence T75, a variable that is dependent upon the speed of the pupil to return to its baseline diameter. Also, the number of trials had a small effect on some pupillary parameters. This could be mitigated by controlling light conditions to keep a normal ambient light interval (300–1500 lx) [25], limiting the number of trials, and using proper measurement technique (keeping the eyelids open manually to avoid excessive blinking). Another solution to controlling light conditions is to perform measurements in complete darkness, although this may be uncomfortable and requires time for adaptation that may not be suited for everyday clinical use. We find that three trials seem to produce reliable serial measurements without affecting the qLRP parameters to a large degree. Although we did not identify significant differences between eyes, we propose that qLRP measurements should be performed on the same eye for serial evaluation.

The disease severity of patients (MCI vs. mild dementia) did not seem to influence the test-retest reliability of qLRP. This means that longitudinal qLRP could perform equally well in both early and later-stage disease, although we did not include patients with an MMSE total score lower than 20. Thus, we cannot generalize findings to patients with moderate-severe dementia for which the neurodegenerative processes might be hypothesized to influence qLRP parameters more severely. Surprisingly, we did not find an increased variability when dichotomizing groups according to CSF-total tau levels, which indicates that the presence of active neurodegenerative processes does not influence the variability of qLRP.

We found a high degree of correlation between qLRP variables. The baseline pupil diameter was correlated with all but one qLRP parameter (latency). This has previously been found [7,15] and it has therefore been proposed to only use relative measures such as relative change in pupillary size. Further, the observed correlation also led us to investigate whether baseline pupil diameter could explain some of the inter-visit variability. Indeed, adjusting for baseline pupil diameter reduced residuals of the models for peak pupil diameter, delta, and constriction velocity. This could indicate that baseline pupil diameter, as a marker of arousal and/or affective state, has a general effect on these light reflex properties [32–34]. Another explanation could be that all these parameters are correlated to a higher-order affective state, which is an aspect we could not investigate since we did not have data relating to the affective state of the patients.

We only investigated metrics that were provided by the manufacturer to increase the reproducibility of our study and to resemble a clinical setting. For example, we did not report on qLRP parameters such as the maximum pupillary constriction acceleration, which has previously been found to be altered in patients with AD [1], and we cannot conclude on the test-retest reliability of this metric.

An interesting possible application of qLRP is disease tracking in Alzheimer's disease, as the light reflex has been found to be altered in patients with AD also at a pre-symptomatic stage [11,35]. We found that patients that qLRP could be reliably measured equally well in patients

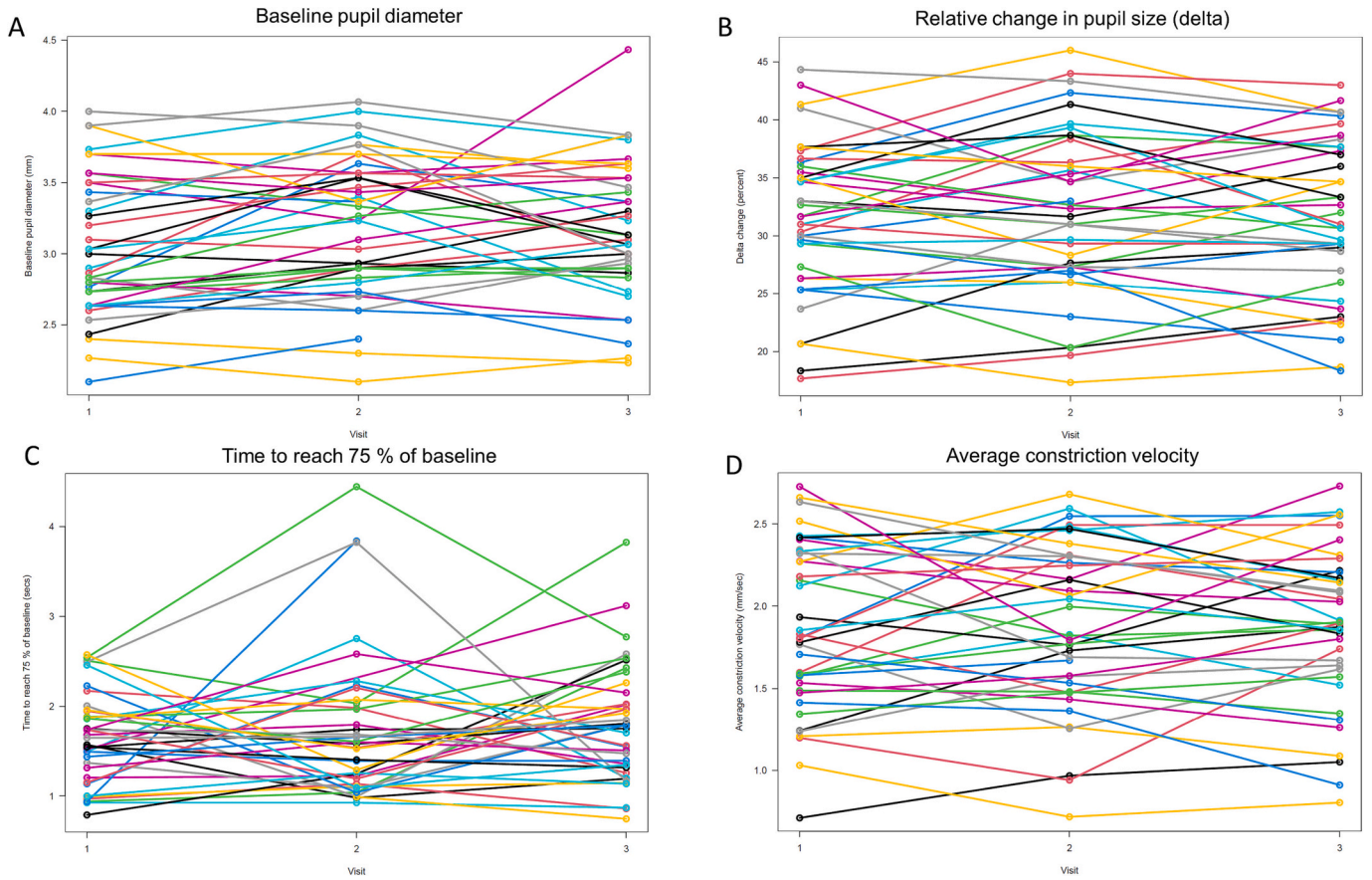


Fig. 4. Spaghetti plots of pupillometric measurements at each visit (Panel A: Baseline pupil diameter, B: Delta change, C: T75, D: Average constriction velocity).

with an AD and non-AD diagnosis, and our study, therefore, supports this proposed application.

Previous studies have focused on cross-sectional measurements in specific disease groups, such as patients with AD and patients with Parkinson's disease with and without cognitive impairment [1], which are disorders frequently encountered in memory clinics. While some groups have shown excellent separation of disease vs. healthy groups [11], others have failed to replicate these results [36,37]. It has been proposed recently that longitudinal measurements may better capture neurodegenerative processes in these diseases [2,3], the further exploration of which the present study supports.

We did not find a statistically significant correlation between MMSE score and qLRP variables, which has previously been shown in patients with AD [38]. Also, age would be expected to correlate with qLRP parameters, which we could not show. There are several explanations for this. First, the study was not powered to this aspect, as we were mainly interested in the test-retest reliability and not cross-sectional associations. Second, our cohort consisted of a mixture of dementia etiologies. This again limited power in detecting significant correlations with MMSE specifically. Third, the MMSE total scores were quite restricted and MMSE is generally not considered a sensitive test of cognitive decline. Finally, qLRP may not be marker of disease severity but perhaps a trait marker of disease, which our study was not designed to investigate. As such, qLRP still presents a promising approach, and a study in a large cross-sectional cohort conducted by our group has shown correlations between MMSE, age, and qLRP parameters (Gramkow et al., unpublished). Efforts should be made in future studies to evaluate whether qLRP could be used according to a cut point, which effectively would aid the clinician in determining the presence of a disease such as Alzheimer's disease, in which qLRP has shown the most promising results [1,11,38]. The present study had a purposefully limited follow-up period, and future studies should evaluate a longitudinal design allowing for longer individual trajectories with repeat measurements of qLRP.

The strengths of this study are the prospective design, the relatively large sample size, and a cohort that represents the heterogeneous profile of memory clinic patients. We also performed measurements in out-of-lab clinical settings reflecting real-world conditions, where qLRP is expected to be applied. As such, external factors that could influence measurements (e.g., scheduled lumbar puncture) were allowed in the research design. This likely increases the external validity of our results.

We are also aware of certain limitations pertaining to our study. We had a small, but noteworthy drop-out rate and the application of complete case analysis may limit the generalizability of our results on the test-retest reliability. Also, the cohort had relatively high MMSE total scores, and we did not include patients with moderate-severe dementia. It is possible that these more advanced disease states could lead to higher intra-subject variability which we could not investigate in this study. To add to this, we only included patients who had lumbar puncture performed, which might have induced selection bias.

In conclusion, most pupillometric parameters, except T75, showed good-excellent test-retest reliability and limited intra-subject, short-term variability. We propose that a mean of three measurements is used for analysis as this method gives the smallest amount of variability. Pupillometry could be reliably used regardless of the age and sex of subjects, the amount of neurodegeneration present, a diagnosis of AD, and the use of antidepressant medication. Considering our results, we find that qLRP is a reliable tool that can be used in patients undergoing evaluation in a memory clinic. This work paves the way for future studies that could evaluate changes in qLRP related to neurodegenerative processes or pharmacological interventions.

CRediT authorship contribution statement

Mathias Holsey Gramkow: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Frederikke Kragh Clemmensen:** Conceptualization, Data curation, Project administration, Writing – review & editing, Validation, Funding acquisition. **Gunhild Waldemar:** Resources, Supervision, Validation, Writing – review & editing, Funding acquisition. **Steen Gregers Hasselbalch:** Conceptualization, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Kristian Steen Frederiksen:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors have no conflict of interest to report.

Acknowledgments

Funding for this project was received from Frimodt-Heineke Fonden. The salary of Mathias Holsey Gramkow was supported by Michaelsen Fonden and the Danish Ministry of Health. The funders had no role in the design or reporting of the study. We would like to thank Dr. Marwan Othman for invaluable advice on the practical and analytical aspects of quantitative pupillometry.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jns.2023.122856.

References

[1] P.S. Chougule, R.P. Najjar, M.T. Finkelstein, N. Kandiah, D. Milea, Light-induced pupillary responses in Alzheimer's disease, Front. Neurol. 10 (APR) (2019) 1–12.
[2] V. Troiani, The future of quantitative pupillometry in health and disease, Clin. Auton. Res. 30 (1) (2020) 11–12. Available from: https://doi.org/10.1007/s10286-019-00655-3.
[3] G.P. Van Stavern, Current opinion neurology: visual pathway biomarkers in Alzheimer's disease, Curr. Opin. Neurol. 33 (1) (2020) 79–86.
[4] T.K. Koo, M.Y. Li, A guideline of selecting and reporting Intraclass correlation coefficients for reliability research, J. Chiropr. Med. 15 (2) (2016) 155–163. Available from: https://doi.org/10.1016/j.jcm.2016.02.012.
[5] N. Balakrishnan, Methods and Applications of Statistics in Clinical Trials. Methods and Applications of Statistics in Clinical Trials. Somerset, Wiley, 2014, pp. 1–961 (p. (Wiley Series Methods and Applications of Statistics; vol. 1)).
[6] C.H. van Dyck, C.J. Swanson, P. Aisen, R.J. Bateman, C. Chen, M. Gee, et al., Lecanemab in early Alzheimer's disease, N. Engl. J. Med. 388 (1) (2023) 9–21. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2212948.
[7] W.R. Taylor, J.W. Chen, H. Meltzer, T.A. Gennarelli, C. Kelbch, S. Knowlton, et al., Quantitative pupillometry, a new technology: normative data and preliminary observations in patients with acute head injury - technical note, J. Neurosurg. 98 (1 SUPPL) (2003) 205–213. Available from: https://thejns.org/view/journals/j-neurosurg/98/1/article-p205.xml.
[8] L.F.M. Scinto, M. Frosch, C.K. Wu, K.R. Daffner, N. Gedi, C. Geula, Selective cell loss in Edinger-Westphal in asymptomatic elders and Alzheimer's patients, Neurobiol. Aging 22 (5) (2001) 729–736. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0197458001002354.
[9] L.F.M. Scinto, C.K. Wu, K.M. Firla, K.R. Daffner, D. Saroff, C. Geula, Focal pathology in the Edinger-Westphal nucleus explains pupillary hypersensitivity in Alzheimer's disease, Acta Neuropathol. 97 (6) (1999) 557–564.
[10] I.A. Mavroudis, M.G. Manani, F. Petrides, C. Petsoglou, S.N. Njau, V.G. Costa, et al., Dendritic and spinal alterations of neurons from Edinger-Westphal nucleus in Alzheimer's disease, Folia Neuropathol. 52 (2) (2014) 197–204.
[11] D.F. Fotiou, V. Stergiou, D. Tsiptsios, C. Lithari, M. Nakou, A. Karlovasitou, Cholinergic deficiency in Alzheimer's and Parkinson's disease: evaluation with pupillometry, Int. J. Psychophysiol. 73 (2) (2009) 143–149. Available from: https://doi.org/10.1016/j.ijpsycho.2009.01.011.
[12] D.M. Bittner, I. Wieseler, H. Wilhelm, M.W. Riepe, N.G. Müller, Repetitive pupil light reflex: potential marker in Alzheimer's disease? J. Alzheimers Dis. 42 (4) (2014) 1469–1477. Available from: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JAD-140969.
[13] B. Nyholm, L. Obling, C. Hassager, J. Grand, J. Møller, M. Othman, et al., Superior reproducibility and repeatability in automated quantitative pupillometry compared to standard manual assessment, and quantitative pupillary response parameters present high reliability in critically ill cardiac patients, PLoS One 17 (7 July) (2022) 1–15.
[14] K.A. Jellinger, Basic mechanisms of neurodegeneration: a critical update, J. Cell. Mol. Med. 14 (3) (2010) 457–487.
[15] C. Kelbsch, T. Strasser, Y. Chen, B. Feigl, P.D. Gamlin, R. Kardon, et al., Standards in pupillography, Front. Neurol. 10 (February) (2019).
[16] C. Martinez, R.W. Jones, S. Rietbrock, Trends in the prevalence of antipsychotic drug use among patients with Alzheimer's disease and other dementias including those treated with antidementia drugs in the community in the UK: a cohort study, BMJ Open 3 (1) (2013).
[17] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, et al., The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimers Dement. 7 (3) (2011) 270–279. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3312027&tool=pmcentrez&rendertype=abstract.
[18] G.M. McKhann, D.S. Knopman, H. Chertkow, B.T. Hyman, C.R. Jack Jr., C. H. Kawas, et al., The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimers Dement. 7 (3) (2011) 263–269.
[19] I.G. McKeith, B.F. Boeve, D.W. Dickson, G. Halliday, J.-P. Taylor, D. Weintraub, et al., Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB consortium, Neurology 89 (1) (2017) 88–100.
[20] M.L. Gorno-Tempini, A.E. Hillis, S. Weintraub, A. Kertesz, M. Mendez, S.F. Cappa, et al., Classification of primary progressive aphasia and its variants, Neurology. 76 (11) (2011) 1006–1014.
[21] S.M. Greenberg, A. Charidimou, Diagnosis of cerebral amyloid angiopathy evolution of the Boston criteria, Stroke. 49 (2) (2018) 491–497.
[22] P. Sachdev, R. Kalaria, J. O'Brien, I. Skoog, S. Alladi, S.E. Black, et al., Diagnostic criteria for vascular cognitive disorders: A VASCOG statement, Alzheimer Dis. Assoc. Disord. 28 (3) (2014) 206–218.
[23] B. Winblad, K. Palmer, M. Kivipelto, V. Jelic, L. Fratiglioni, L.O. Wahlund, et al., Mild cognitive impairment - beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment, J. Intern. Med. 256 (3) (2004) 240–246.
[24] G.M. McKhann, D.S. Knopman, H. Chertkow, B.T. Hyman, C.R. Jack, C.H. Kawas, et al., The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging and the Alzheimer's Association workgroup, Alzheimers Dement. 7 (3) (2011) 263–269. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21514250.
[25] L. Dębska, A. Białek, Lighting conditions as the occupational health related issue – case study, MATEC Web Conf. 354 (2022) 00059.
[26] M.F. Folstein, S.E. Folstein, P.R. McHugh, "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician, J. Psychiatr. Res. 12 (3) (1975) 189–198.
[27] D.F. Fotiou, C.G. Brozou, A.B. Haidich, D. Tsiptsios, M. Nakou, A. Kabitsi, et al., Pupil reaction to light in Alzheimer's disease: evaluation of pupil size changes and mobility, Aging Clin. Exp. Res. 19 (5) (2007) 364–371.
[28] R. Prettyman, P. Bitsios, E. Szabadi, Altered pupillary size and darkness and light reflexes in Alzheimer's disease, J. Neurol. Neurosurg. Psychiatry 62 (6) (1997) 665–668 (Available from: /pmc/articles/PMC1074161/?report=abstract).
[29] H. Schielzeth, N.J. Dingemanse, S. Nakagawa, D.F. Westneat, H. Allegue, C. Teplitsky, et al., Robustness of linear mixed-effects models to violations of distributional assumptions, Methods Ecol. Evol. 11 (9) (2020) 1141–1152.
[30] K.O. McGraw, S.P. Wong, Forming inferences about some Intraclass correlation coefficients, Psychol. Methods 1 (1) (1996) 30–46.
[31] R Core Team, R, A Language and Environment for Statistical Computing, Available from: https://www.r-project.org/, 2016.
[32] C. Strauch, C. Wang, W. Einhäuser, S. Van der Stigchel, M. Naber, Pupillometry as an integrated readout of distinct attentional networks, Trends Neurosci. 45 (8) (2022) 635–647. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0166223622000972.
[33] S.R. Steinhauer, G. Hakerem, The pupillary response in cognitive psychophysiology and schizophrenia, Ann. N. Y. Acad. Sci. 658 (1) (1992) 182–204. Available from, https://doi.org/10.1111/j.1749-6632.1992.tb22845.x.
[34] P. Raiturkar, A. Kleinsmith, A. Keil, A. Banerjee, E. Jain, Decoupling light reflex from pupillary dilation to measure emotional arousal in videos, in: Proceedings of the ACM Symposium on Applied Perception, Association for Computing Machinery, New York, NY, USA, 2016, pp. 89–96 (SAP '16). Available from: https://doi.org/10.1145/2931002.2931009.
[35] S. Frost, L. Robinson, C.C. Rowe, D. Ames, C.L. Masters, K. Taddei, et al., Evaluation of cholinergic deficiency in preclinical Alzheimer's disease using Pupillometry, J. Ophthalmol. (2017), 7935406 ([cited 2021 Oct 13];2017. Available from: /pmc/articles/PMC5574262/).
[36] G.P. Van Stavern, L. Bei, Y.-B.B. Shui, J. Huecker, M. Gordon, Pupillary light reaction in preclinical Alzheimer's disease subjects compared with normal ageing

controls, Br. J. Ophthalmol. 103 (7) (2019) 971–975. Available from: http://bjo.bmj.com/.

[37] A. Kawasaki, S. Ouanes, S.V. Crippa, J. Popp, Early-stage Alzheimer's disease does not alter pupil responses to colored light stimuli, Lee C, editor, J. Alzheimers Dis. 75 (4) (2020) 1191–1201. Available from: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JAD-200120.

[38] D. Fotiou, A. Kaltsatou, D. Tsiptsios, M. Nakou, Evaluation of the cholinergic hypothesis in Alzheimer's disease with neuropsychological methods, Aging Clin. Exp. Res. 27 (5) (2015) 727–733.

STUDY II

Diagnostic performance of light reflex pupillometry in Alzheimer's disease

Gramkow, M. H., Clemmensen, F. K., Sjælland, N. S., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S. (2024). Diagnostic performance of light reflex pupillometry in Alzheimer's disease. Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring, 16(3), 1–11.

DOI: https://doi.org/10.1002/dad2.12628

The supplementary material related to this journal article can be accessed at:

https://doi.org/10.1002/dad2.12628

or by scanning the QR-code below

Received: 12 March 2024 | Revised: 4 July 2024 | Accepted: 5 July 2024

DOI: 10.1002/dad2.12628

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Diagnostic performance of light reflex pupillometry in Alzheimer's disease

Mathias Holsey Gramkow[1] | Frederikke Kragh Clemmensen[1] | Nikolai Sulkjær Sjælland[1] | Gunhild Waldemar[1,2] | Steen Gregers Hasselbalch[1,2] | Kristian Steen Frederiksen[1,2]

[1]Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark

[2]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Correspondence
Mathias Holsey Gramkow, Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital – Rigshospitalet, Inge Lehmanns Vej 8, Copenhagen, DK-2100, Denmark.
Email: mathias.holsey.gramkow@regionh.dk

Funding information
Frimodt-Heineke Fonden; Absalon Foundation; Rigshospitalets Forskningsfond; The Danish Alzheimer Society Research Foundation; Ellen Mørchs Fond; Overretssagfører L. Zeuthens Mindelegat

Abstract

Easily applied diagnostic tools such as digital biomarkers for Alzheimer's disease (AD) are urgently needed due to the recent approval of disease-modifying therapies. We aimed to determine the diagnostic performance of hand-held, quantitative light reflex pupillometry (qLRP) in patients with AD in a proof-of-concept, cross-sectional study. Participants underwent qLRP at a university memory clinic from August 2022 to October 2023. We fitted multivariable logistic regression models with qLRP, sex, and age as predictors evaluated with area under the receiver operating characteristics curve (AUROC). In total, 107 patients with AD, 44 patients with mixed AD and vascular cognitive dysfunction (VCD), 53 patients with dementia with Lewy bodies (DLB), and 50 healthy controls (HCs) were included. Our diagnostic models showed similar discriminatory ability (AUROC range 0.74-0.81) when distinguishing patients with AD from HCs and other dementias. The qLRP seems promising as a bedside digital biomarker to aid in diagnosing AD.

KEYWORDS
Alzheimer's disease, diagnosis, digital biomarker, light reflex, proof-of-concept, quantitative pupillometry

Highlights

- We demonstrated the diagnostic performance of qLRP in Alzheimer's disease.
- The diagnostic models were robust in sensitivity analyses.
- qLRP may assist in the bedside diagnostic evaluation of Alzheimer's disease.

1 | BACKGROUND

The prevalence of Alzheimer's disease (AD) is estimated to double every 20 years through 2050,[1] and recently, breakthrough disease-modifying therapies have been approved by the US Food and Drug Administration (FDA).[2,3] These events are likely to lead to an increased burden on diagnostic services. As a consequence, there is a need for low-cost, easily applied diagnostic assessment tools such as digital biomarkers for AD.[4,5]

Quantitative light reflex pupillometry (qLRP) can provide objective measures of the pupillary action in response to light stimuli. It can be carried out with a point-and-shoot, hand-held digital device, which

has shown excellent test-retest reliability in memory clinic cohorts, as recently demonstrated.[6] The light reflex is mainly subserved by the cholinergic preganglionic cell group of the Edinger-Westphal nucleus (EWn) which receives modulatory input from the locus coeruleus located inferior to the EWn in the mesencephalon.[7] The EWn is also connected with cortical areas, such as the frontal eye field, and the colliculus superior.[8] Cholinergic brain circuits are affected in AD as cholinergic neurons are particularly vulnerable to the pathological insults of AD, namely the build-up of $A\beta_{1\text{-}42}$-rich plaques and neurofibrillary tangles consisting of tau leading to neurodegeneration.[9,10] Beyond general cholinergic dysfunction, neuropathological studies have provided evidence of specific neuronal damage in the EWn in patients with AD.[11,12]

Several studies have shown a high discriminatory ability of qLRP in AD cohorts, but studies were mainly small with differing methods, and relevant diagnostic group comparators were not always included.[13–15]

Other approaches for exploring the pupillary response in AD have focused on the influence of cognitive load on pupil dilation to distinguish between healthy and AD giving credence to the hypothesis of utilizing the pupil as a biomarker in AD.[16]

We sought to develop a diagnostic application of hand-held qLRP in AD in a proof-of-concept setting in a large, representative cohort of memory clinic patients with high diagnostic certainty, including relevant comparators such as patients with mixed AD (+vascular pathology), vascular cognitive dysfunction (VCD), patients with dementia with Lewy bodies (DLB), and healthy controls (HCs). We hypothesized that qLRP could provide sensitive and specific diagnostic measures and could aid in distinguishing AD from both HCs and patients with other relevant pathologies, which could mimic AD in the clinic, namely patients with cerebrovascular disease and DLB.

RESEARCH IN CONTEXT

1. **Systematic review**: A systematic PubMed search was performed to identify studies investigating quantitative light reflex pupillometry (qLRP) in Alzheimer's disease (AD). Few studies included relevant differential diagnostic groups to evaluate the performance of qLRP in an envisioned use case, namely memory clinics.
2. **Interpretation**: We demonstrated the performance of hand-held quantitative light reflex pupillometry in the diagnosis of Alzheimer's disease in a real-world setting, and identified its usefulness in differentiating dementia disorders, thus proving the potential of the tool in the diagnostic assessment of AD.
3. **Future directions**: Future studies should aim at validating these findings in other settings, such as primary care, along with assessing its added value to known biomarkers for AD.

2 | METHODS

2.1 | Study design

Single-center, proof-of-concept, cross-sectional diagnostic study.

2.2 | Participants

We included patients seen for diagnostic work-up due to cognitive complaints in a tertiary memory clinic at a university hospital (Danish Dementia Research Centre at Rigshospitalet) in the period January 2022-October 2023 and HCs. Diagnostic assessment at the Danish Dementia Research Centre (DDRC) comprises examination by a doctor and nurse preferably with an informant/caregiver present. As a minimum patients undergo cognitive testing with two routine screening tests (Mini-Mental State Examination[17] and Addenbrooke's Cognitive Examination[18]), a structural scan (computed tomography or magnetic resonance imaging), and routine blood sampling. The center utilizes state-of-the-art neuroimaging techniques to ascertain various aspects of neurodegenerative pathology (e.g., 18-F-FDG-positron emission tomography [PET], dopamine agonist tracer scans, amyloid PET), cerebrospinal fluid biomarker analysis, genetic counseling, and neuropsychological testing if deemed relevant. The diagnosis is given following a multi-disciplinary (neurology, geriatrics, psychiatry, neuropsychology, nursing care) consensus conference.

Patients were included after diagnosis (with few exceptions, see below) if they fulfilled the following inclusion criteria (1) a diagnosis of AD (both mild cognitive impairment [MCI] and dementia stages),[19,20] AD with mixed pathology (cerebrovascular),[20] cognitive dysfunction due to cerebrovascular disease,[21] or MCI or DLB,[22,23] (2) An MMSE total score > 15, (3) able to provide informed consent, (4) able to cooperate to investigations with pupillometry, and none of the following exclusion criteria: (1) diagnosis of concurrent neurological (other than AD) or psychiatric disorder (mild depression allowed), (2) excessive alcohol intake > 5 units per day or substance use, (3) concurrent participation in interventional studies, (4) other known brain disorder which could cause cognitive dysfunction, (5) severe ophthalmological disorders (severe cataract, severe glaucoma, age-related macular degeneration, vitreous body bleed, vitreous body collapse, retinal bleed), which may interfere with pupillometry, as evaluated by the investigator. Mild eye disease was allowed, defined as glaucoma receiving monotherapy glaucoma medication, cataract treatment with uncomplicated cataract surgery, or similar uncomplicated disorders as evaluated by the primary investigator (M.G.) due to the high prevalence of these disorders, especially age-related cataract (~50% in +75-year olds[24]). No restriction was put on the medication that patients could receive (including acetylcholinesterase inhibitors, antidepressants, and hypnotics). The reference test for this diagnostic study was the diagnostic criteria applied as listed above. The HCs were included if they fulfilled the following inclusion criteria: (1) able to cooperate with

the investigations, (2) normal cognitive function as evaluated by the investigator, (3) age between 50 and 90 years, and none of the listed exclusion criteria applied to patients.

The HCs were recruited via a list of healthy participants recruited for previous finished trials and clinical studies and through advertising. No compensation was given to participants. Eleven participants were included before diagnosis during diagnostic work-up (eight patients later diagnosed with AD, two with DLB, and one patient with VCD) as part of a previous study.[6] The pupillometry measurements were not used to diagnose these patients and the diagnosis was given within 1 month after inclusion. The study referenced[6] showed no changes within 1 month for pupillary light response (PLR) parameters. We registered the study at clinicaltrials.gov (NCT05175664) and obtained permission from the regional ethics committee before the commencement of the study (file number: H-21045068). The tenets of the 1975 Helsinki Declaration were followed, and written informed consent was obtained from all participants. For participants with cognitive impairment, a caregiver had to be present during obtainment of written consent. The memory clinic physician was always consulted to ensure capacity to provide consent. The study is reported according to the TRIPOD reporting guidelines (see Supplementary Material for TRIPOD checklist).[25]

2.3 | Procedure

2.3.1 | Quantitative light reflex pupillometry

The hand-held PLR-3000 (NeurOptics®) Pupillometer system was used to measure the light reflex (for description of procedure see Supplementary Material).

2.3.2 | Primary outcome measures

The PLR-3000 provides the following parameters: baseline pre-stimulus pupil diameter (mm), peak constriction pupil diameter (mm), delta change between baseline and peak pupil diameter (%), latency (msec), average constriction velocity (mm/sec), maximum constriction velocity (mm/sec), average dilation velocity (mm/sec), and time to reach 75% of baseline value after peak constriction (T75) (sec).

2.3.3 | Pre-processing of data

The method for pre-processing follows the same procedure as described previously (see Supplementary Material).[6]

2.3.4 | Cognitive testing and demographic data

The MMSE[17] assesses global cognitive function and was administered by a trained nurse at an initial evaluation visit to the memory clinic for patients, whereas the investigator administered the MMSE for the HCs. In addition to the MMSE, patients underwent Addenbrooke's Cognitive Examination and a tailored diagnostic work-up, in some cases including neuropsychological testing, whereas this was done for HCs. A history of ophthalmological conditions, subjective visual acuity (to screen for major ocular disorders), medications, and eye surgery were taken from the patient and caregiver as well as extracted from medical files. HCs filled out a questionnaire, where these items were addressed as well. Demographic data, information on medication use, and eye surgery were extracted from medical files and entered into an electronic database (REDCap, Vancouver).

2.3.5 | Statistical analysis

R (ver. 4.2.2) was used for all analyses.[26] The source code used to produce the output can be found at https://github.com/Matgram/dqLRP/tree/main.

The sample size was estimated from a power calculation with the following parameters: alpha 0.05 and power 80%. As a surrogate measure of diagnostic performance, the study was powered to detect a difference of 0.19 mm/sec in the qLRP variable constriction velocity assuming a standard deviation of 0.33 mm/sec as reported in healthy individuals using the ClinCalc power calculator (https://clincalc.com/stats/samplesize.aspx).[27]

We calculated Spearman's correlation coefficient, *rho*, for correlations between all pupillometric and demographic variables.

We fitted multivariable logistic regression class prediction models with AD set as the case versus HCs, patients with cerebrovascular disease (mixed AD + VCD, and VCD, separately), and DLB in separate models. Three models were investigated for each class distinction, namely a model incorporating all qLRP parameters as mentioned in the previous section (Model 1), a sparse model with age and sex (excluding qLRP parameters with multicollinearity) (Model 2), and third an exploratory model, where qLRP parameters included in the sparse model were fitted with cubic splines using the *splines* package in *R* (Model 3). We considered a complete case analysis in all analyses. Assumption of no multicollinearity was assessed by examining variance inflation factors and was violated for Model 1. Linearity assumption was checked by visual inspection of residuals of sparse models and the influence of outliers was checked by inspecting differences in fits and DFBETAs. The performance of models was evaluated by comparing their area under the receiver operating characteristics curve (AUROC) and a corresponding 95% confidence interval ($CI_{95\%}$) using the DeLong test. The AUROC takes on values between 0.5 (chance level) and 1 (perfect discrimination). We investigated thresholds that maximized sensitivity and specificity respectively in a two-cutpoint procedure, where the first and the latter measure were set as a minimum criterion of 85% while maximizing the other measure resulting in a "rule-out" lower, sensitive threshold of the model as well as a "rule-in" higher, specific threshold and an interposed "grey zone." This was done to provide clinicians with cutpoints with the highest diagnostic applicability. A $CI_{95\%}$ for each corresponding measure was calculated using

TABLE 1 Cohort characteristics.

Parameter	Alzheimer's disease (*N* = 107)	Mixed AD (+vascular pathology) (*N* = 20)	Vascular cognitive dysfunction (*N* = 24)	Dementia with Lewy bodies (*N* = 53)	Healthy controls (*N* = 50)	*p*-Value
Age (years)						<0.001[a]
Mean (SD)	75.6 (7)	80.4 (5.2)	80.5 (5.1)	76.2 (6.5)	71.4 (7.9)	
Sex						<0.001[b]
Female	55 (51.4%)	12 (60%)	8 (33%)	8 (15.1%)	31 (62%)	
Male	52 (48.6%)	8 (40%)	16 (67%)	45 (84.9%)	19 (38%)	
Educational level (years)						<0.001[a]
Mean (SD)	11.8 (3.6)	11.4 (4.2)	9.8 (2.8)	11.8 (4)	14.8 (2.7)	
Range	7-20	7-17	7-17	7-19	6-18	
MMSE total score						<0.001[a]
Mean (SD)	25.8 (2.6)	22.4 (3.2)	25.5 (4)	25.3 (3.7)	29.1 (1.2)	
Range	19-30	18-28	12[c]-30	16-30	25-30	
Disease severity						
MCI	28 (25.9%)	0 (0%)	10 (42%)	5 (9.4%)	–	
Mild	75 (69.4%)	11 (55%)	9 (38%)	41 (77.4%)	–	
Moderate	5 (4.6%)	9 (45%)	5 (21%)	7 (13.2%)	–	
Medication with pupillary effect						<0.001[b]
No	95 (89.6%)	14 (70%)	16 (67%)	34 (66.7%)	45 (90%)	
Yes	11 (10.4%)	6 (30%)	8 (33%)	17 (33.3%)	5 (10%)	
Mild chronic eye disease						0.016[b]
No	69 (64.5%)	10 (50%)	12 (50%)	20 (37.7%)	32 (64%)	
Yes	38 (35.5%)	10 (50%)	12 (50%)	33 (62.3%)	18 (36%)	
Cataract surgery history						0.023[b]
No	81 (75.7%)	11 (55%)	16 (67%)	35 (66%)	44 (88%)	
Yes	26 (24.3%)	9 (45%)	8 (33%)	18 (34%)	6 (12%)	

[a]One-way ANOVA.
[b]Pearson's χ-squared test.
[c]One patient with cerebrovascular disease was included with an MMSE total score of 12, evaluated as partly due to a slight language barrier. Percentages may add up to more than 100% due to rounding. N missing < 5 not shown.
Abbreviations: AD = Alzheimer's disease; ANOVA, analysis of variance; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination.

the *boot_ci* function (*N* = 2000 runs and out-of-bag estimates) from the *cutpointr* package.

Sensitivity analyses were performed to assess the influence of variables that could confound the discriminatory ability of qLRP, namely the presence of mild eye disease, the use of medication that could exert pupillary influence, history of cataract surgery, and lastly the performance in early (MCI stage) and dementia stage in dichotomized groups. Further, internal validation with 10-fold 1000 times cross-validation was performed for Model 2 using the *caret* package in *R*.

To use the prediction model, qLRP parameters along with age and sex can be entered as coefficients in the modeling scenarios using the provided web applications (see Supplementary Material).

The alpha was set at 0.05 and differences considered significant below this level, and only two-tailed tests were used.

3 | RESULTS

The cohort characteristics are shown in Table 1. Patients with AD were generally older than HCs and younger than patients with mixed AD and cerebrovascular disease. There was a skewed distribution of men and women, as patients with AD had both sexes equally represented whereas patients with DLB were mainly male. Patients with AD were classified as MCI or mild dementia in 26% and 69% of cases, respectively, reflecting a relatively early disease stage. Patients with AD and HCs had equal proportions receiving medication that could influence the pupil, whereas a higher proportion was found for patients with cerebrovascular disease and DLB. All patients had a higher proportion than HCs of individuals who had a history of cataract surgery. MMSE total scores were different between groups, and patients with mixed pathology had the lowest MMSE score.

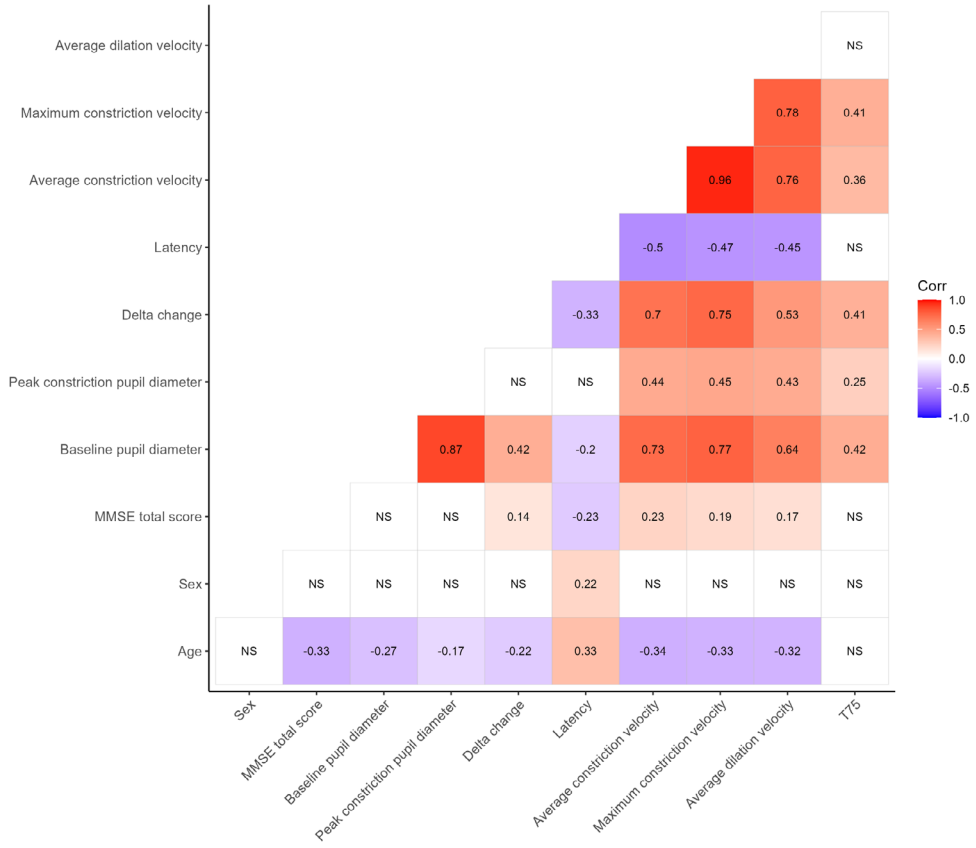


FIGURE 1 A correlogram showing Spearman's rank correlation coefficients for significant correlations. Correlation is shown for the direction of male sex (female sex is in the opposite direction). NS, not significant.

In correlation analysis (shown in Figure 1), we found moderate-strong positive correlations between the qLRP parameters baseline pupil diameter, peak constriction pupil diameter, average and maximum constriction velocity, and delta. MMSE total score correlated weakly positively with delta, constriction/dilation velocity, and negatively with latency. Age was weakly correlated negatively with all qLRP parameters, except latency (positive correlation) and T75 (no correlation). Male sex was weakly, positively correlated with latency.

All qLRP parameters differed between diagnostic groups, except T75 ($p = 0.34$) (Table 2 and Figures 2–3).

3.1 | Diagnostic models

3.1.1 | AD versus HC

The model which also incorporated age and sex showed diagnostic ability slightly higher than the model without age and sex (AUROC 0.74, $CI_{95\%}$: 0.66-0.83) and the cross-validated estimate was only slightly lower, indicating external validity of the model (Table 3). When we applied a two-cutpoint procedure to Model 2, it showed a specificity of 40% with a sensitivity at 85%. In the Supplementary Material (Figures S1-S8), the associated negative and positive predictive values are shown for varying disease prevalence.

3.1.2 | AD versus DLB

The diagnostic model with only qLRP parameters had the ability to discriminate between AD and DLB above chance level (AUROC 0.67, $CI_{95\%}$: 0.58-0.76), and the addition of age and sex improved the model slightly. Model 3, which explored a non-linear association, showed the highest diagnostic ability (AUROC 0.86) but could suffer from overfitting.

3.1.3 | AD versus mixed AD and VCD

When age and sex were added to the model, it showed a similar diagnostic ability to discriminate both mixed AD and VCD from AD, both above chance level (Table 3). The cross-validated AUROCs were lower for both group discriminations (0.73 and 0.7, respectively), which indicated that the models were slightly optimistic. The diagnostic models had similar specificity and sensitivity when applying a two-cutpoint procedure.

3.1.4 | Sensitivity analyses

The results of the sensitivity analyses are reported in the Tables S1-S4.

TABLE 2 Quantitative light reflex pupillometry parameters according to diagnostic group

Parameter	Alzheimer's disease	Mixed AD (+vascular pathology)	Vascular cognitive dysfunction	Dementia with Lewy bodies	Healthy controls	*p*-Value
Baseline pupil diameter (mm)						0.009[a]
Mean (SD)	3.03 (0.51)	2.95 (0.5)	3.06 (0.69)	2.87 (0.6)	3.27 (0.56)	
Range	1.9-4.33	2.13-4.03	1.87-4.53	1.8-4.6	2.13-5.13	
Peak constriction pupil diameter (mm)						0.023[a]
Mean (SD)	2.05 (0.33)	2.14 (0.35)	2.16 (0.48)	1.97 (0.33)	2.18 (0.32)	
Range	1.23-3	1.43-2.9	1.33-3.27	1.3-3.07	1.43-3.13	
Delta change (percent)						<0.001[a]
Mean (SD)	31.8 (6)	27.1 (7)	29 (5.4)	30.4 (6.1)	32.8 (4.5)	
Range	15-43	12.3-38.7	17.3-38.3	16.7-46	21.1-40.7	
Latency (sec)						<0.001[a]
Mean(SD)	0.240(0.027)	0.263 (0.031)	0.255 (0.027)	0.251 (0.025)	0.227 (0.021)	
Range	0.177-0.31	0.21-0.333	0.2-0.313	0.2-0.31	0.2-0.29	
Average constriction velocity (mm/sec)						<0.001[a]
Mean (SD)	1.88 (0.52)	1.6 (0.54)	1.7 (0.58)	1.65 (0.59)	2.18 (0.46)	
Range	0.53-3.62	0.71-2.64	0.59-2.58	0.67-3.92	1.14-3.24	
Maximum constriction velocity (mm/sec)						<0.001[a]
Mean (SD)	2.85 (0.78)	2.4 (0.78)	2.65 (0.87)	2.56 (0.89)	3.25 (0.72)	
Range	0.85-5.23	1.03-3.64	0.99-4.11	0.92-5.78	1.8-4.96	
Average dilation velocity (mm/sec)						0.009[a]
Mean (SD)	0.94 (0.24)	0.84 (0.34)	0.87 (0.24)	0.88 (0.27)	1.022 (0.19)	
Range	0.33-1.54	0.34-1.49	0.47-1.31	0.42-1.92	0.68-1.45	
Time to reach 75% of baseline (sec)						0.339[a]
Mean (SD)	1.62 (0.65)	1.45 (0.46)	1.472 (0.52)	1.63 (0.99)	1.78 (0.69)	
Range	0.71-4.19	0.9-2.34	0.97-3.55	0.73-7.29	0.85-3.87	

[a]One-way ANOVA.

Abbreviations: AD, Alzheimer's disease; ANOVA, analysis of variance;

4 | DISCUSSION

This study investigated the diagnostic performance of qLRP in diagnosing AD in a memory clinic cohort. We found that our models performed with an ability above chance level when distinguishing AD from HCs, DLB, and patients with cerebrovascular disease. The diagnostic models were robust to sensitivity analyses and suspected confounding ocular factors did not seem to drive the discriminatory ability of qLRP.

Our results indicate that qLRP might prove useful as a diagnostic aide, both when distinguishing healthy from patients with AD as well as in the differential diagnostic process of distinguishing different dementia etiologies, where the highest discriminatory ability was found for cerebrovascular disease. We acknowledge that this tool should not be used as a standalone diagnostic biomarker for AD, and we did not evaluate the added value of qLRP to current diagnostic tests, which remains to be elucidated in future studies. We have not identified studies, which have previously investigated the ability of qLRP to distinguish between dementia etiologies and our study provides the first evidence that qLRP parameters may be affected in both patients with DLB and patients with cerebrovascular disease. We mainly found more exaggerated slowing of the pupillary reflex as compared with AD, which was surprising. We found the highest AUROCs when we applied qLRP in the AD vs. cerebrovascular disease scenario. It seems that cerebrovascular disease may be distinguished with some certainty from AD using only qLRP. While we could not provide data that explain the driving factor behind the altered qLRP parameters in cerebrovascular disease, one hypothesized explanation could be extant white matter lesions, which are prevalent in cerebrovascular disposed individuals, and their anatomical localization.[28] A recent study investigated whether strategic infarctions might influence pupillary measurements but did not identify differences between groups, hinting at other possible explanations such as white matter hyperintensities.[29] The proposed pathophysiology of white matter hyperintensities, which includes axonal loss and demyelination,[30] could explain why certain qLRP parameters were shown to be affected, but further research is needed to establish this association.

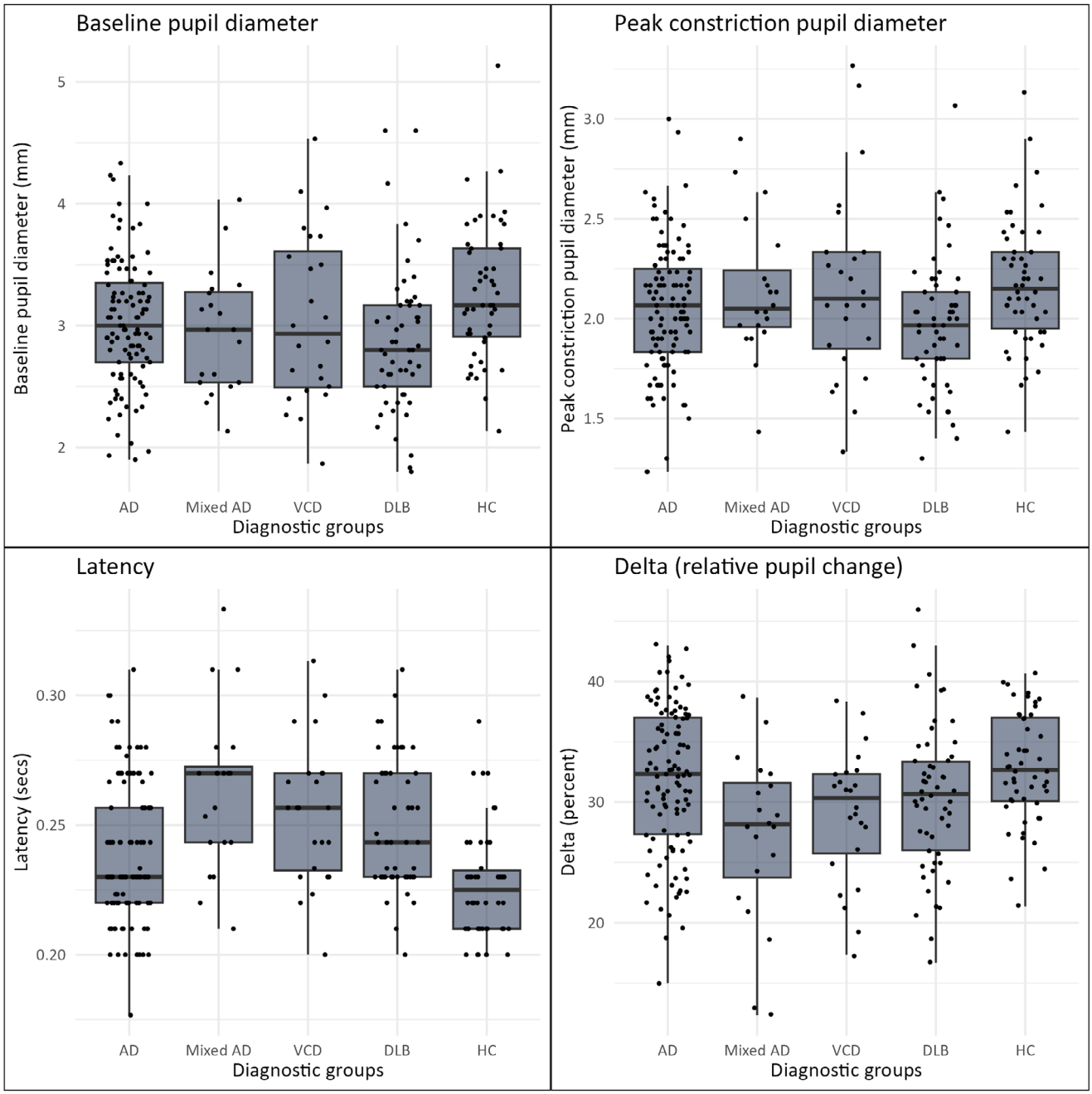


FIGURE 2 Boxplots with overlayed dot-plots of qLRP parameters. The horizontal stroke in the box indicates the median, and the boxes indicate the 25% and 75% interquartile range (IQR). Whiskers are 1.5 × IQR. AD, Alzheimer's disease; DLB, dementia with Lewy bodies; HC, healthy controls; mixed AD, Alzheimer's disease with mixed pathology (AD+vascular pathology); qLRP, quantitative light reflex pupillometry; VCD, vascular cognitive dysfunction.

qLRP was able to distinguish with some certainty AD from DLB, which is an important differential diagnostic to AD, as DLB shows a high degree of co-pathology and shares symptoms with AD.[31,32] Previous studies showed that Parkinson's disease, which shares neuropathological features with DLB, could be distinguished from HCs using qLRP,[33] but no studies have to our knowledge investigated the potential of qLRP in distinguishing AD and DLB. The mechanism that drives the diagnostic ability of qLRP for this discrimination is not known in detail, but autonomic dysfunction could be hypothesized to play a role as the light reflex is under the influence of the autonomic nervous system.[34] As such it could be assumed that autonomic dysfunction, which is pronounced in DLB and part of the diagnostic criteria,[22] could drive the effect observed on qLRP parameters, and qLRP could be a sensitive marker in gauging this aspect of the disease, although more research is needed.

We provided sensitivity and specificity thresholds of our diagnostic models to increase clinical applicability. This two-cutpoint method and the different diagnostic models could be applied in specific

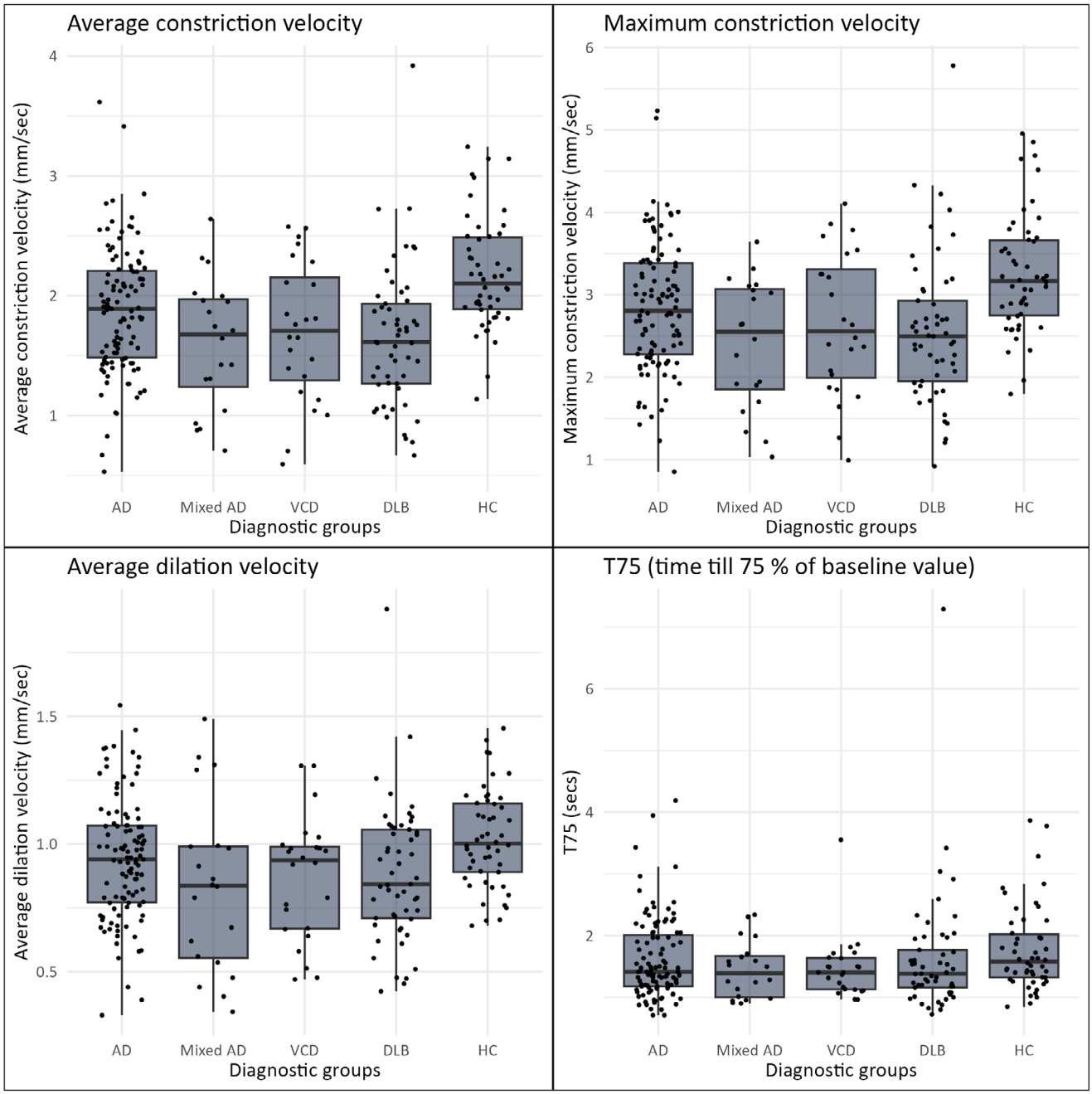


FIGURE 3 Boxplots with overlayed dot-plots of qLRP parameters. The horizontal stroke in the box indicates the median, and the boxes indicate the 25% and 75% interquartile range (IQR). Whiskers are 1.5 × IQR. AD, Alzheimer's disease; DLB, dementia with Lewy bodies; HC, healthy controls; mixed AD, Alzheimer's disease with mixed pathology (AD+vascular pathology); qLRP, quantitative light reflex pupillometry; VCD, vascular cognitive dysfunction.

situations, for example, at an initial triage entry-point to memory clinics such as primary care providers and in the setting of diagnostic centers (i.e., memory clinics). Applying both thresholds in a primary care setting, the sensitive threshold would indicate a "rule-out" level, whereby patients falling below the threshold have little chance of having AD vs. a healthy condition (non-neurodegenerative), and the upper threshold could indicate that the presence of AD may be likely and a referral to a specialist center should be made. The resulting interposed "gray zone" may benefit from further testing before a decision is made, and ultimately represents a downside to our cutpoint optimization. The other diagnostic models (vs. cerebrovascular disease and DLB) may have relevance in more specialized settings where an ability to differentiate dementia disorders is needed. These testing scenarios would need to be evaluated in future studies to evaluate the added value of qLRP in these particular settings.

Previous studies have shown a remarkable discriminatory ability of qLRP for distinguishing AD vs. healthy, with near-perfect AUC.[13–15] We could not replicate these results, and the reasons for this could

TABLE 3 Performance of diagnostic models.

Parameter	AUROC [95% CI]	Specificity at minimum 85% sensitivity [95% CI] (cutpoint)	Sensitivity at minimum 85% specificity [95% CI] (cutpoint)	Cross-validated AUROC
AD vs. HC Model 1	0.72 [0.64-0.8]			
AD vs. HC Model 2	0.74 [0.66-0.83]	40% [19-68%] (0.52)	38% [16-72%] (0.81)	0.68
AD vs. HC Model 3	0.81 [0.74-0.88]			
AD vs. DLB Model 1	0.67 [0.58-0.76]			
AD vs. DLB Model 2	0.75 [0.67-0.83]	47% [21-67%] (0.52)	46% [18-71%] (0.81)	0.69
AD vs. DLB Model 3	0.86 [0.79-0.92]			
AD vs. Mixed AD + VCD Model 1	0.75 [0.66-0.83]			
AD vs. Mixed AD + VCD Model 2	0.8 [0.72-0.88]	57% [31-79%] (0.59)	52% [36-78%] (0.82)	0.73
AD vs. Mixed AD + VCD Model 3	0.899 [0.85-0.95]			
AD vs. VCD Model 1	0.73 [0.61-0.85]			
AD vs. VCD Model 2	0.81 [0.72-0.9]	63% [25-89%] (0.72)	58% [33-85%] (0.86)	0.7
AD vs. VCD Model 3	–			

Notes: Model 1 = all qLRP parameters as explaining variables, Model 2 = Sparse model (excluding peak constriction and maximum velocity qLRP variables) with age and sex as covariates, Model 3 = as Model 2 with cubic spline terms fitted for all qLRP parameters. Cutpoint refers to the probability threshold of the logistic regression model prediction. Model algorithm did not converge.

Abbreviations: AD, Alzheimer's disease; AUROC, area under the receiver operating characteristics curve; qLRP, quantitative light reflex pupillometry; VCD, vascular cognitive dysfuntion.

be manifold: first, our cohort differed, as we included a larger cohort of patients with frequent ocular disorders, and allowed medication that could have affected the pupillary response, with a wide MMSE range, indicating a representative sample of a memory clinic, which is heterogenous.[35] Also, the pupillometry system applied differed between some previous studies and ours. We used a hand-held, research-grade pupillometer, while other groups applied in-house developed, stationary pupillometers with a faster sampling rate resulting in smaller measurement errors (see Table S5 for review of previous qLRP studies) and a lower percentage of discarded samples, which may explain the differences observed. In that regard, the low-cost and quick application of hand-held qLRP seems advantageous in the clinic, while the experimental approach seems more useful in a proof-of-mechanism setting. Further, recent studies have shown promise in targeting specific melanopsin retinal ganglion cells of the retina. This approach seems interesting given the fact that patients with AD show selective disruption of function in these cells.[36]

While a substantial proportion of patients across diagnostic groups had ocular pathology, which is common in older age and populations of patients with dementia,[37] these factors did not affect the diagnostic ability of qLRP in sensitivity analyses. Earlier studies on qLRP have, with few exceptions,[38] excluded patients with even mild ocular disorders.[13–15,39–47] This could have induced selection bias in previous studies, which seemingly can be avoided without influencing the discriminatory ability of qLRP to a large degree.

Our study had several strengths. First, we examined a large cohort of patients and relevant comparators in a real-world setting, reflecting the setting where one use case of qLRP seems relevant. Second, we applied broad inclusion criteria which included participants with mild ocular disorders, thus enhancing the generalizability of our results; and third, we applied a commercially available pupillometer, which increases the reproducibility of our results.

We also acknowledge several limitations of our study. First, we included a healthy control group without cognitive complaints. For a true diagnostic ability, an ideal comparison group for the detection of AD might be a subjective cognitive decline group with adequate follow-up and diagnostic work-up to rule out incipient dementia. Second, our VCD group was small which led to insufficient data in certain modeling situations. Third, we did not perform ophthalmological examination beyond ophthalmoscopy without a specific focus on fundus pathology,

which may have led to some patients erroneously not having relevant eye pathology, such as optic nerve pathology, described. This could have been mitigated by the inclusion of fundus exam and/or optical coherence tomography and represents a shortcoming of this study. Fourth, we did not validate our results in an independent dataset; however, we did include an out-of-sample estimate of the AUROC by using cross-validation. Fifth, we could not account for senile miosis, and while we ran sensitivity analyses excluding patients with mild eye disorders, this could also represent a considerable source of residual confounding; however, this may have been mitigated by the inclusion of age-matched controls. Sixth, as we mainly used data from the right eye, a between-eyes correlation was not considered which may have reduced the statistical power.

In this proof-of-concept study, we have shown that qLRP may serve as a supportive digital, bedside biomarker for AD in a specialized memory clinic setting, where qLRP could be used to assist in distinguishing between patients with AD, healthy individuals, patients with cerebrovascular disease, and DLB. Our findings may be used to validate qLRP as a sensitive digital, diagnostic biomarker for AD, and future studies should apply qLRP in relevant settings in an unbiased manner along with other diagnostic assessment modalities to validate our findings.

ACKNOWLEDGMENTS

The authors have nothing to report. Funding for this project was received from Frimodt-Heineke Fonden, Absalon Foundation, Rigshospitalets Forskningsfond and Overretssagfører L. Zeuthens Mindelegat. The salary of Mathias Holsey Gramkow was supported by Michaelsen Fonden, The Danish Alzheimer Society Research Foundation, and Ellen Mørchs Fond. The funders had no role in the design or reporting of the study.

CONFLICT OF INTEREST STATEMENT

S.G.H. has taught courses for Novo Nordisk with honoraria paid directly to the Danish Dementia Research Center. The remaining authors have no conflicts of interest. Author disclosures are available in the supporting information.

CONSENT STATEMENT

All human subjects provided informed consent.

ORCID

Mathias Holsey Gramkow https://orcid.org/0000-0002-3827-6489

REFERENCES

1. Gustavsson A, Norton N, Fast T, et al. Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's Dement*. 2023;19(2):658-670. doi:10.1002/alz.12694
2. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's disease. *N Engl J Med*. 2023;388:9-21. doi:10.1056/NEJMoa2212948
3. Sims JR, Zimmer JA, Evans CD, et al. Donanemab in Early Symptomatic Alzheimer Disease: the TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *Jama*. 2023;330:512-527. doi:10.1001/jama.2023.13239
4. Hampel H, Au R, Mattke S, et al. Designing the next-generation clinical care pathway for Alzheimer's disease. *Nat Aging*. 2022;2:692-703. doi:10.1038/s43587-022-00269-x
5. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *Npj Digit Med*. 2019;2:9. doi:10.1038/s41746-019-0084-2
6. Gramkow MH, Clemmensen FK, Waldemar G, Hasselbalch SG, Frederiksen KS. Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort. *J Neurol Sci*. 2024;456:122856. doi:10.1016/j.jns.2023.122856
7. McDougal DH, Gamlin PD. Autonomic control of the eye. *Compr Physiol*. 2015;5:439-473. doi:10.1002/cphy.c140014.Autonomic
8. Peinkhofer C, Knudsen GM, Moretti R, Kondziella D. Cortical modulation of pupillary function: systematic review. *PeerJ Comput Sci*. 2019;2019:e6882. doi:10.7717/peerj.6882
9. Bartus RT. On neurodegenerative diseases, models, and treatment strategies: lessons learned and lessons forgotten a generation following the cholinergic hypothesis. *Exp Neurol*. 2000;163:495-529. doi:10.1006/exnr.2000.7397
10. Xia Y, Maruff P, Doré V, et al. Longitudinal trajectories of basal forebrain volume in normal aging and Alzheimer's disease. *Neurobiol Aging*. 2023;132:120-130. doi:10.1016/j.neurobiolaging.2023.09.002
11. Scinto LFM, Frosch M, Wu CK, Daffner KR, Gedi N, Geula C. Selective cell loss in Edinger-Westphal in asymptomatic elders and Alzheimer's patients. *Neurobiol Aging*. 2001;22:729-736. doi:10.1016/S0197-4580(01)00235-4
12. Mavroudis IA, Manani MG, Petrides F, et al. Dendritic and spinal alterations of neurons from Edinger-Westphal nucleus in Alzheimer's disease. *Folia Neuropathol*. 2014;52:197-204. doi:10.5114/fn.2014.43791
13. Fotiou DF, Brozou CG, Haidich AB, et al. Pupil reaction to light in Alzheimer's disease: evaluation of pupil size changes and mobility. *Aging Clin Exp Res*. 2007;19:364-371. doi:10.1007/BF03324716
14. Fotiou D, Kaltsatou A, Tsiptsios D, Nakou M. Evaluation of the cholinergic hypothesis in Alzheimer's disease with neuropsychological methods. *Aging Clin Exp Res*. 2015;27:727-733. doi:10.1007/s40520-015-0321-8
15. Frost S, Robinson L, Rowe CC, et al. Evaluation of cholinergic deficiency in preclinical Alzheimer's disease using pupillometry. *J Ophthalmol*. 2017;2017:1-8. doi:10.1155/2017/7935406
16. Granholm EL, Panizzon MS, Elman JA, et al. Pupillary responses as a biomarker of early risk for Alzheimer's disease. *J Alzheimer's Dis*. 2017;56:1419-1428. doi:10.3233/JAD-161078
17. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198. doi:10.1016/0022-3956(75)90026-6
18. Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*. 2000;55:1613-1620. doi:10.1212/01.wnl.0000434309.85312.19
19. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:270-279. doi:10.1016/j.jalz.2011.03.008
20. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:263-269. doi:10.1016/j.jalz.2011.03.005
21. Sachdev P, Kalaria R, O'Brien J, et al. Diagnostic criteria for vascular cognitive disorders: a VASCOG statement. *Alzheimer Dis Assoc Disord*. 2014;28:206-218. doi:10.1097/WAD.0000000000000034

22. McKeith IG, Boeve BF, Dickson DW, et al. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. *Neurology*. 2017;89:88-100. doi:10.1212/WNL.0000000000004058
23. McKeith IG, Ferman TJ, Thomas AJ, et al. Research criteria for the diagnosis of prodromal dementia with Lewy bodies. *Neurology*. 2020;94:743-755. doi:10.1212/WNL.0000000000009323
24. Purola PKM, Nättinen JE, Ojamo MUI, et al. Prevalence and 11-year incidence of common eye diseases and their relation to health-related quality of life, mental health, and visual impairment. *Qual Life Res*. 2021;30:2311-2327. doi:10.1007/s11136-021-02817-1
25. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162:55-63. doi:10.7326/M14-0697
26. R Core Team. R: A Language and Environment for Statistical Computing 2021.
27. Taylor WR, Chen JW, Meltzer H, et al. Quantitative pupillometry, a new technology: normative data and preliminary observations in patients with acute head injury—Technical note. *J Neurosurg*. 2003;98:205-213. doi:10.3171/jns.2003.98.1.0205
28. Botz J, Lohner V, Schirmer MD. Spatial patterns of white matter hyperintensities: a systematic review. *Front Aging Neurosci*. 2023;15:1-13. doi:10.3389/fnagi.2023.1165324
29. Peinkhofer C, Martens P, Grand J, et al. Influence of strategic cortical infarctions on pupillary function. *Front Neurol*. 2018;9. doi:10.3389/fneur.2018.00916
30. Habes M, Erus G, Toledo JB, et al. White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain*. 2016;139:1164-1179. doi:10.1093/brain/aww008
31. Gibson LL, Abdelnour C, Chong J, Ballard C, Aarsland D. Clinical trials in dementia with Lewy bodies: the evolving concept of co-pathologies, patient selection and biomarkers. *Curr Opin Neurol*. 2023;36:264-275. doi:10.1097/WCO.0000000000001173
32. Slaets S, Theuns J, Sleegers K, et al. Amyloid pathology influences Aβ1-42 cerebrospinal fluid levels in dementia with lewy bodies. *J Alzheimer's Dis*. 2013;35:137-146. doi:10.3233/JAD-122176
33. Fotiou DF, Stergiou V, Tsiptsios D, Lithari C, Nakou M, Karlovasitou A. Cholinergic deficiency in Alzheimer's and Parkinson's disease: evaluation with pupillometry. *Int J Psychophysiol*. 2009;73:143-149. doi:10.1016/j.ijpsycho.2009.01.011
34. Idiaquez J, Roman GC. Autonomic dysfunction in neurodegenerative dementias. *J Neurol Sci*. 2011;305:22-27. doi:10.1016/j.jns.2011.02.033
35. Gramkow MH, Gjerum L, Koikkalainen J, et al. Prognostic value of complementary biomarkers of neurodegeneration in a mixed memory clinic cohort. *PeerJ*. 2020;2020:e9498. doi:10.7717/peerj.9498
36. Romagnoli M, Amore G, Avanzini P, Carelli V, La Morgia C. Chromatic pupillometry for evaluating melanopsin retinal ganglion cell function in Alzheimer's disease and other neurodegenerative disorders: a review. *Front Psychol*. 2023;14:1-11. doi:10.3389/fpsyg.2023.1295129
37. Chua J, Zhang Z, Wong D, et al. Age-related eye diseases in individuals with mild cognitive impairment and Alzheimer's disease. *Front Aging Neurosci*. 2022;14. doi:10.3389/fnagi.2022.933853
38. Ferrario E, Molaschi M, Villa L, Varetto O, Bogetto C, Nuzzi R. Is videopupillography useful in the diagnosis of Alzheimer's disease? *Neurology*. 1998;50:642-644. doi:10.1212/WNL.50.3.642
39. Romagnoli M, Stanzani Maserati M, De Matteis M, et al. Chromatic Pupillometry Findings in Alzheimer's Disease. *Front Neurosci*. 2020;14:1-10. doi:10.3389/fnins.2020.00780
40. Prettyman R, Bitsios P, Szabadi E. Altered pupillary size and darkness and light reflexes in Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 1997;62:665-668. doi:10.1136/jnnp.62.6.665
41. Fotiou F, Fountoulakis KN, Tsolaki M, Goulas A, Palikaras A. Changes in pupil reaction to light in Alzheimer's disease patients: a preliminary report. *Int J Psychophysiol*. 2000;37:111-120. doi:10.1016/S0167-8760(00)00099-4
42. Tales A, Troscianko T, Lush D, Haworth J, Wilcock GK, Butler SR. The pupillary light reflex in aging and Alzheimer's disease. *Aging Clin Exp Res*. 2001;13:473-478.
43. Granholm E, Morris S, Galasko D, Shults C, Rogers E, Vukov B. Tropicamide effects on pupil size and pupillary light reflexes in Alzheimer's and Parkinson's disease. *Int J Psychophysiol*. 2003;47:95-115. doi:10.1016/S0167-8760(02)00122-8
44. Frost S, Kanagasingam Y, Sohrabi H, et al. Pupil response biomarkers for early detection and monitoring of Alzheimer's disease. *Curr Alzheimer Res*. 2013;10:931-939. doi:10.2174/15672050113106660163
45. Bittner DM, Wieseler I, Wilhelm H, Riepe MW, Müller NG. Repetitive Pupil Light Reflex: potential Marker in Alzheimer's Disease? *J Alzheimer's Dis*. 2014;42:1469-1477. doi:10.3233/JAD-140969
46. Kawasaki A, Ouanes S, Crippa SV, Popp J. Early-stage Alzheimer's disease does not alter pupil responses to colored light stimuli. *J Alzheimer's Dis*. 2020;75:1191-1201. doi:10.3233/JAD-200120
47. La Morgia C, Mitolo M, Romagnoli M, et al. Multimodal investigation of melanopsin retinal ganglion cells in Alzheimer's disease. *Ann Clin Transl Neurol*. 2023;10:918-932. doi:10.1002/acn3.51773

SUPPORTING INFORMATION

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

How to cite this article: Gramkow MH, Clemmensen FK, Sjælland NS, Waldemar G, Hasselbalch SG, Frederiksen KS. Diagnostic performance of light reflex pupillometry in Alzheimer's disease. *Alzheimer's Dement*. 2024;16:e12628. https://doi.org/10.1002/dad2.12628

STUDY III

Diagnostic performance of actigraphy in Alzheimer's disease using a machine learning classifier – a cross-sectional memory clinic study

Gramkow, M. H., Brink-Kjær, A., Clemmensen, F. K., Sjælland, N. S., Waldemar, G., Jennum, P., Hasselbalch, S. G., & Frederiksen, K. S. (2025). Diagnostic performance of actigraphy in Alzheimer's disease using a machine learning classifier – a cross-sectional memory clinic study. Alzheimer's Research & Therapy 2025 17:1, 17(1), 1–12.
DOI: https://doi.org/10.1186/S13195-025-01751-5

The supplementary material related to this journal article can be accessed at:
https://doi.org/10.1186/S13195-025-01751-5
or by scanning the QR-code below

Gramkow *et al. Alzheimer's Research & Therapy* (2025) 17:111
https://doi.org/10.1186/s13195-025-01751-5

Alzheimer's Research & Therapy

RESEARCH Open Access

Diagnostic performance of actigraphy in Alzheimer's disease using a machine learning classifier – a cross-sectional memory clinic study

Mathias Holsey Gramkow[5*], Andreas Brink-Kjær[2,3], Frederikke Kragh Clemmensen[1], Nikolai Sulkjær Sjælland[1], Gunhild Waldemar[1,4], Poul Jennum[3,4], Steen Gregers Hasselbalch[1,4] and Kristian Steen Frederiksen[1,4]

Abstract

Background Movement patterns, activity levels and circadian rhythm are altered in Alzheimer's disease (AD) and can be assessed by actigraphy using wearable sensors. We aimed to determine the diagnostic performance of actigraphy in AD in a memory clinic population by using a machine-learning classifier.

Methods In our single-center cross-sectional study, 70 patients with AD (MCI-moderate dementia), dementia with Lewy bodies (DLB) ($N=29$) and cerebrovascular disease (CVD) ($N=23$), and 48 elderly healthy controls were included. Participants underwent actigraphy at home using two body-worn sensors (SENS Motion®) for 1 week. We derived movement patterns (walking, running, resting, etc.) from raw accelerometry data using a proprietary algorithm. By evaluating the movement patterns during day and nighttime, we calculated 510 activity-related features, including robustness and fragmentation of the circadian rhythm. These features were used to train a machine learning (ML) classifier using logistic regression. We evaluated the performance of our classifier by assessing the accuracy and precision of predictions.

Results We found that movement patterns as well as the robustness and fragmentation of the circadian rhythm differed significantly between groups. During the daytime, patients with AD performed less moderate activity and walked less than the healthy group. While we achieved a modest accuracy of 68.8% for differentiating AD and healthy, the performance was highest (accuracy: 80–89%; precision: 69–84%) when ML was applied to actigraphy data to differentiate dementia etiologies (AD vs. DLB + AD vs. CVD).

Conclusion Actigraphy accurately identifies different dementia etiologies and could serve as a supplement to diagnostic investigations in patients with suspected AD for differential diagnostic purposes.

Keywords Actigraphy, Activity types, Physical activity, Alzheimer's disease, Diagnosis, Digital biomarker

*Correspondence:
Mathias Holsey Gramkow
mathias.holsey.gramkow@regionh.dk

Full list of author information is available at the end of the article

Introduction

Wearable digital health technologies represent a promising approach for assisting in the diagnosis of Alzheimer's disease (AD) [1]. There is a large unmet need for easily applicable diagnostic markers, due to an expected increase in the prevalence of AD and recently approved disease-modifying therapies [2–4]. Actigraphy can capture physical activity levels in a home environment [5], has been used previously to assess circadian function and behavioral changes associated with AD [6–10], and thus may serve diagnostic purposes. As recently proposed [11], a paradigmatic shift is needed to accommodate the disease burden of AD. Digital health technologies, such as actigraphy, are expected to play a major role in this transition towards a digitized memory clinic and could be add-on diagnostic tools.

A major challenge for diagnosing AD correctly is the differential diagnoses versus other dementia disorders, such as dementia with Lewy bodies (DLB), and vascular cognitive dysfunction (VCD). Patients with DLB and VCD are frequently evaluated in memory clinics [12] and patients with DLB show a high degree of co-pathology with AD making a correct diagnosis difficult [13]. Tools to help differentiate these entities in a clinical setting are therefore needed. These tools should preferably be less costly and more easily accessible.

Activity levels exhibit a decline in patients with AD and dementia in general [8, 14], and increased fragmentation of the circadian rhythm has been shown already in the preclinical stages of AD [15, 16]. A number of behavioral changes, due to neuropsychiatric symptoms such as apathy, are prevalent at an early stage of the disease, which can be reflected in a lower overall activity level [6]. A larger cohort study from 2012 showed that decreased activity levels are associated with increased risk of developing AD and that lower activity levels were predictive of greater cognitive decline [17].Whether qualitative differences exist relating to which activity types (e.g., walking, running, cycling) are particularly affected is yet to be explored, although a recent study highlighted differences for moderate-intensity activity, being significantly lower in individuals with mild AD [18]. As actigraphy offers an inexpensive method of assessing these aspects and has shown promise in detecting salient features of AD [19], the technology may be used to assist in diagnosing AD. An additional aspect is the ability of actigraphy to capture natural behavior outside clinic settings. Traditional scales for assessing activities of daily living (ADL) are limited by the need for proxy rating [20]. As such, actigraphy may be envisioned as a method for capturing and tracking day-to-day variability as a digital biomarker for general ADL functioning.

Due to the multidimensionality of actigraphic data, hypothesis-driven approaches, while valid for certain research questions, may overlook important elements of time-distributed complex data [21]. As such, machine learning provides a method for selecting relevant parameters in predictive modeling, without this pitfall. In the present study, we investigated a machine learning approach for predicting an AD diagnosis in a well-described memory clinic cohort, using actigraphy-derived activity categories and circadian rhythm with appropriate comparators, namely aged healthy controls, patients with DLB, and patients with cerebrovascular disease. We hypothesized that actigraphy could differentiate healthy and disease states, and that actigraphy had differential diagnostic potential.

Materials and methods

Study design

Single-center, cross-sectional diagnostic study.

Participants

This study is part of an investigation of digital health technologies, from which results on quantitative light reflex pupillometry have been reported [22]. The recruitment procedure has been described previously, and the inclusion of participants in the present study is shown in a flowchart in Fig. 1. Briefly, patients were diagnosed at a university memory clinic and underwent comprehensive diagnostic assessment. The diagnosis was given following a multi-disciplinary (neurology, geriatrics, psychiatry, neuropsychology, nurses) consensus conference.

Patients were included if they fulfilled the inclusion criteria: (1) a diagnosis of AD (both mild cognitive impairment (MCI) and dementia stages) [23, 24], AD with mixed pathology (cerebrovascular co-pathology) [24], cognitive dysfunction due to cerebrovascular disease [25], or dementia with Lewy bodies (MCI and dementia) [26, 27], (2) An MMSE total score > 15, (3) able to provide informed consent and none of the following exclusion criteria: (1) diagnosis of concurrent neurological (other than AD) or psychiatric disorder (mild depression allowed), (2) excessive alcohol intake > 5 units per day or substance use, (3) concurrent participation in interventional studies, (4) other known brain disorder which could cause cognitive dysfunction. Healthy controls were included if they fulfilled the following inclusion criteria (1) ability to cooperate with the investigations, (2) normal cognitive function as evaluated by the investigator, (3) age between 50 and 90 years, and none of the listed exclusion criteria applied to patients. Patients were not institutionalized, except for one patient with DLB.

The study (IPATCH-AD) was registered at clinicaltrials.gov (Clinical Trial No.: NCT05175664, registration date: 4th of January 2022) and carried out with permission from the regional ethics committee (file number: H-21045068). The tenets of the 1975 Helsinki

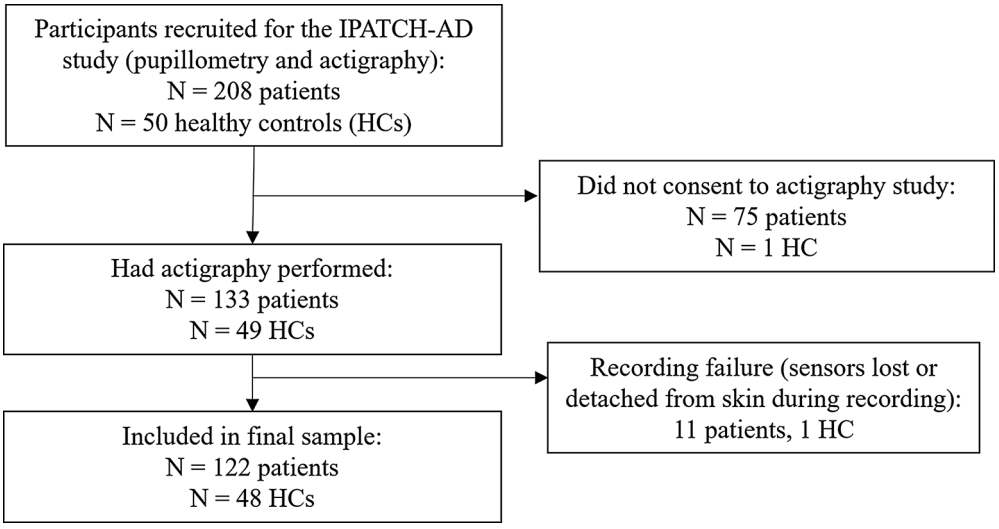


Fig. 1 Flow chart for the actigraphy part of the IPATCH-AD study

Declaration were followed, and written informed consent was obtained from all participants ensuring the capacity to provide consent after consulting the memory clinic physician responsible for the patient.

Cognitive testing and demographic data

The Mini-Mental State Examination (MMSE) [28], assessing global cognitive function, was administered by the investigator for the healthy controls. MMSE score, administered by a doctor at the most recent visit to the memory clinic was used for patients. Demographic data, including data on age, sex, educational level, and information on medication use that could influence the circadian rhythm (antidepressants, antipsychotics, hypnotics and sedatives, and anxiolytics), was extracted from the medical files and entered into an electronic database (REDCap, Vancouver). Healthy controls filled out a questionnaire, where this was addressed as well. Patients and controls were also asked to gauge their maximal walking distance in a single session (answers were capped at 10.000 m) as a measure of functional capacity.

Procedure

Actigraphy

Patients and controls had actigraphy performed in their home environment using the SENS® Motion Actigraph System for a full 7-day week as an opt-in part of IPATCH-AD (NCT05175664) (for flow chart, see Fig. 1). The system consists of an accelerometer with a 12.5 Hz sampling rate that measures acceleration (-4-+4 *g*) in the *xyz*-coordinate system, a chronometer and a thermometer embedded in a 47 × 22 × 4.5 mm large unit incorporated in a medical adhesive patch (3M®). One sensor was applied on the sternum or below the mid part of the clavicula, to indicate upper body position, and one sensor was applied on the lateral thigh (10 cm proximal to the lateral femoral epichondylum) of the participant to capture movement.

Pre-processing

The SENS Motion system uses a proprietary algorithm for classifying activity, evaluated for each 5-second epoch. This algorithm has previously been shown to have high concordance with validated movement behavior in elderly patients [29]. The activity categories are (1) upright standing, (2) sporadic walking, 2) walking, 4) running (high intensity), 5) moderate intensity (brisk walking) 6) lying rest, 7) lying movement, 8) sitting. The exact definition of each activity category is described in the Supplementary Material.

The amount of each activity category over a 7-day wear week from which an actigram can be constructed depicting a daily activity pattern is illustrated in Fig. 2.

Feature extraction

The SENS software provides activity categories and activity count measures in 5.128-second, 15-minute, and 1-hour windows. The 15-minute resolution was chosen as a compromise between temporal resolution and variability of data.

These measures were subsequently analyzed in three time windows (24-hour period: 10 PM – 10 PM, night: 12 midnight – 6 AM, day: 9 AM – 9 PM). In these periods, the following features were calculated for each of the activity categories: (a) mean, (b) percentage of windows with above $x \in \{5,\ 10,\ 25\,\%\}$ of time spent in activity category, (c) intra-daily variability:

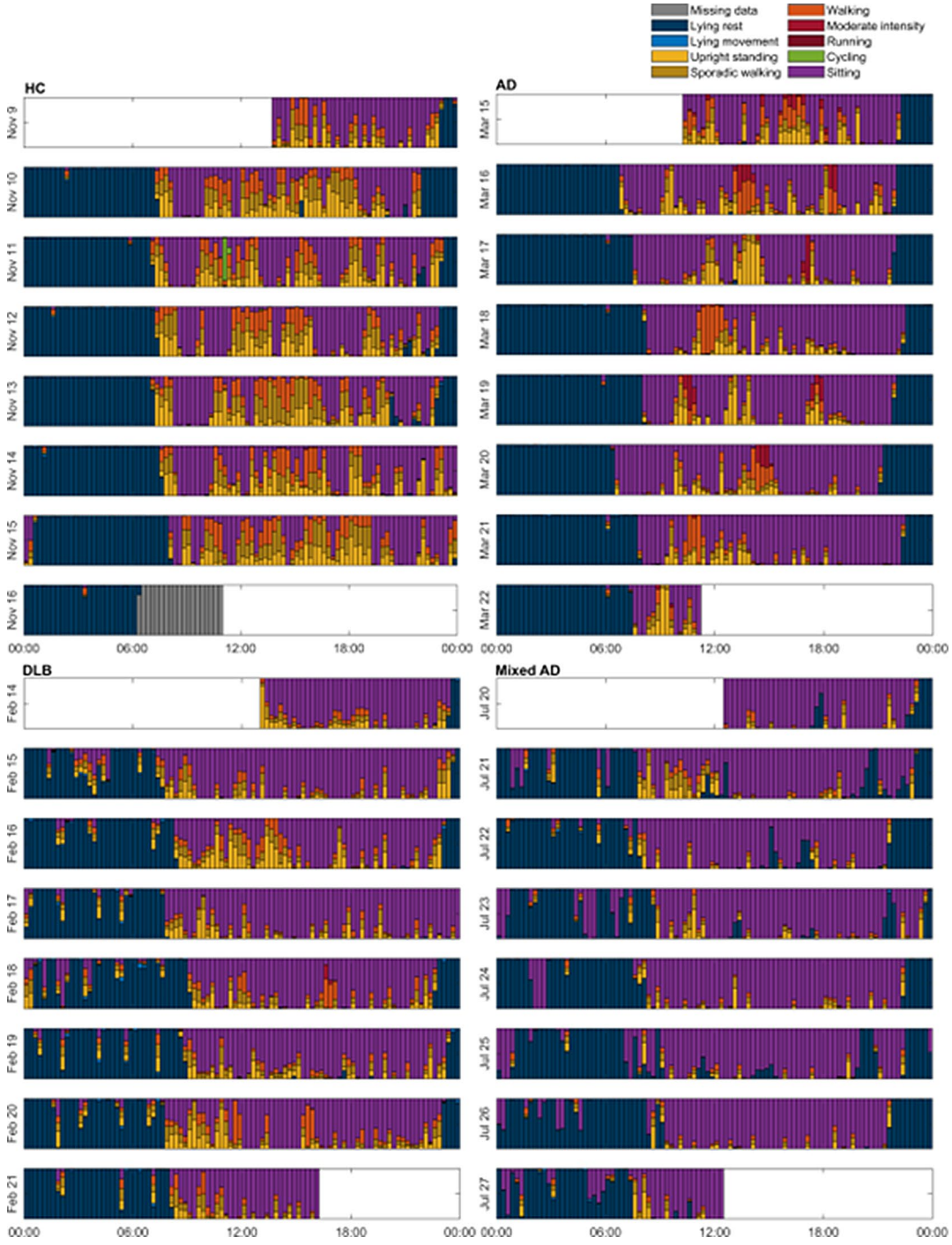


Fig. 2 Actigram from a healthy control (HC), a patient with Alzheimer's disease (AD), dementia with Lewy bodies (DLB), and mixed AD (+cerebrovascular pathology) showing the distribution of activity patterns during a 7-day wear period

$$IV = \frac{n \cdot \sum_{i=2}^{n} (x_i - x_{i-1})^2}{(n-1) \cdot \sum_{i=1}^{n} \left(\bar{x} - x_i\right)^2} \quad (1)$$

where x is the activity categories, $\bar{x}$ is the average of x, and n is the length of x. The activity count measures were summarized using (a) mean, (b) mean during sitting activity category, (c) activity during the 10 most active hours (+timing), (d) activity during the 5 least active hours (+timing), (e) relative amplitude: (M10 - L5) / (M10 + L5), (f) intra-daily variability (Eq. 1). In total, this adds up to 255 features that were computed for each 24-hour period.

To compute summary statistics, these were summarized using average and variance for all available 24-hour periods per subject resulting in 510 features. Moreover, for activity count measures, the average curve of each 24-hour cycle was computed for each diagnostic group. Summary statistics (mean and standard deviation) for all calculated features for each diagnostic group are available in the Supplementary Material.

Machine learning

Machine learning was carried out using logistic regression combined with sequential feature selection filtered by a minimum redundancy maximum relevance algorithm. All models were fitted and evaluated in a nested leave-one out cross-validation setup, and selected features were evaluated in the inner loop. No further regularization was used. As a comparison, a model using a single feature, the average intensity count during the 24-hour period was also used. As the mixed AD and vascular cognitive dysfunction groups were similar on several parameters, these were combined for the machine learning analyses. The evaluation metrics were sensitivity, specificity, accuracy, precision and the F1 score:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

The F1-score, which is the harmonic mean of precision and sensitivity, is commonly used to assess the performance of machine learning models. It penalizes extreme measures of either precision or sensitivity [30]. Feature importance was evaluated by examining the average absolute values of logistic regression coefficients.

Statistical analysis

Python, R (ver. 4.2.2) and MATLAB (R2019b, The MathWorks, Natick, Massachusetts, USA) were used for all analyses.

Quantitative variables are expressed by their mean and standard deviation. Variables were assessed for normality by visually inspecting histograms. Parametric and non-parametric tests were used to test group differences as appropriate. As actigraphic parameters exhibited non-normal distribution, a Kruskal-Wallis rank sum test was used to evaluate group differences. Performance confidence intervals were estimated using Clopper–Pearson's method. Performance for the logistic regression was evaluated at the default thresholds and at an optimized threshold by finding the point closest to (0,1) on the ROC curve. We imposed an alpha level of 0.05 and used only two-tailed tests.

Results

As shown in Table 1, a total of 122 patients and 48 healthy controls were included. There were significant differences between diagnostic groups for age, sex, and educational level, while no significant group differences were found for the use of antidepressant, antipsychotic, hypnotic, or anxiolytic medication (Table 1). Disease groups were older than the control group, and a higher proportion of patients with DLB were male. The majority of patients across disease groups had MCI or mild dementia. MMSE total scores were lower for dementia groups and lowest in the mixed dementia group. There were more patients with DLB or vascular cognitive dysfunction (mixed AD and pure vascular dementia) who had objective gait impairment. Likewise, the subjective walking distance was lower in dementia groups, with the lowest found for the mixed dementia group.

When we examined the overall proportion of the time spent performing each activity type there were significant between-group differences for the lying rest category, which was numerically slightly higher in disease groups (Table 2). When examining average intensity counts, as an overall measure of all activity types, these were highest for healthy controls (Fig. 3). Patients with dementia exhibited less walking activity, with the most pronounced differences between healthy controls and mixed dementia (Fig. 4). Significant between-group differences were found for relative amplitude reflecting the stability of the circadian rhythm, with disease groups showing a less robust circadian rhythm, although AD and healthy controls were similar on this parameter. A similar pattern was found for the intra-daily variability, indicating fragmentation of circadian rhythm, with a numerically more fragmented rhythm for disease groups, which again was less pronounced for the AD group. Averaging activity counts throughout the measurement period for each participant category revealed overall lower activity levels throughout the day with the lowest activity level for the mixed dementia group (Fig. 3).

The subjective maximal walking distance, an analog measure of functional capacity, followed activity levels closely (Table 1).

An in-depth examination of nighttime and daytime distribution of activity categories revealed only slight differences between groups (Supplementary Tables 1 and 2).

Diagnostic machine learning classifier

AD vs. healthy controls

When distinguishing between AD and healthy controls, the machine learning classifier, which used all 510 features, performed similarly to the single feature classifier using only the overall daily activity levels (accuracy: 66.1% vs. 68.8%, F1: 75% vs. 76.4%) (Fig. 5 and Supplementary Table 3). Both achieved an F1-score above 75%.

Table 1 Cohort characteristics

	Alzheimer's disease (*N* = 70)	Dementia with Lewy bodies (*N* = 29)	Mixed AD (+ CVD) (*N* = 8)	Vascular cognitive dysfunction (*N* = 15)	Healthy control (*N* = 48)	*P*-value
Age (years), mean (SD)	75.3 (7.1)	76.2 (6.4)	81.8 (3.3)	78.5 (5.9)	71.1 (7.8)	< 0.001[1]
Sex, n (%)						
Female	33 (47.1%)	4 (13.8%)	4 (50.0%)	5 (33.3%)	30 (62.5%)	< 0.001[2]
Male	37 (52.9%)	25 (86.2%)	4 (50.0%)	10 (66.7%)	18 (37.5%)	
MMSE total score, mean (SD)	25.5 (2.7)	26.2 (2.9)	22.6 (2.8)	26.5 (2.9)	29.1 (1.1)	< 0.001[3]
Educational level (years), mean (SD)	11.8 (3.7)	13 (4)	13.3 (4.7)	10.5 (2.9)	15 (2.5)	< 0.001[3]
Disease severity, n(%)						
Mild cognitive impairment	21 (30.0%)	3 (10.3%)	0 (0.0%)	8 (53.3%)	-	< 0.001[2]
Mild dementia	46 (65.7%)	25 (86.2%)	5 (62.5%)	4 (26.7%)	-	
Moderate dementia	3 (4.3%)	1 (3.4%)	3 (37.5%)	3 (20.0%)	-	
Objective gait impairment, n (%)						
No	63 (90.0%)	11 (37.9%)	4 (50.0%)	8 (53.3%)	45 (93.8%)	< 0.001[2]
Yes	7 (10.0%)	17 (58.6%)	4 (50.0%)	6 (40.0%)	3 (6.2%)	
Subjective maximum walking distance (meters), mean (SD)	6137 (3205)	4598 (3473)	1775 (1639)	3314 (3303)	7523 (3144)	< 0.001[3]
Hypnotic medication, n (%)						
Yes	1 (1.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0.697[2]
Previous stroke, n (%)						
Yes	2 (2.9%)	1 (3.4%)	2 (25.0%)	5 (33.3%)	0 (0.0%)	< 0.001[2]
Sedatives or analgesic medication, n (%)						
Yes	1 (1.4%)	2 (6.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0.126[2]
Antidepressant medicine, n (%)						
Yes	6 (8.6%)	4 (13.8%)	1 (12.5%)	3 (20.0%)	3 (6.2%)	0.429[2]
Antipsychotic medication, n (%)						
Yes	1 (1.4%)	1 (3.4%)	0 (0.0%)	1 (6.7%)	0 (0.0%)	0.519[2]
Antidementia medication, n (%)						
Yes	29 (41.4%)	22 (75.9%)	6 (75.0%)	0 (0.0%)	0 (0.0%)	< 0.001[2]

All variables had < 5 missing values. Frequency is calculated for participants without missing data

[1]ANOVA

[2]Fisher's Exact Test

[3]Kruskal-Wallis rank sum test

Sensitivity was similar for both models (85.7%) with slightly higher specificity for the single feature model (43.8%) when optimizing the discriminatory threshold, although the confidence intervals overlapped.

Healthy controls vs. disease groups

When we examined the predictive ability of actigraphy to distinguish between all disease groups (AD, DLB, CVD) and healthy controls, the model performed with an accuracy of 75.3%, a precision of 76.3%, and an F1-score of 84.7% (95% CI: 78.4–89.8) for the optimized threshold (Fig. 5 and Supplementary Table 3). Compared to the single feature classifier, sensitivity was higher (95.1% vs. 86.1%), while the other performance measures (accuracy, precision and F1-score) indicated no additional information when choosing from the full set of features (Fig. 5 and Supplementary Table 3).

AD vs. DLB

The machine learning classifier outperformed the single feature classifier on all performance measures, except for specificity, achieving > 80% accuracy in predicting AD vs. DLB using all available actigraphy features. The F1-score also indicated a better overall performance for the machine learning model, maybe due to the substantially higher sensitivity (62.1% vs. 24.1%) (Fig. 5 and Supplementary Table 4).

AD vs. mixed AD (+ vascular pathology) and VCD

The full-feature set machine learning model, which had a sensitivity of 69.6% and specificity of 95.7%, outperformed the single feature model on sensitivity and achieved an accuracy of 89% with an F1-score of 76.2%, for which the confidence intervals overlapped with that of the single feature classifier (Fig. 5 and Supplementary Table 4). Performance for the default threshold classifier is shown in Supplementary Fig. 1.

Table 2 The mean overall time spent in activity categories, the activity level and circadian indices for each diagnostic group

	Alzheimer's disease (*n* = 70)	Dementia with Lewy bodies (*n* = 29)	Mixed AD (+CVD) (*n* = 8)	Vascular cognitive dysfunction (*n* = 15)	Healthy control (*n* = 48)	*P*-value
Activity categories						
Lying rest, minutes/day	591 ± 102.7	625.4 ± 119.2	568.9 ± 101	633.1 ± 107.3	539.7 ± 88.8	**0.005**
Lying movement, minutes/day	5.7 ± 3.8	5.4 ± 3.6	6.4 ± 3.2	5.1 ± 3.4	6.3 ± 3.3	0.3
Upright standing, minutes/day	120 ± 53.6	116.8 ± 48.7	145.3 ± 65.3	89 ± 42.6	117.3 ± 40.1	0.05
Sporadic walking, minutes/day	116.3 ± 40	98.5 ± 37.3	89.4 ± 35.4	82.1 ± 52.2	108.1 ± 32.1	**0.002**
Running, minutes/week	3.1 ± 20.7	4.4 ± 15.5	0.3 ± 0.6	0.1 ± 0.3	3.8 ± 10.4	**0.047**
Cycling, minutes/week	17.6 ± 40.6	60.3 ± 99.8	1.8 ± 5.1	11.2 ± 18.5	42.6 ± 65.9	**0.001**
Sitting, minutes/day	489.8 ± 112.8	486.4 ± 110.3	560.9 ± 151.4	545.6 ± 138	528.2 ± 96.3	**0.01**
Sit-to-stand, number/day	61 ± 19.7	65 ± 25.1	50 ± 21.8	52 ± 13.9	62 ± 15.8	0.06
Activity level and circadian indices						
Overall intensity, intensity count	31.7 ± 10.6	29.1 ± 13.8	19.5 ± 10.6	21.4 ± 11.2	39.1 ± 12.2	**< 0.001**
Most active 10 h, intensity count	926 ± 331	832 ± 414.9	548.6 ± 356.7	602.0 ± 353.4	1135.4 ± 401.8	**< 0.001**
Least active five hours, intensity count	22.2 ± 13.6	36.0 ± 30.4	43.5 ± 11.3	34.9 ± 25.5	24.5 ± 12.7	**0.002**

Values are expressed as the mean ± standard deviation and are given as average number of minutes per day or week spent in an activity category as classified by the algorithm. Intensity counts are unitless. P-values from Kruskal-Wallis test. **Bold** indicates p-value below significance level

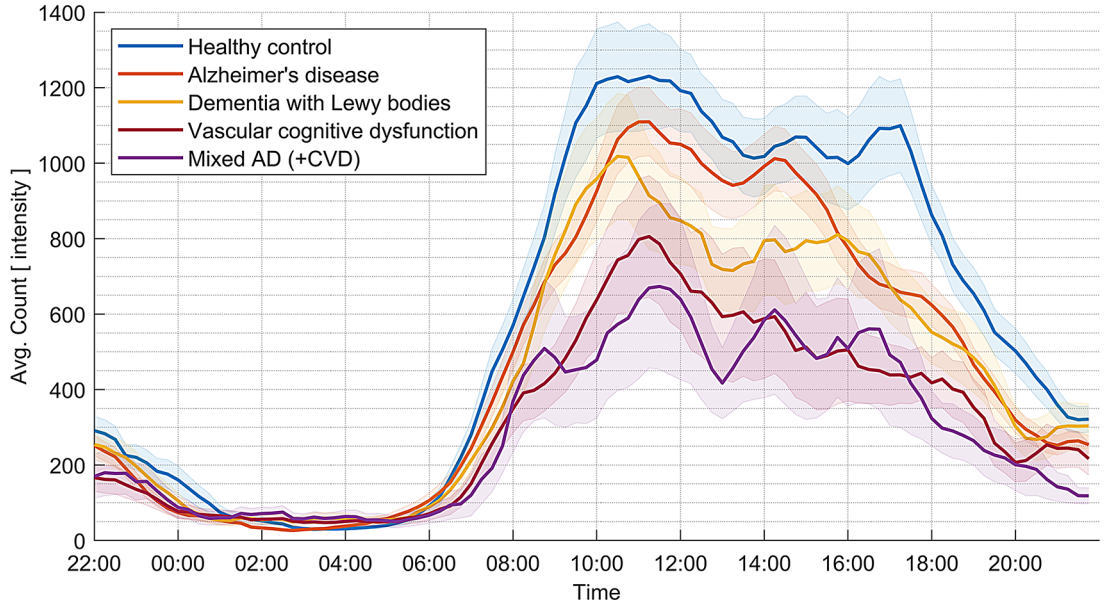


Fig. 3 Average intensity count during the 24-hour period for each diagnostic group. The curve is based on the 15-minute resolution and smoothed within the closest hour. Shaded areas represent the standard error of the mean

Feature importance

Feature importance plots for all machine learning classifiers are shown in Supplementary Fig. 2. The absolute regression coefficients were numerically high for the healthy vs. disease classification. This indicated that the machine learning model may have experienced overfitting or little discriminatory value in the variables, as was also shown in the forest plot. Nonetheless, nighttime activity was one of the most important features for classification. For the DLB vs. AD ML classifier, the robustness of the number of steps, as a proxy for activity, was most important, and the distribution for lying

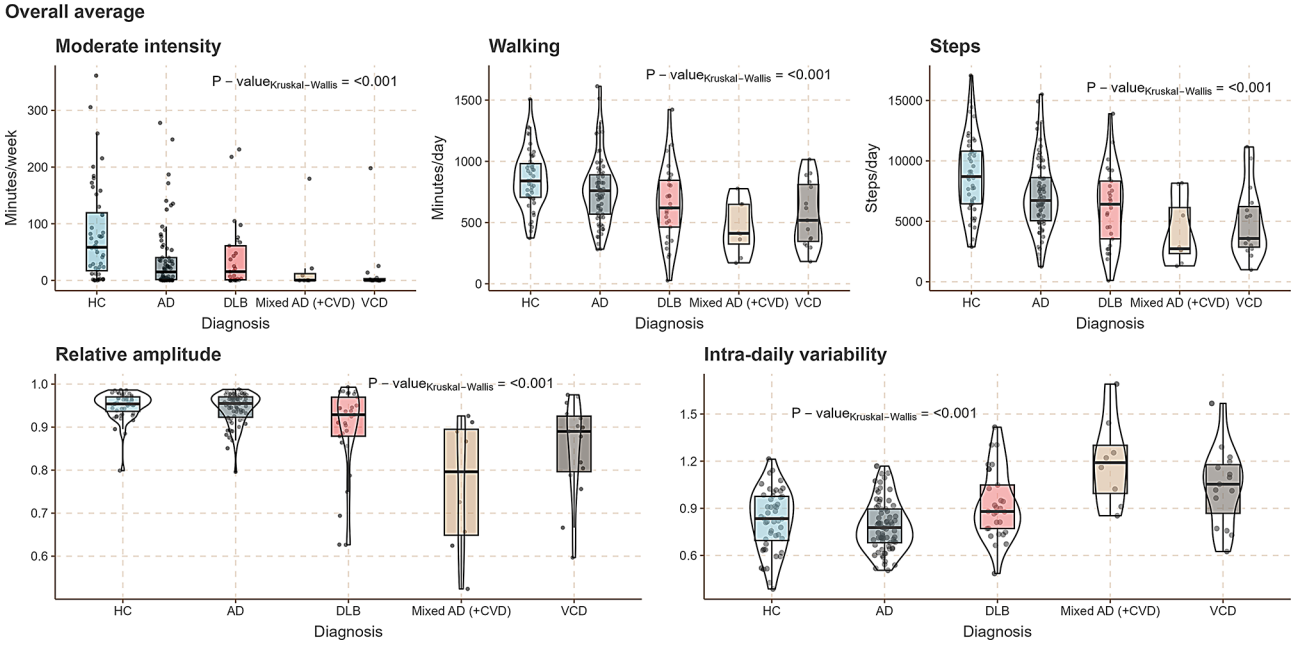


Fig. 4 Box- violin-, and dotplots showing the proportion of time spent in activity categories and circadian indices for all groups. Boxplots show the median and inter-quartile ranges (IQR), whiskers are 1,5 x IQR. HC = healthy control, AD = Alzheimer's disease, DLB = dementia with Lewy bodies, Mixed AD (+ CVD) = mixed Alzheimer's disease (+ cerebrovascular disease), VCD = vascular cognitive dysfunction. P-values represent between-group differences

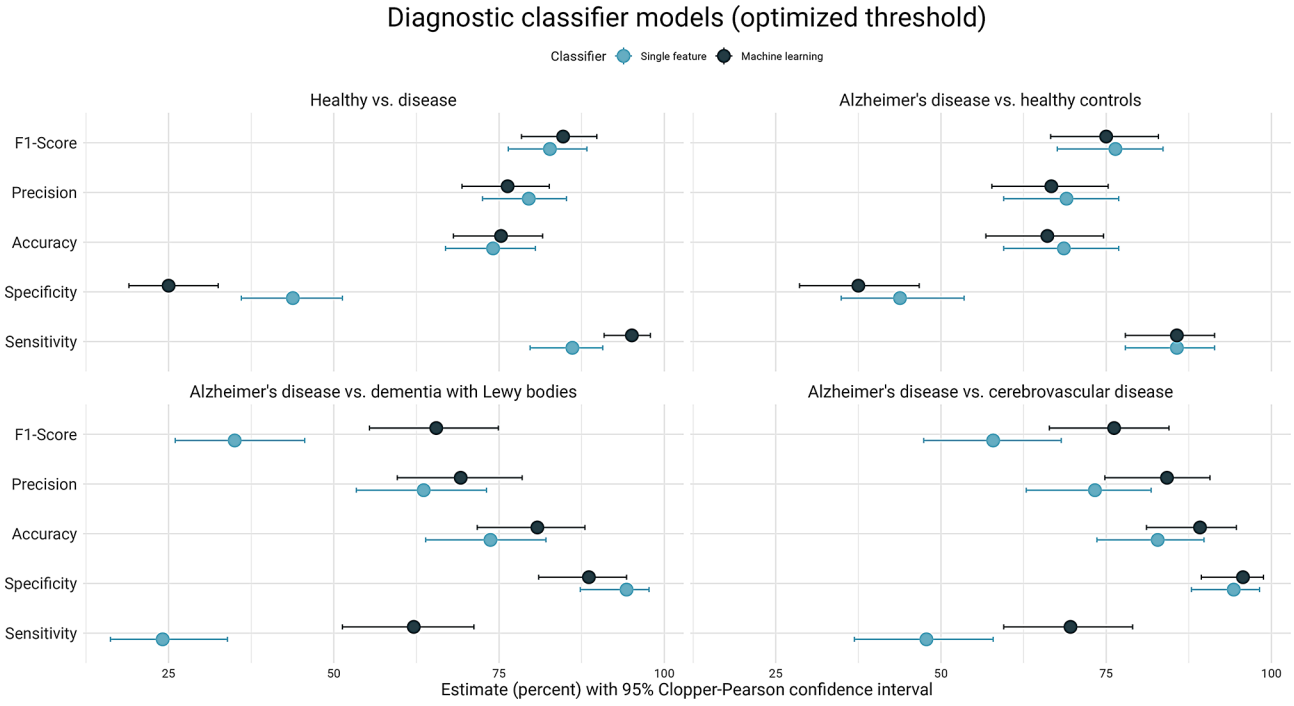


Fig. 5 Performance in differentiating groups in a leave-one-out scheme using either logistic regression (machine learning) or thresholding of average intensity count during the day (single feature). The probability threshold was optimized by finding the point closest to (0,1) on the ROC curve

rest showed high importance as well for this model. For the vascular vs. AD classification, the distribution and variability of walking activity as well as nighttime sitting activity were indicative of the disease class.

Output probability

We modeled the output probability of or ML models for healthy and dementia groups, separately, to test for possible confounders (age, sex, medication). The results of this analysis are shown in Supplementary Tables 5 and 6. This

showed no concern for confounding in the control group (no significant predictors), while in the dementia group, age and the use of sedatives or analgesic medication had a significant effect on the output probability, indicating possible confounding, although the estimates were small.

Discussion

We have evaluated the diagnostic performance of actigraphy coupled with an activity-type algorithm using machine learning to differentiate AD from aged, healthy controls and relevant disease comparators in a well-described memory clinic cohort. We generally found the highest diagnostic predictive ability with our machine learning approach, when the classifier was applied to differentiate between dementia etiologies, whereas this approach did not add to the differentiating ability when comparing overall disease groups or AD with healthy subjects. We found indications that the circadian rhythm was less robust and more fragmented in dementia groups when compared to aged, healthy controls, although the AD group was similar to the healthy group. The circadian phase was preserved across groups.

Interestingly, we found that combining actigraphy with a machine learning approach was most advantageous for differentiating dementia etiologies, as this approach achieved significantly higher sensitivity and higher accuracy when distinguishing both AD vs. DLB and AD vs. CVD compared to only using the overall daily mean activity level (single feature). To our knowledge, no other studies have evaluated actigraphically derived activity types for AD diagnosis. Utilizing the full scope of data, including nighttime periods, the machine learning approach performed seemingly better than single-feature classification, indicating that movement/activity patterns and subtle nighttime differences between groups might have improved the differentiation of etiologies. This was reflected by the most important features being fragmentation of walking activity, maybe indicating shorter walking bouts for the CVD group, and sitting activity during nighttime. The fragmentation of walking activity fits well with the literature describing a decrease in walking activity in stroke survivors [31], although the time distribution of walking has not been investigated previously. Although the existing literature is sparse on the comparison of AD to other dementias with actigraphy, a previous study reported no differences between AD and multi-infarct dementia [32], a finding we could not replicate, as we found that actigraphy was able to differentiate these etiologies with high accuracy. This may be explained by differences in our analysis method and two-sensor setup with a higher sampling frequency. Specifically, we applied an activity category algorithm, further classifying the movement patterns associated with each dementia etiology. Also, as diagnostic criteria for vascular cognitive dysfunction and AD have changed since the publication by Mishima et al., this may be reflected in the differences observed. Other groups have found associations between hallmarks of vascular cognitive dysfunction, such as higher levels of white matter hyperintensities and small vessel disease, and lower activity levels, which is in accordance with our results [33, 34].

The neuropathological underpinnings of a disturbed circadian rhythm, possibly also affecting activity levels, are mainly situated in the suprachiasmatic nucleus, where a lower number of neurons has been shown in patients with AD [35, 36]. Also, a recent study showed that hypopigmentation of the locus coeruleus, a key modifier of circadian function, was associated with cortical AD pathology at autopsy [37]. While we found lower activity levels in the AD groups relative to controls, we could not confirm the presence of a distinct circadian rhythm dysfunction in AD, which has frequently been reported in clinical [38–40] and pre-clinical stages [41], although not consistently [42]. Rather, patients with DLB and CVD exhibited a less robust and more fragmented circadian rhythm, which may have increased the discriminatory power of our models when differentiating dementia etiologies. Indeed, circadian fragmentation and short bouts of lying rest were one of the most important features in our models. While sleep disturbances are well-described in DLB, such as REM sleep behavior disorder, which is part of the diagnostic criteria for the disease [26], a distinct circadian dysfunction has not been characterized in larger studies [43]. Surprisingly, we did not detect differences in nighttime behavior as groups were largely similar during this period of recording. This suggests that thigh-worn actigraphy may not capture salient nighttime pathology in DLB, where for example RBD is highly prevalent [43]. To this end, wrist-worn actigraphy represents a more promising approach [21], which is also supported by the finding that REM sleep without atonia is more pronounced in the arms in comparison to the chin and legs [44]. In our study, we did not extract sleep metrics from the thigh-worn actigraphy due to limitations in the used algorithm. While we did not extract sleep metrics, our analysis of activity in the nighttime window may serve as a proxy for sleep-related behavior, although we cannot provide validation for this aspect. Future studies should evaluate sleep metrics alongside activity patterns to fully elucidate the discriminatory ability of thigh-worn actigraphy. Literature on the direct comparison of AD and DLB on actigraphic parameters measured in clinically diagnosed DLB is scarce, although post-mortem determined Lewy body co-pathology in AD patients has shown to be associated with circadian dysfunction [45]. Only one smaller study has looked at the discriminatory power of actigraphy in the distinction between AD and DLB [46], and our results are in accordance with these

findings, namely a larger degree of variability in nighttime to daytime activity reflecting less robustness of circadian rhythm in the DLB group, as was also reflected by the robustness of number of steps being the top feature for AD vs. DLB classification. As co-pathology with AD, especially cerebral deposition of amyloid, is prevalent in DLB [47] it is crucial to develop tools, besides fluid biomarkers, that can aid in discriminating these etiologies in the memory clinic. To this end, actigraphy may represent a promising approach, and coupling this approach with machine learning seems useful. As such, our results provide a proof-of-concept for the application of actigraphy in a memory clinic setting for differential diagnostic purposes.While we investigated the diagnostic capability of actigraphy at the cross-sectional level, there is also an avenue for assessing activity-related changes as a proxy for disease progression in longitudinal design. These analyses are planned for a follow-up sub-cohort of the present study (ClinicalTrials.gov ID NCT05175664).

Groups were mostly similar regarding medication use that could influence the circadian rhythm and activity levels, but significant differences were found for subjective walking distance and gait impairment. This means that some differences in actigraphic parameters could have been driven by impaired functional capacity in certain diagnostic groups and important diagnostic information can thus be surveyed by asking simple questions about subjective walking distance, although the study was not designed to investigate this aspect. The DLB and CVD groups were older than healthy controls and the AD group meaning that age-related functional decline and circadian changes [41] could have confounded the results for comparisons with these groups. This was also indicated by investigating output probability of our ML model, showing that age was a significant predictor in the dementia groups. We also did not find a significant circadian phase shift, meaning that patients kept the circadian rhythm intact, possibly by aligning with known entrainers (light, social activity, temperature, etc.) [48, 49].

We found slightly lower overall activity levels in AD compared to healthy controls, possibly relating to less time spent walking and doing moderate-intensity activities (e.g., brisk walking), which was also found in a recent study [18], and more time spent resting for AD patients. The higher mean overall activity levels detected in the healthy controls could be due to a sustained afternoon activity level, as healthy controls showed an activity peak in the late afternoon, a pattern that was absent in dementia groups. Although the mean overall activity levels as detected by the actigraph were able to differentiate the groups with some certainty, a machine learning approach may not be useful for the distinction of AD and healthy controls.

As such, the application of actigraphy in a memory clinic setting seems viable for differential diagnostic purposes: the technology can be applied early in the diagnostic process and could weave seamlessly into the clinic as an add-on to traditional biomarkers (cognitive testing, cerebrospinal fluid samping, and neuroimaging) as it provides information on the natural behavior and circadian rhythm outside the clinic. In addition, it is associated with low costs and a low patient burden, which is preferred due to an expected increase in referrals to memory clinics in part due to expected demographic changes with increasing cases of dementia and due to recently approved disease-modifying therapies for AD [20, 50]. Together, this makes actigraphy an interesting digital tool to be explored further in memory clinic settings; this could be for disease-tracking, remote monitoring of treatment response, or early screening purposes, which are beyond the scope of this paper. However, the application of a ML approach should warrant caution regarding how results are communicated to patients, given that the algorithms applied are not always easily interpretable. With our chosen ML model, feature importance can be examined, while models such as neural networks are less easy to interpret [51]. Improving interpretability of ML models is an ongoing task and one that will need special refinement when it comes to informing patients with cognitive impairment.

The strengths of our study are the large, well-described cohort of patients and the relevant setting of the memory clinic, where using actigraphs with a high sampling rate could be envisioned as part of a diagnostic work-up. Further, we applied machine learning, utilizing the full scope of data rather than focusing on single variables, and ensured a qualitative assessment of activity levels by applying an activity-type algorithm from two body-worn sensors. All but one patient were non-institutionalized, meaning that our results reflected the natural behavior of participants, which increases the generalizability of our findings.

Limitations also apply to our study. First, we applied a pragmatic definition of day- and nighttime periods, which may have influenced our results, however, changing the periods slightly did not alter our results. Second, our choice of sensor placement, namely thigh and sternum, could have hindered the detection of subtle nighttime motor activity reflecting REM sleep behavior disorder, which is prevalent in DLB, and may be more easily detected using wrist-worn actigraphy. On the other hand, the chosen sensor placement was essential for correct activity category classification and for accurately determining physical activity levels. Third, the study lacked external validation in an independent cohort, preferably patients evaluated before diagnosis, however, we provide unbiased estimates from cross-validated

analyses to adjudicate this. Fourth, relevant differential diagnoses such as MCI of unknown etiology or depression were not included. Fifth, the diagnostic groups were not balanced on demographics, which could have influenced our results, as age was a significant predictor of the output probability, albeit only for the dementia group. Sixth, we did not evaluate the added value of actigraphy to traditional biomarkers (e.g., cerebrospinal fluid, EEG, neuroimaging), however, assessing the stand-alone value of a novel biomarker seems crucial before conducting an add-on study design.

We have explored the clinical applicability of actigraphy in a memory clinic setting using a machine-learning approach. We found several actigraphic parameters that were altered in the dementia stage in accordance with recent literature. While our machine learning approach was less robust in differentiating healthy from disease, we did find that actigraphy was able to differentiate dementia etiologies with high accuracy, although part of the performance may have been driven by subtle age-differences. Actigraphy could therefore potentially be used for differential diagnostic purposes in memory clinics, provided age differences are taken into account. Future studies should aim to validate these findings in other memory clinic cohorts and evaluate actigraphy alongside traditional biomarkers.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13195-025-01751-5.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Funding for this project was received from A.P. Møller og Hustru Chastine Mc-Kinneys Fond til almene Formaal, Frimodt-Heineke Fonden, Vera og Carl Johan Michaelsens Legat, Alzheimer-forskningsfonden, Kong Christian den Tiendes Fond, and Ellen Mørchs Fond. Andreas Brink-Kjaer and Poul Jennum were supported by Innovation Foundation Denmark. The funders had no role in the design or reporting of the study.

Author contributions

Conceptualization: MG, FC, SH, KF, GW; Methodology: MG, FC, AB, SH, PJ, KF; Validation: MG, AB, FC, NS, SH, GW, PJ, KF; Formal analysis: AB, MG; Investigation: MG, FC, NS; Resources: GW; Writing - Original Draft: MG; Writing - Review & Editing: MG, AB, FC, NS, SH, GW, PJ, KF; Visualization: AB, MG; Supervision: KF, SH, GW, PJ; Project administration: KF, SH, GW; Funding acquisition: MG, FC, KF, SH, GW.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Human ethics statement and consent to participate

The Scientific Ethics Committees for the Capital Region of Denmark (file number: H-21045068) gave permission to conduct the study before it was initiated. The study was conducted in accordance with the Declaration of Helsinki. All participants gave informed written and oral consent to participate.

Competing interests

Kristian Steen Frederiksen is Editor-In-Chief at Alzheimer's Research and Therapy.

Author details

[1]Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
[2]Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark
[3]Department of Clinical Neurophysiology, Danish Center for Sleep Medicine, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
[4]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[5]Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital – Rigshospitalet, Inge Lehmans Vej 8, Copenhagen DK-2100, Denmark

Received: 17 December 2024 / Accepted: 29 April 2025
Published online: 21 May 2025

References

1. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. NPJ Digit Med. 2019;2(1):9.
2. Gustavsson A, Norton N, Fast T, Frölich L, Georges J, Holzapfel D, et al. Global estimates on the number of persons across the Alzheimer's disease continuum. Alzheimer's Dement. 2023;19(2):658–70.
3. Sims JR, Zimmer JA, Evans CD, Lu M, Ardayfio P, Sparks JD, et al. Donanemab in early symptomatic alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. JAMA. 2023;330(6):512–27.
4. van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M et al. Lecanemab in Early Alzheimer's Disease. New England Journal of Medicine [Internet]. 2023;388(1):9–21. Available from: http://www.nejm.org/doi/https://doi.org/10.1056/NEJMoa2212948
5. Martin JL, Hakim AD. Wrist actigraphy. Chest [Internet]. 2011 Jun 1 [cited 2024 Oct 1];139(6):1514–27. Available from: /pmc/articles/PMC3109647/.
6. Kuhlmei A, Walther B, Becker T, Müller U, Nikolaus T. Actigraphic daytime activity is reduced in patients with cognitive impairment and apathy. Eur Psychiatry. 2013;28(2):94–7.
7. David R, Mulin E, Friedman L, Le Duff F, Cygankiewicz E, Deschaux O, et al. Decreased daytime motor activity associated with apathy in alzheimer disease: an actigraphic study. Am J Geriatric Psychiatry. 2012;20(9):806–14.
8. Musiek ES, Xiong DD, Holtzman DM. Sleep, circadian rhythms, and the pathogenesis of Alzheimer Disease. Exp Mol Med [Internet]. 2015 [cited 2021 Oct 26];47(3):e148. Available from: http://www.pmc/articles/PMC4351409/
9. Brown BM, Peiffer JJ, Sohrabi HR, Mondal A, Gupta VB, Rainey-Smith SR et al. Intense physical activity is associated with cognitive performance in the elderly. Transl Psychiatry [Internet]. 2012;2(11):e191-6. Available from: https://doi.org/10.1038/tp.2012.118
10. Ju YES, McLeland JS, Toedebusch CD, Xiong C, Fagan AM, Duntley SP et al. Sleep quality and preclinical Alzheimer disease. JAMA Neurol [Internet]. 2013;70(5):587–93. Available from: https://doi.org/10.1016/j.earlhumdev.2015.09.003%5Cnhttp://dx.doi.org/10.1016/j.earlhumdev.2014.01.002%5Cnhttp://dx.doi.org/10.1016/S0378-3782(12)70006-3%5Cnhttp://www.sciencedirect.com/science/article/pii/S2341287914000763%5Cn.
11. Gramkow MH, Waldemar G, Frederiksen KS. The Digitized Memory Clinic. Nat Rev Neurol [Internet].:1–9. Available from: https://doi.org/10.1038/s41582-024-01033-y
12. Cañabate P, Martínez G, Rosende-Roca M, Moreno M, Preckler S, Valero S, et al. Social representation of dementia: an analysis of 5,792 consecutive cases evaluated in a memory clinic. J Alzheimer's Disease. 2017;58(4):1099–108.
13. Gibson LL, Abdelnour C, Chong J, Ballard C, Aarsland D. Clinical trials in dementia with lewy bodies: the evolving concept of co-pathologies, patient selection and biomarkers. Curr Opin Neurol. 2023;36(4):264–75.
14. James BD, Boyle PA, Bennett DA, Buchman AS. Total daily activity measured with actigraphy and motor function in community-dwelling older persons with and without dementia. Alzheimer Dis Assoc Disord [Internet]. 2012;26(3):238–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21946015

15. Musiek ES, Bhimasani M, Zangrilli MA, Morris JC, Holtzman DM, Ju YESES. Circadian Rest-Activity pattern changes in aging and preclinical alzheimer disease. JAMA Neurol. 2018;75(5):582–90.
16. Leng Y, Musiek ES, Hu K, Cappuccio FP, Yaffe K. Association between circadian rhythms and neurodegenerative diseases. Lancet Neurol [Internet]. 2019 Mar 1 [cited 2021 Oct 13];18(3):307–18. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1474442218304617
17. Buchman AS, Boyle PA, Yu L, Shah RC, Wilson RS, Bennett DA. Total daily physical activity and the risk of AD and cognitive decline in older adults. Neurol [Internet]. 2012;78(17):1323–9. Available from: NS -.
18. Varma VR, Watts A. Daily Physical Activity Patterns During the Early Stage of Alzheimer's Disease. Journal of Alzheimer's Disease [Internet]. 2016;55(2):659–67. Available from: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JAD-160582
19. Camargos EF, Louzada FM, Nóbrega OT. Wrist actigraphy for measuring sleep in intervention studies with Alzheimer's disease patients: Application, usefulness, and challenges. Sleep Med Rev [Internet]. 2013;17(6):475–88. Available from: https://doi.org/10.1016/j.smrv.2013.01.006
20. Gramkow MH, Waldemar G, Frederiksen KS. The Digitized Memory Clinic. Nat Rev Neurol [Internet]. 2024; Available from: https://doi.org/10.1038/s41582-024-01033-y
21. Brink-Kjaer A, Gupta N, Marin E, Zitser J, Sum-Ping O, Hekmat A, et al. Ambulatory detection of isolated Rapid-Eye-Movement sleep behavior disorder combining actigraphy and questionnaire. Mov Disord. 2023;38(1):82–91.
22. Gramkow MH, Clemmensen FK, Sjælland NS, Waldemar G, Hasselbalch SG, Frederiksen KS. Diagnostic performance of light reflex pupillometry in Alzheimer's disease. Alzheimer's Dementia: Diagnosis Assess Disease Monit. 2024;16(3):1–11.
23. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):270–9.
24. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr., Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):263–9.
25. Sachdev P, Kalaria R, O'Brien J, Skoog I, Alladi S, Black SE, et al. Diagnostic criteria for vascular cognitive disorders: A VASCOG statement. Alzheimer Dis Assoc Disord. 2014;28(3):206–18.
26. McKeith IG, Boeve BF, Dickson DW, Halliday G, Taylor JP, Weintraub D, et al. Diagnosis and management of dementia with lewy bodies: fourth consensus report of the DLB consortium. Neurology. 2017;89(1):88–100.
27. McKeith IG, Ferman TJ, Thomas AJ, Blanc F, Boeve BF, Fujishiro H, et al. Research criteria for the diagnosis of prodromal dementia with lewy bodies. Neurology. 2020;94(17):743–55.
28. Folstein MF, Folstein SE, McHugh PR. Mini-mental State. A practical method for grading the cognitive State.of patients for the clinician. J Psychiatr Res. 1975;12(3):189–98.
29. Ghaffari A, Rahbek O, Lauritsen REK, Kappel A, Kold S, Rasmussen J. Criterion Validity of Linear Accelerations Measured with Low-Sampling-Frequency Accelerometers during Overground Walking in Elderly Patients with Knee Osteoarthritis. Sensors [Internet]. 2022 Jul 15 [cited 2024 Oct 1];22(14):5289. Available from: https://www.mdpi.com/1424-8220/22/14/5289/htm
30. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep [Internet]. 2022;12(1):1–9. Available from: https://doi.org/10.1038/s41598-022-09954-8
31. Handlery R, Regan EW, Stewart JC, Pellegrini C, Monroe C, Hainline G et al. Predictors of daily steps at one year post stroke: A secondary analysis of a randomized controlled trial. Stroke [Internet]. 2021 May 1 [cited 2025 Mar 13];52(5):1768. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8085080/
32. Mishima K, Okawa M, Satoh K, Shimizu T, Hozumi S, Hishikawa Y. Different manifestations of circadian rhythms in senile dementia of Alzheimer's type and multi-infarct dementia. Neurobiol Aging. 1997;18(1):105–9.
33. Fleischman DA, Yang J, Arfanakis K, Arvanitakis Z, Leurgans SE, Turner AD, et al. Physical activity, motor function, and white matter hyperintensity burden in healthy older adults. Neurology. 2015;84(13):1294–300.
34. Sommer R, Yu L, Schneider JA, Bennett DA, Buchman AS, Lim ASP. Disrupted Rest-Activity Rhythms and Cerebral Small Vessel Disease Pathology in Older Adults. Stroke [Internet]. 2021;52(7):2427–31. Available from: https://www.ahajournals.org/doi/https://doi.org/10.1161/STROKEAHA.120.030870
35. Wang JL, Lim AS, Chiang WY, Hsieh WH, Lo MT, Schneider JA et al. Suprachiasmatic neuron numbers and rest-activity circadian rhythms in older humans. Ann Neurol [Internet]. 2015;78(2):317–22. Available from: https://onlinelibrary.wiley.com/doi/https://doi.org/10.1002/ana.24432
36. Yaffe K, Falvey CM, Hoang T. Connections between sleep and cognition in older adults. Lancet Neurol. 2014;13(10):1017–28.
37. Van Egroo M, van Someren EJW, Grinberg LT, Bennett DA, Jacobs HIL. Associations of 24-Hour Rest-Activity rhythm fragmentation, cognitive decline, and postmortem locus coeruleus hypopigmentation in Alzheimer's disease. Ann Neurol. 2024;95(4):653–64.
38. Hatfield CF, Herbert J, Van Someren EJW, Hodges JR, Hastings MH. Disrupted daily activity/rest cycles in relation to daily cortisol rhythms of home-dwelling patients with early Alzheimer's dementia. Brain [Internet]. 2004;127(5):1061–74. Available from: https://academic.oup.com/brain/article-lookup/doi/https://doi.org/10.1093/brain/awh129
39. Mulin E, Zeitzer JM, Friedman L, Le Duff F, Yesavage J, Robert PH, et al. Relationship between apathy and sleep disturbance in mild and moderate Alzheimer's disease: an actigraphic study. J Alzheimer's Disease. 2011;25(1):85–91.
40. La Morgia C, Ross-Cisneros FN, Koronyo Y, Hannibal J, Gallassi R, Cantalupo G et al. Melanopsin retinal ganglion cell loss in Alzheimer disease. Ann Neurol [Internet]. 2016 Jan 1 [cited 2021 Oct 7];79(1):90–109. Available from: https://pubmed.ncbi.nlm.nih.gov/26505992/
41. Musiek ES, Bhimasani M, Zangrilli MA, Morris JC, Holtzman DM, Ju YES. Circadian Rest-Activity Pattern Changes in Aging and Preclinical Alzheimer Disease. JAMA Neurol [Internet]. 2018;75(5):582–90. Available from: https://doi.org/10.1001/jamaneurol.2017.4719
42. La Morgia C, Mitolo M, Romagnoli M, Stanzani Maserati M, Evangelisti S, De Matteis M, et al. Multimodal investigation of melanopsin retinal ganglion cells in Alzheimer's disease. Ann Clin Transl Neurol. 2023;10(6):918–32.
43. Elder GJ, Lazar AS, Alfonso-Miller P, Taylor JP. Sleep disturbances in lewy body dementia: A systematic review. Int J Geriatr Psychiatry. 2022;37(10).
44. Röthenbacher A, Cesari M, Doppler CEJ, Okkels N, Willemsen N, Sembowski N et al. RBDtector: an open-source software to detect REM sleep without atonia according to visual scoring criteria. Sci Rep [Internet]. 2022;12(1):1–14. Available from: https://doi.org/10.1038/s41598-022-25163-9
45. Harper DG, Stopa EG, McKee AC, Satlin A, Fish D, Volicer L. Dementia severity and lewy bodies affect circadian rhythms in alzheimer disease. Neurobiol Aging. 2004;25(6):771–81.
46. Brown K, Shutes-David A, Payne S, Jankowski A, Seto E, Wilson K, et al. Actigraphy and sleep-monitoring devices to help differentiate dementia with lewy bodies from Alzheimer's disease. Alzheimer's Dement. 2022;18(S6):1–3.
47. Pichet Binette A, Mammana A, Wisse L, Rossi M, Strandberg O, Smith R et al. Associations between misfolded alpha-synuclein aggregates and Alzheimer's disease pathology in vivo. Alzheimer's and Dementia [Internet]. 2024; Available from: http://www.ncbi.nlm.nih.gov/pubmed/39258841
48. Canazei M, Papousek I, Weiss EM. Light intervention effects on circadian activity rhythm parameters and nighttime sleep in dementia assessed by wrist actigraphy: A systematic review and Meta-Analysis. Gerontologist. 2022;62(10):E614–28.
49. Pavlova M. Circadian Rhythm Sleep-Wake Disorders. Continuum (Minneap Minn) [Internet]. 2017;23(4, Sleep Neurology):1051–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28777176
50. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res. 2019;21(11):e16273.
51. Xu Q, Xie W, Liao B, Hu C, Qin L, Yang Z et al. Interpretability of Clinical Decision Support Systems Based on Artificial Intelligence from Technological and Medical Perspective: A Systematic Review. J Healthc Eng [Internet]. 2023 [cited 2025 Mar 12];2023:9919269. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9918364/

Publisher's note

STUDY IV

Prognostic value of light reflex pupillometry in Alzheimer's disease – a longitudinal cohort study

Gramkow, M. H., Clemmensen, F. K., Lindberg, U., Law, I., Henriksen, O. M., Waldemar, G., Hasselbalch, S. G., & Frederiksen, K. S.: Prognostic value of light reflex pupillometry in Alzheimer's disease – a longitudinal cohort study

(**Accepted** June 22, 2025, in Alzheimer's Research and Therapy)

Accepted June 22, 2025, in Alzheimer's Research and Therapy

Prognostic value of light reflex pupillometry in Alzheimer's disease – a longitudinal cohort study

Authors: Mathias Holsey Gramkow[1], Frederikke Kragh Clemmensen[1], Ulrich Lindberg[2,3], Ian Law[3,4], Otto Mølby Henriksen[3], Gunhild Waldemar[1,4], Steen Gregers Hasselbalch[1,4], Kristian Steen Frederiksen[1,4]

Affiliations

[1]Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital - Rigshospitalet, Denmark. [2]Functional Imaging Unit, Department of Clinical Physiology and Nuclear Medicine, Copenhagen University Hospital - Rigshospitalet, Denmark. [3]Department of Clinical Physiology and Nuclear Medicine, Copenhagen University Hospital - Rigshospitalet, Denmark. [4]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

Running title

Pupillometry in Alzheimer's disease

Corresponding author

Mathias Holsey Gramkow, Danish Dementia Research Centre, Department of Neurology, Copenhagen University Hospital – Rigshospitalet, Inge Lehmanns Vej 8, Copenhagen, DK-2100, Denmark.

Telephone No. (+45) 3545 5201. Email address: mathias.holsey.gramkow@regionh.dk.

Abstract

Background: The pupillary light reflex (PLR) has been indicated as a biomarker in Alzheimer's disease and may be suitable for easy prognostication. We sought to evaluate the prognostic potential of quantitative light reflex pupillometry (qLRP) in early AD.

Methods: At baseline, 3-months, 12-months and at 18/24-months follow-up (FU), we carried out qLRP with a hand-held pupillometer (PLR-3000, NeurOptics®). We assessed clinically evaluated progression, Mini Mental-State Examination (MMSE), visual progression on [^{18}F]FDG-PET and Clinical Dementia Rating Sum-of-Boxes (CDR-SoB) at 1-year FU. Logistic and linear regression models were fitted with baseline qLRP and the short-term dynamic qLRP change from baseline visit to 3 months as predictors with adjustment for age and sex. We evaluated logistic regression models by the cross-validated area under the receiver operating curve (AUC).

Results: A decrease in resting pupillary diameter was associated with a higher risk of clinically evaluated progression (odds ratio 4.3, 95 % confidence interval (CI): 1.2-16.9) and predicted this outcome with an associated AUC of 0.65. Less relative pupillary change after light stimulus measured at baseline was associated with cognitive decline on MMSE (β = -5.1, 95 % CI: -1.6 – -8.6, p=0.004) and a decrease in this variable from baseline–3 months could predict visual progression of [^{18}F]FDG-PET (AUC 0.63). There were no significant associations between qLRP and changes in the CDR-SoB, although estimates were in the same direction.

Discussion: qLRP holds promise as a prognostic, bedside, digital biomarker in Alzheimer's disease, possibly reflecting changes in the arousal state as a disease progression marker. Our results await confirmation in larger cohorts.

Keywords: prognosis, quantitative pupillometry, Alzheimer's disease, bedside prognostication

Introduction

There exist no approved prognostic markers for Alzheimer's disease (AD) and current candidates may be too cumbersome for patients to undergo or are not available outside highly specialized centers[1]. The disease course of AD is highly variable[2] and patients and physicians are thus left without important information at the early stages. This prognostic information could potentially be used for advance care planning, treatment selection and for directing clinical follow-up. As disease-modifying treatments are becoming available as treatment options[3,4], markers to inform treatment decisions are urgently needed both from a patient and clinician perspective[5]. Digital biomarkers such as light reflex pupillometry offer unobtrusive continuous monitoring of symptoms or features related to AD pathology and may hold prognostic information[6]. Quantitative light reflex pupillometry (qLRP) is a digital bedside biomarker, measuring the static and dynamic properties of the pupillary action in response to brief light stimuli[7], and has recently demonstrated promising diagnostic capability in AD[8]. The marker is shown to change longitudinally in patients harboring amyloid[9], but whether it holds prognostic information is not known[10].

Cross-sectional studies[9,11–14] have demonstrated that in clinical AD, the pupillary constriction is slower and pupil size is smaller compared to healthy aged-matched controls. This has been taken as evidence that the pupillary deficits are caused in part by parasympathetic failure[14]. Key regulators of the pupillary response include the Edinger Westphal nucleus (EWn)[15], locus coeruleus (LC)[16], and the dorsal raphe nucleus[17], which are located in proximity to each other in the mesencephalon. These nuclei all show neuronal changes related to AD pathology, and specific brain stem affection in Alzheimer's disease could be the cause of an altered pupillary response[18–21]. There has been an increasing focus on the role of pathology in the LC in Alzheimer's disease as a predilection site for neurofibrillary tangles predating cortical deposition[21–23]. Moreover, significant associations between LC changes and cognitive decline have been found in AD[24] Elevated metabolism of the locus

coeruleus is associated with resilience to cognitive decline[25], indicating that pupillary function may be used to probe this aspect. In addition, studies that target the pupil dilation response under cognitive paradigms have also shown an association with amnestic mild cognitive impairment and polygenic risk scores for AD[26,27]. Peinkhofer et al.[28] have reviewed the cortical modulation of pupillary function and underscored the influence of several cortical regions, such as the insular cortex, frontal eye field and the prefrontal cortex. McDougal & Gamlin[15] provide an extensive review on the autonomic control of the eye, highlighting the pivotal role of the parasympathetic pathway in regulating the pupil. A review by Chougule et al.[14] likewise posited that parasympathetic dysfunction was the main explanation for the "sluggish" constriction of the pupil observed in several studies in patients with AD. Together, this makes an interesting case for exploring the pupillary response as a prognostic biomarker in Alzheimer's disease. Studies investigating pupillometry in cohorts of early AD, representing a population potentially eligible for anti-amyloid treatment, are essentially missing from the literature and represent a current gap in knowledge. In this regard, both baseline measurement and short-term (months) dynamic changes in the pupillary response could be indicative of future decline, with the latter perhaps holding more information related to the disease course.

In this exploratory, prospective longitudinal study we aimed to investigate the prognostic value of baseline and short-term dynamic changes detected by quantitative light reflex pupillometry in a well-characterized cohort of early AD patients evaluated in a tertiary university memory clinic setting. We hypothesized that light reflex pupillometry is prognostic in early AD and that short-term dynamic pupillary changes can predict clinically evaluated progression within the first year after diagnosis and is predictive of progression of neurodegenerative changes as observed on [^{18}F]-fluorodeoxyglucose position emission tomography ([^{18}F]FDG-PET). As this was an exploratory study, our results can only be interpreted as hypothesis-generating.

Methods and Participants

Design

Prospective, longitudinal cohort study.

Participants

We recruited patients from a tertiary university memory clinic from January 2022 to June 2023. Diagnostic assessment at our center is comprehensive and includes examination by a physician and nurse. All patients undergo cognitive testing with two routine screening tests (Mini-Mental State Examination and Addenbrooke's Cognitive Examination), a structural scan (computed tomography or magnetic resonance imaging), and routine blood sampling (e.g., vitamin B_{12} status). State-of-the-art neuroimaging techniques to ascertain various aspects of neurodegenerative pathology (e.g., dopamine agonist tracer scans, amyloid PET), cerebrospinal fluid biomarker analysis, genetic counseling, and neuropsychological testing, are used if deemed necessary. Patients are diagnosed following a multi-disciplinary (neurology, geriatrics, psychiatry, neuropsychology, nursing) consensus conference. The inclusion criteria were: 1) a diagnosis of MCI [29] or mild-moderate dementia due to AD [30] with and without confirmation of amyloid positivity, 2) a caregiver willing to participate as an informant, 3) a Mini Mental-State Examination total score higher than 19, 4) a baseline diagnostic hybrid brain [^{18}F]FDG-PET/magnetic resonance imaging (MRI) or [^{18}F]FDG-PET/computed tomography (CT), and 5) ability to cooperate to the investigations. The exclusion criteria were: 1) Current alcohol or substance abuse, 2) other major neurologic or psychiatric disorder, or 3) active participation in drug trials. This study was part of a main study (TRACK-AD) that investigated blood-based biomarkers and digital biomarkers (actigraphy and pupillometry). The TRACK-AD study was registered at clinicaltrials.gov (Clinical Trial No.: NCT05175664, registration date: 4th of January 2022). The main study was designed to both investigate prognostic and disease-tracking markers in early AD. We followed the tenets of the 1975 Helsinki Declaration.

Procedure

Patients were included at the time of their diagnosis and followed for 18-24 months with scheduled visits at baseline, 3 months, 12 months, and 18/24 months (see study flow chart and overview in Supplementary Figures 1 and 2). At baseline and follow-up visits, a clinical evaluation, pupillometry, actigraphy, blood sampling and cognitive testing as well as Clinical Dementia Rating (CDR) were performed (see Supplementary Material for study overview). As a screening procedure, all patients underwent ophthalmoscopy to check for obscurities of the lens and vitreous body before inclusion in the present study. The disease severity was established at each visit and classified as MCI due to AD, or mild, moderate or severe dementia according to the NIA-AA criteria for MCI due to AD[29] and the International Statistical Classification of Diseases and Related Health Problems 10th revision criteria for dementia[31]. For the present study, we focused on the immediate prognostic information of light reflex pupillometry (within one-year follow-up), in line with a previous study on neurodegenerative biomarkers[32]. All study procedures were performed on the same day, except follow-up neuroimaging which was performed approx. two weeks prior to the 1-year visit.

Demographic data

Data on demographic variables (e.g., age, sex, educational level), and relevant comorbidities such as eye disease were extracted from the electronic health record. We also obtained a history from the caregivers and participants of any relevant eye disease and inspected medical files for prescription medication for eye-related disorders. Medication that could influence the pupil was noted according to Kelbsch et al[33] and included both eye medication (e.g., glaucoma, allergy, other) and systemic treatment (e.g., opioids, dopaminergic agents, cholinergic medication). Eleven patients were excluded post-hoc in the present study due to severe eye disease (e.g., severe glaucoma, cataract or retinopathy) due to interference with pupillometry. The presence of severe eye disease was evaluated by the investigator (M.G.).

Clinically evaluated progression

The primary outcome was clinically evaluated progression as rated at the 1-year follow-up visit by an experienced dementia physician. The physician determined the disease course leading up to each of the follow-up visits as either progressive or stable. The clinician was not blinded to clinical information and light reflex pupillometry nor follow-up scans, but the decision was based on the sum of clinical information (clinical impression, cognitive testing, information provided by the informant), and the rating was not necessarily carried out by the same physician at each visit. This approach was similar to that applied in a recent study investigating prognosis in a mixed memory clinic cohort[32]. The short-term dynamic pupillary changes from baseline to 3-months FU were not calculated until after the study had completed.

Cognitive scales

Rating on the CDR scale, a semi-quantitative interview-based scale rating cognitive and functional decline[34], was carried out by an experienced rater as part of the clinical examination at baseline and at the 1-year FU visit. Only the CDR Sum-of-Boxes score (CDR-SoB) was used for analysis. The Mini-Mental State Examination (MMSE), a brief cognitive screening test[35], was administered by a nurse during the initial diagnostic work-up if this was within 4 months of the baseline visit. Otherwise, the investigator administered these tests at the baseline visit. At all subsequent visits, the investigators (M.G. and F.K.C) administered the cognitive tests and rating scales.

[^{18}F]FDG-PET

All [^{18}F]FDG-PET images were obtained on clinical PET/CT or PET/MRI systems according to international standard operating procedure[36]. Baseline and follow-up scans were obtained on an identical system in 73 % of cases (all Siemens PET/MRI) or comparable systems (baseline Siemens PET/CT and follow-up PET/MR) in the rest of cases, except for two cases with baseline scan performed on a GE PET/CT system and follow-up on Siemens PET/MR. Scans were assessed using

statistical surface projections (Siemens Syngo.via, MI neurology, Erlangen, Germany) rated by two experienced nuclear medicine physicians (OMH, IL) in a blinded fashion with a forced decision between progressive or stable cerebral metabolism as a proxy of neurodegeneration. As such, the outcome was visually noticeable neuroimaging disease progression. Projections of baseline and follow-up scans were presented side-by-side in random order to minimize forecast bias. Scans where randomization led to reverse chronological order and were rated as progressed initially, were changed to stable after unblinding of order. If there was disagreement between raters, a consensus was reached referencing the original [^{18}F]FDG-PET images fused to MRI or CT after unblinding the order.

Quantitative light reflex pupillometry

The procedure used for qLRP has been described previously[8], where baseline visit cross-sectional qLRP data from this cohort are also reported. Briefly, a hand-held research grade pupillometer (PLR-3000, NeurOptics®) was used with customized settings to measure the resting pupillary diameter and pupillary light reflex (PLR). qLRP was carried out by the investigators (M.G. and F.K.C) at the baseline and subsequent visits under ambient lighting in the examination rooms of the memory clinic. The following parameters are given by the pupillometer: Resting pupil diameter (before light stimulus, mm), peak constriction pupil diameter (mm), latency (time in seconds after stimulus till constriction occurs), delta (relative percentage change between baseline and peak constriction diameter), average and maximum constriction velocity (mm/s), average dilation velocity (mm/s), and T75 (time in seconds to reach 75 % of baseline value after light stimulus). A table summarizing the physiological interpretation of each parameter is included as Supplementary Table 1. The same procedure was applied at subsequent follow-up visits, ensuring similar ambient lighting within a predefined range (approx. 500-800 lux) using dimmer light switches and measurement of the light intensity near the head of the participant using the Light Meter app (using the most recent version) on an iPhone 12 or higher.

193 **Data quality**

194 The method for quality checking pupillometry data has been described previously[8]. Briefly,
195 recordings flagged by the pupillometer were visually inspected for blinking artifacts during the
196 constriction phase and discarded if present. In the present study, a total of 156 recordings (11 % of all
197 recordings) were flagged by the pupillometer and 99 recordings (7 % of all recordings) were discarded
198 after visual inspection.

199 **Statistical analysis**

200 All statistical analysis was carried out using R (ver. 4.4.2) with the integrated development
201 environment RStudio (ver. 2024.09.1)[37].

202 Quantitative variables are presented by their means and standard deviations or median and range as
203 appropriate. Normality was assessed by qq-plots and equality of variance was assessed with an equal
204 variance test. Linear models were fitted for CDR SoB and MMSE as the dependent variables and
205 each of the eight qLRP parameter (both entered as baseline measurements and short-term dynamic
206 changes) as the independent variables in both crude and adjusted models. We used the mean of the
207 first three recordings as the unit of analysis for all qLRP metrics. The short-term dynamic change was
208 calculated by subtracting the baseline measurement from the 3-month measurement. The adjustment
209 differed for models using baseline measurements (baseline models) and short-term dynamic changes
210 (dynamic change models) as predictors. Baseline models were adjusted for age, sex, and resting
211 pupillary diameter. Dynamic change models were additionally adjusted for the baseline measurement
212 of the included qLRP parameter to reflect relative short-term dynamic changes. Models that survived
213 the above-mentioned adjustment were further adjusted for medication that could affect the pupil and
214 the presence of mild eye disease, separately. The CDR-SoB was log-transformed using the natural
215 logarithm to alleviate violation of the linearity assumption. As such, coefficients from these models
216 should interpreted by 1 unit increase in the predictor variable corresponding to the percentage change

in the CDR SoB score. Logistic regression models were fitted with clinically evaluated progression and visual [^{18}F]FDG-PET progression as outcomes. The outcome was coded for 1-year clinically evaluated progression as progression occurring at either the 3-month or 1-year FU. Prediction error for the logistic regression models were evaluated by the area under the receiver operating characteristics curve (AUC) with an associated 95 % confidence interval using the DeLong method. To assess robustness of these estimates on unseen data, a five-fold, 10 times cross-validation was performed using the *caret* package in R for the best models as indicated by lowest Akaike information criterion (AIC) value identified using the *MuMIn* package in R, which tests all combinations of univariable models. Cut points were found by optimizing the Youden index and the sensitivity and specificity were reported with an associated 95 % CI calculated with bootstrap estimation using the *cutpointr* package in R. Relevant assumptions (linearity, normally distributed residuals, and absence of significant heteroscedasticity) were met for all linear and logistic regression models. Partial regression plots were constructed for significant predictors in the linear models showing the regression line with the associated 95 % prediction interval, keeping all other variables at the mean value and categorical variables to reference values. A p-value<0.05 was considered significant and only two-tailed tests were used. Due to the exploratory nature of the study, no adjustments for multiple comparisons were made, and our results are only interpreted as hypothesis-generating.

Results

In Table 1, patient characteristics for the included 86 patients are shown stratified according to clinically evaluated progression status at the 1-year follow-up visit. Patients were followed for a median of 1.66 years (inter-quartile range: 1.479-1.96 years). During the study period, there was a loss to follow-up of 17 patients and one patient changed diagnosis at the 3-month visit and was excluded from analysis (see study flow chart in Supplementary Figure 1).

Patients who progressed were significantly older (p=0.021). There was an equal distribution of the sexes and groups were balanced at baseline on the MMSE and CDR-SoB. No differences were observed for mild, chronic eye disease or medication that could influence pupillary function. There was an equal proportion of patients receiving treatment with a cholinergic drug at the 3-month visit. qLRP parameters were highly correlated, and baseline pupillary diameter at baseline was significantly correlated with all other qLRP variables (Figure 1). When stratifying patients cross-sectionally at baseline according to disease severity, there was a tendency towards a slowing of the pupillary response (p=0.021 for baseline maximum constriction velocity) with increasing disease severity. There was an increased variability, as appreciated visually, of the pupillary response in the mild dementia group (Figures 3 and 4).

Table 1. Cohort characteristics at the baseline visit stratified according to progression status at 1-year follow-up.

	Stable (N=39)	Progression (N=47)	Total (N=86)	p-value
Age (years)				**0.021**[1]
Mean (SD)	73.8 (7.4)	77.3 (6.4)	75.7 (7)	
Range	53.3 - 89.9	63.7 - 88.2	53.3 - 89.9	
Sex				0.709[2]
Male	19 (48.7%)	21 (44.7%)	40 (46.5%)	
Female	20 (51.3%)	26 (55.3%)	46 (53.5%)	
Educational level (years)				0.315[3]
Mean (SD)	11.2 (3.4)	12.2 (4)	11.7 (3.7)	
Range	7 - 20	7 - 20	7 - 20	
Disease severity				0.542[2]
MCI	10 (25.6%)	12 (25.5%)	22 (25.6%)	
Mild	28 (71.8%)	35 (74.5%)	63 (73.3%)	
Moderate	1 (2.6%)	0 (0.0%)	1 (1.2%)	
MMSE total score				0.758[1]
Mean (SD)	25.4 (2.6)	25.6 (2.4)	25.5 (2.5)	
Range	20 - 30	21 - 30	20 - 30	
Clinical Dementia Rating Sum-of-Boxes score				0.157[3]
Mean (SD)	2.5 (1.1)	3 (1.6)	2.8 (1.4)	
Range	0.5 - 5	1 - 9.5	0.5 - 9.5	
Mild, chronic eye disease				0.237[2]
No	28 (71.8%)	28 (59.6%)	56 (65.1%)	
Yes	11 (28.2%)	19 (40.4%)	30 (34.9%)	
Medication with pupil effect				0.133[4]
No	33 (84.6%)	45 (95.7%)	78 (90.7%)	
Yes	6 (15.4%)	2 (4.3%)	8 (9.3%)	
AChEI treatment at 3-month FU				0.265[2]
No	17 (43.6%)	15 (31.9%)	32 (37.2%)	
Yes	22 (56.4%)	32 (68.1%)	54 (62.8%)	
Visual progression on PET				
N missing	6	12		0.029[2]
No	21 (63.6%)	13 (37.1%)	34 (50 %)	
Yes	12 (36.4%)	22 (62.9%)	34 (50 %)	

[1]Unpaired T-test (Welch), [2]Pearson's χ^2-test,[3]Wilcoxon rank sum test, [4]Fisher's exact test. MMSE: Mini Mental-State Examination, ACE: Addenbrooke's Cognitive Examination, AChEI: acetylcholinesterase inhibitor. **Bold** indicates significance.

Prognostic models – clinically evaluated progression (1-year)

For short-term dynamic changes in qLRP parameters, several parameters (resting and peak constriction pupillary diameter, average and maximum constriction velocity) showed a statistically significant difference between stable and progressed patients (Figure 2). Progressors had smaller resting pupil sizes (p=0.018) and exhibited slower average constriction velocity (p=0.015) with a trend towards smaller delta values (p=0.054) (Figure 2), whereas baseline qLRP parameters were not significantly different between groups (Supplementary Table 3 and Supplementary Figure 3). Several short-term dynamic qLRP changes, such as resting and peak constriction pupillary diameter and the velocity of constriction, were univariately associated with progression at 1-year FU and could predict this outcome above chance level (Table 2 and Figure 5). Following adjustment, only decreases in the resting pupillary diameter remained associated with clinically evaluated progression (Table 2). The multivariable models had similar AIC values (data not shown), i.e., there was no additive value by combining predictors, and therefore the best model was chosen based on the variable surviving adjustment for age and sex. Dynamic changes in the resting pupillary diameter had an associated cross-validated AUC of 0.66 (SD: 0.13), a sensitivity of 98 % [95 % CI: 15-100 %], and a specificity of 31 % [95 % CI: 13-91 %] for predicting clinically evaluated progression. This present study reports a 1-year progression rate of 54%, which would in turn provide a negative predictive value of 93 % and a positive predictive value of 63% for predicting 1-year progression using short-term dynamic changes in the resting pupillary diameter.

Table 2. Logistic regression models with 1-year clinically evaluated progression (progressor vs. stable at 1-year follow-up) as the outcome and short-term dynamic changes (3-month Δ) as predictors. Odds ratios are displayed as reciprocal to indicate the association of a 1 unit decrease in the qLRP variables.

	Univariable		Adjusted	
	Odds ratio [95 % CI]	**p-value**	**Odds ratio [95 % CI]**	**p-value**
Resting pupil diameter (3-month Δ)	3.9 [1.3, 14]	**0.022**	4.28 [1.24, 16.9]	**0.028**[a]
Peak constriction pupil diameter (3-month Δ)	13.6 [1.3, 192.1]	**0.039**	3.02 [0.001, 70.6]	0.7
Delta, relative pupillary change (3-month Δ)	1.09 [0.99, 1.21]	0.078	1.03 [0.89, 1.22]	0.6
Latency (3-month Δ)	-	0.3	-	0.8
Average constriction velocity (3-month Δ)	3.62 [1.28, 11.4]	**0.020**	1.16 [0.09, 14.8]	>0.9
Maximum constriction velocity (3-month Δ)	2.5 [1.23, 5.59]	**0.016**	2.21 [0.34, 14.5]	0.4
Average dilation velocity (3-month Δ)	12.1 [0.7, 262.6]	0.10	2.53 [0.02, 303.7]	0.7
T75 (3-month Δ)	1.64 [0.86, 3.5]	0.2	1.99 [0.67, 6.65]	0.2

Adjustment was age, sex, baseline qLRP variable and change in pupil diameter from baseline-3-months (except for in [a]). The estimate for odds ratio and the 95 % CI for latency is not shown, as it was extremely high due to little variation in the variable. **Bold** indicates significance.

Clinical Dementia Rating Sum-of-Boxes

We evaluated the association between qLRP measurements at the baseline visit as well as 3-month dynamic changes with the annualized change in the CDR-SoB. The variance of the CDR-SoB score was large (see Supplementary Figure 4), and also larger than for MMSE (data not shown). We did not find any significant associations between qLRP and CDR-SoB (see Supplementary Tables 4 and

298 5). Nonetheless, the estimates were in a similar direction as those observed for clinically evaluated
299 progression.

300 **Cognitive decline – MMSE**

301 Baseline measurements of both peak constriction diameter and the delta variable were significantly
302 associated with cognitive decline on MMSE, also when adjusting for relevant confounders (Table 3
303 and Figure 6). For 3-month changes in qLRP, no statistically significant associations were found.
304 (Supplementary Table 6).

305 **Table 3.** Results of a linear regression model showing the association between MMSE annualized
306 change (1-year) and qLRP variables (baseline measurements). Adjustment was for age, sex and
307 resting pupil diameter, except for the model with this variable as the predictor.

	Univariable			Adjusted		
Characteristic	**Beta**	**95% CI**	**p-value**	**Beta**	**95% CI**	**p-value**
Resting pupil diameter (mm) (baseline)	0.45	-0.67, 1.6	0.4			
Peak constriction pupil diameter (mm) (baseline)	2.0	0.18, 3.8	**0.031**	5.1	1.6, 8.6	**0.004**
Delta, relative pupillary change (percent (baseline))	-0.11	-0.21, -0.01	**0.024**	-0.16	-0.26, -0.05	**0.004**
Latency (s) (baseline)	-1.8	-24, 20	0.9			
Average constriction velocity (mm/s) (baseline)	-0.20	-1.3, 0.88	0.7			
Maximum constriction velocity (mm/s) (baseline)	-0.04	-0.77, 0.69	>0.9			
Average dilation velocity (mm/s) (baseline)	1.8	-0.60, 4.1	0.14			
T75 (s) (baseline)	-0.53	-1.5, 0.39	0.3			

308

309

310

Visual [^{18}F]FDG-PET progression

We evaluated the predictive value of qLRP on progression as assessed on a visual [^{18}F]FDG-PET read, whereby visual changes as shown on a cortical projection image were detectable by two experienced nuclear medicine physicians. In general, the agreement between clinical and neuroimaging progression was poor, indicated by Cohen's $\kappa = 0.27$, however there was some overlap between clinical progressors and neuroimaging progressors (Table 1). Only short-term dynamic decreases in the delta qLRP variable (relative pupillary change)were predictive of progression on [^{18}F]FDG-PET (cross-validated AUC 0.64, 95 % CI 0.5-0.07), but this variable was not significantly associated with neuroimaging progression in the logistic regression models (see Figure 7 and Supplementary Tables 7 and 8). The associated sensitivity was 68 % [95 % CI: 25-100 %] and the specificity was 62 % [95 % CI: 17-83 %]. The directions of estimates for the logistic regression models were similar to clinically evaluated progression for short-term dynamic changes.

Confounder analysis

While no statistically significant differences were present at baseline between progressors and stable patients for medication that could affect the pupil and mild eye disease, there was a numerically lower proportion of progressors receiving these types of medications. There was also a numerically higher proportion of patients with mild eye disease in the progressor group. This could lead to confounding and we therefore performed a further multivariable adjustment including these variables. The analyses showed that the association between short-term dynamic changes in the resting pupillary diameter and clinically evaluated progression could be explained by adjusting for medication that could affect the pupil but not when only adjusting for eye disease (see Supplementary Table 9). Taking medication that could affect the pupil was not however associated with either baseline (Unpaired T-test p=0.697) nor short-term dynamic changes (Unpaired T-test p=0.101) in the resting state pupil. The associations

were unaltered and the statistical significance was intact when adjusting for these confounders in models with MMSE annualized change as the outcome (data not shown).

Discussion

In this study we evaluated the prognostic performance of qLRP in early Alzheimer's disease. Our results showed that short-term dynamic changes in the pupillary resting diameter were predictive of progression at 1-year follow-up using clinically evaluated progression as an outcome. This indicates the possible usefulness of qLRP as a bedside prognostic tool. This was further corroborated by associations with cognitive decline as measured with MMSE and cross-sectional differences in qLRP between disease stages, while significant associations with functional decline as measured on the CDR Sum-of-Boxes were not found, maybe due to high variance in the CDR-SoB. Short-term dynamic changes in the qLRP delta variable were predictive of progression on visual [^{18}F]FDG-PET read indicating that qLRP marker may also be used to predict global neurodegenerative changes as observed with neuroimaging.

We identified the resting pupillary diameter, a general measure of arousal[38], to be predictive of clinically evaluated progression in early Alzheimer's disease. As hinted previously, neuropathology of the brain stem in the dorsal raphe nucleus may explain why arousal and disease progression are connected in AD, as early pathology is present in the brain stem before cortical areas such as the transentorhinal cortex[20,39]. Further, brain areas associated with the presence of apathy, a marker of rapid cognitive decline[40], are involved in arousal systems[41]. These notions align with our main finding of the resting pupillary diameter as a marker of future progression. This is to our knowledge the first study to investigate qLRP as a bedside prognostic marker, and as such represents the first proof-of-concept for using pupillometry to gauge prognostic aspects in AD. Previous studies[9,42] have generally identified a slowing of the contracting phase after a light stimulus in AD cross-sectionally[14], although few studies adjusted for resting pupillary diameter and age of participants. In our study, constriction velocity was not associated with clinically evaluated progression after adjustment for resting pupillary

diameter, which seems appropriate, as we have previously shown the resting pupillary diameter to be strongly correlated with all pupillary metrics[8,43], indicating a hierarchical physiological system. Our adjustment for age was justified by the significant baseline correlations found between this variable and several qLRP parameters, which has also been demonstrated by others[44]. As the progressors were older on average, this was an important confounder to consider and indeed most estimates diminished following adjustment for this factor.

Generally, the changes observed on the group level were small and some of the patients in the stable group showed short-term increases in pupillary diameter, which we cannot explain from a physiological perspective. This could indicate that the marker exhibits more variability than previously shown[43], and should warrant some caution in the interpretation of our results. We found a high negative predictive value for short-term dynamic changes in the resting pupil diameter, providing the ability to reassure patients of stable disease course within a year. However, the positive predictive value was only slightly higher than the base rate of progression, indicating that it should not be used for indicating a high likelihood of progression. Indeed, the AUC was only moderate and the standard deviation for the cross-validated AUC was high, indicating low robustness of our findings. This could be owing to the either a small effect size or the moderate number of progressors, that was borderline for the conservative rule of thumb of 10 events per variable[45]. This was likewise reflected in the confounder analysis where adjustment for medication that could affect the pupil overturned the significance of the association, diminishing the belief of a true association. We also observed a tendency for patients at later disease stages to exhibit a higher variability of the marker at baseline, indicating that the constriction velocity may become unstable as the disease progresses. This suggests that qLRP may best reflect underlying disease progression in the very early (MCI) stages of AD, although we did not investigate this aspect due to our limited sample size. Future studies should investigate whether the volatility of qLRP measurements could serve as a stage marker for AD.

We showed that the qLRP variable delta was predictive of visual progression on PET, while we could not replicate the finding for resting pupillary diameter as shown for clinically evaluated progression. This suggests that the outcomes were not completely congruent, and indeed the agreement between clinical and neuroimaging progression in this study was poor. The more global patterns observed on visual [^{18}F]FDG-PET reads may not reflect subtle changes in strategic areas important for cognition and independent living and some patients show little impact on ADL function even when larger areas show cumulative hypometabolism, especially in earlier disease stages[46]. This could be mitigated by applying voxel-wise statistics, which we intend to do in future analyses.

While we observed significant differences in dynamic changes of qLRP metrics and disease stage, we could not establish an association with cognition at the baseline visit, which has previously been shown[42]. This could be explained by the early disease stage, as reflected by the higher mean MMSE, resulting in a small variation in the variable, as MMSE exhibits a ceiling effect[47]. This correlation should preferably be examined in a longitudinal set-up where changes in one variable could be correlated with changes in the other. This would increase power, which is lower for cross-sectional comparisons[48].

Our cohort resembles those of previous studies regarding the rate of progression (approx. 2 points decline per year on the CDR SoB)[49] and we confirm the larger variance usually observed in the CDR score rendering it more suitable for larger cohort studies[2]. We believe that the null findings for this outcome may be owing to the large variation which was confirmed by evaluating the variance of the CDR-SoB, which increased with each visit, more so than for MMSE. It seems that, as the disease progresses, the CDR takes on a larger variability, as we also observed for qLRP. The increasing variability of the CDR and the limited capability of MMSE to gauge progression[50] makes the case that investigating clinically evaluated progression as evaluated by a comprehensive clinical

evaluation is a valid construct that could supplement these investigations. This is not to say that clinically evaluated progression is without faults, as it could be driven by individual factors related to the examining clinician, which we did not characterize in the present study. Looking at several progression outcomes and thereby seeking to validate findings across outcomes could prove a useful method, as we also apply in this study. However, while we validate some findings across the clinical measures, for example regarding the delta variable, other findings could not be replicated limiting the validity of our results. While we do provide adequate adjustment for confounders, there is also a high likelihood of type I error due to the multiple outcomes and our choice of investigating all eight qLRP parameters. Choosing to investigate all eight qLRP metrics was a deliberate choice as no other studies have looked at these variables in a prognostic context. Thus, it naturally follows that our results can only be interpreted as hypothesis-generating.

A strength of our study is the well-characterized, prospective cohort with generalizable inclusion criteria. Also, as the cohort was at a very early stage of disease our results become more applicable in the future of disease-modifying treatment, where the focus is on early initiation to prevent decline. In addition, we provide cross-validated estimates of our prediction models to evaluate robustness and adjust models for relevant confounders such as mild eye disease and medication with a pupillary effect. Last, using a hand-held, clinically validated, and easily applicable tool, our results could be directly translated to clinical practice, however, validation studies and a better understanding of the underlying mechanism are needed, which could be gained through quantitative neuroimaging studies.

We acknowledge that our study is not without limitations. First, we experienced a small, but non-negligible, drop-out, which may have skewed our results and limited power to draw firm conclusions. Second, a complete ophthalmological examination beyond ophthalmoscopy was not done and this may have led to some patients not having relevant eye pathology reported. Although we adjust for mild eye disease, this could have influenced our results. We await the ongoing efforts to evaluate the

444 influence of ophthalmological comorbidities on pupillary responses in persons with dementia[51].
445 Third, we did not seek to validate our results in an independent cohort, however, using cross-
446 validation, we could provide unbiased estimates of predictive power. Fourth, as we tested multiple
447 outcomes, there may be a higher chance of spurious findings, which is why our results should be
448 replicated by others in larger cohorts. Fifth, there is a possibility of confirmation bias stemming from
449 the examiner not being blinded to pupillary read-outs, however, aggregated mean measures and short-
450 term dynamic changes were not calculated until after study completion to limit this bias. Sixth, as
451 showed with cross-validated analyses our prediction was not robust to unseen data, probably due to
452 a moderate sample size and a high variation in the parameters in question, which limits the external
453 validity of our findings.

454 In conclusion, we demonstrated the prognostic potential of qLRP as a bedside digital biomarker in a
455 well-characterized cohort with adequate follow-up. We demonstrated that the most promising marker
456 for predicting clinically evaluated progression is short-term dynamic changes in the resting pupillary
457 diameter, indicating that decreases in arousal levels could be used to predict clinically evaluated
458 progression in early Alzheimer's disease, however the marker was not robust to confounder analysis,
459 and some of the effects may have been driven by slight differences in the proportion of participants
460 taking medication that could influence the pupil. Nonetheless, qLRP seems promising in providing a
461 bedside measure that could be used, along with future prognostic markers, to inform patients and
462 possibly guide treatment selection in Alzheimer's disease. Future studies should evaluate
463 pupillometry in larger cohorts along with quantitative neuroimaging and clinical measures and with
464 adequate adjustment for confounders as well as limit comparisons to corroborate and validate our
465 findings before clinical implementation can be envisioned.

466

Acknowledgments

We would like to thank the patients and their caregivers/informants for participating in this study.

Funding Declaration

Funding for this study was received from A.P. Møller og Hustru Chastine Mc-Kinneys Fond til almene Formaal, Frimodt-Heineke Fonden, Vera og Carl Johan Michaelsens Legat, Alzheimer-forskningsfonden, Kong Christian den Tiendes Fond, and Ellen Mørchs Fond. The funders were not involved in the design or reporting of this study.

Conflicts of Interest Statement

Dr. Frederiksen serves as consultant and/or member of an advisory board for Novo Nordisk, Eisai/Bioarctic, Eli Lilly, Roche Diagnostics (Remuneration paid to institution), serves (or has served) as Principal or Sub-investigator and National Coordinator on several industry-sponsored phase 2 and 3 trials (indication Alzheimer´s disease): Roche, Roche Diagnostics, Biogen, NovoNordisk, Osuka, AbbVie (remuneration paid to institution), serves on the Scientific advisory board (and as lecturer) on the MiCog educational program (supported by an unrestricted grant from Nestlé to the European Geriatric Medicines Society, personal remuneration), speaker and educational activities for Roche, Roche Diagnostics, Eisai/Bioarctic, Eli Lilly, Novo Nordisk, Lundbeck A/S (2023, 2024, 2025) (remuneration paid to institution), BestPractice Nordic, Alzheimerforeningen, Lundbeckfonden, Folkeuniversitetet i Emdrup, Dagens medicin (2023, 2024, 2025) (personal remuneration), serves as Editor-in-Chief for Alzheimer´s Research & Therapy Springer – Nature (from 2024 – 2023-2024 as Associate Editor) (personal remuneration), receives royalties from publications with Springer-Nature and Hans Reitzels Forlag (outside this work). The remaining authors have no conflicts of interest.

Data Availability Statement

The dataset generated in the present study can not be shared publicly due to the Danish Data Protection Law. It is available upon reasonable request from the corresponding author.

Code Availability Statement

The code used to generate the results of this paper is available at https://www.github.com/Matgram/progqLRP.

Ethics Statement

Written informed consent was obtained from all participants before the commencement of the study, and approval (journal No. H-21040317) from the Scientific Ethics Committees for the Capital Region of Denmark was obtained prior to initiation of the study. The study was carried out in compliance with the 1975 Helsinki Declaration.

Author Contributions

Conceptualization: MG, FC, SH, KF, GW; Methodology: MG, FC, IL, OMH, SH, KF; Validation: MG, FC, IL, OMH, UL, SH, GW, KF; Data curation: MG, UL, OMH; Formal analysis: MG; Investigation: MG, FC, OMH, IL; Resources: GW; Writing - Original Draft: MG; Writing - Review & Editing: MG, FC, UL, OMH, IL, SH, GW, KF; Visualization: MG; Supervision: KF, SH, GW; Project administration: KF, SH, GW; Funding acquisition: MG, FC, KF, SH, GW

References

1. Garcia, M. J. *et al.* Prognostic and Predictive Factors in Early Alzheimer's Disease: A Systematic Review. *J Alzheimers Dis Rep* **8**, 203–240 (2024).

2. Barnes, J., Bartlett, J. W., Wolk, D. A., Van Der Flier, W. M. & Frost, C. Disease Course Varies According to Age and Symptom Length in Alzheimer's Disease. *Journal of Alzheimer's Disease* **64**, 631–642 (2018).

3. Sims, J. R. *et al.* Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA* **330**, 512–527 (2023).

4. van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine* **388**, 9–21 (2023).

5. Gramkow, M. H., Waldemar, G. & Frederiksen, K. S. The Digitized Memory Clinic. *Nat Rev Neurol* 1–9 (2024) doi:10.1038/s41582-024-01033-y.

6. Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *NPJ Digit Med* **2**, 9 (2019).

7. Taylor, W. R. *et al.* Quantitative pupillometry, a new technology: Normative data and preliminary observations in patients with acute head injury - Technical note. *J Neurosurg* **98**, 205–213 (2003).

8. Gramkow, M. H. *et al.* Diagnostic performance of light reflex pupillometry in Alzheimer's disease. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring* **16**, 1–11 (2024).

9. Frost, S. *et al.* Evaluation of Cholinergic Deficiency in Preclinical Alzheimer's Disease Using Pupillometry. *J Ophthalmol* **2017**, 1–8 (2017).

10. Troiani, V. The future of quantitative pupillometry in health and disease. *Clinical Autonomic Research* **30**, 11–12 (2020).

11. Fotiou, D. F. *et al.* Pupil reaction to light in Alzheimer's disease: Evaluation of pupil size changes and mobility. *Aging Clin Exp Res* **19**, 364–371 (2007).

12. Frost, S. *et al.* Pupil Response Biomarkers for Early Detection and Monitoring of Alzheimer's Disease. *Curr Alzheimer Res* **10**, 931–939 (2013).

13. Kawasaki, A., Ouanes, S., Crippa, S. V. & Popp, J. Early-Stage Alzheimer's Disease Does Not Alter Pupil Responses to Colored Light Stimuli. *Journal of Alzheimer's Disease* **75**, 1191–1201 (2020).

14. Chougule, P. S., Najjar, R. P., Finkelstein, M. T., Kandiah, N. & Milea, D. Light-induced pupillary responses in Alzheimer's disease. *Front Neurol* **10**, 1–12 (2019).

15. McDougal DH & Gamlin PD. Autonomic control of the eye. *Compr Physiol* **5**, 439–73 (2015).

16. Huang, R. & Clewett, D. The Locus Coeruleus: Where Cognitive and Emotional Processing Meet the Eye. *Modern Pupillometry* 3–75 (2024) doi:10.1007/978-3-031-54896-3_1.

17. Cazettes, F., Reato, D., Morais, J. P., Renart, A. & Mainen, Z. F. Phasic Activation of Dorsal Raphe Serotonergic Neurons Increases Pupil Size. *Current Biology* **31**, 192-197.e4 (2021).

18. Mavroudis, I. A. *et al.* Dendritic and spinal alterations of neurons from Edinger-Westphal nucleus in Alzheimer's disease. *Folia Neuropathol* **52**, 197–204 (2014).

19. Scinto, L. F. M. *et al.* Focal pathology in the Edinger-Westphal nucleus explains pupillary hypersensitivity in Alzheimer's disease. *Acta Neuropathol* **97**, 557–564 (1999).

20. Grinberg, L. T. *et al.* The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer̈s disease. A precocious onset? *Neuropathol Appl Neurobiol* **35**, 406–416 (2009).

21. Theofilas, P. *et al.* Locus coeruleus volume and cell population changes during Alzheimer's disease progression: A stereological study in human postmortem brains with potential implication for early-stage biomarker discovery. *Alzheimer's and Dementia* **13**, 236–246 (2017).

22. Beardmore, R., Hou, R., Darekar, A., Holmes, C. & Boche, D. The Locus Coeruleus in Aging and Alzheimer's Disease: A Postmortem and Brain Imaging Review. *Journal of Alzheimer's Disease* **83**, 5–22 (2021).

23. Beardmore, R. *et al.* Changes in the locus coeruleus during the course of Alzheimer's disease and their relationship to cortical pathology. *Neuropathol Appl Neurobiol* **50**, e12965 (2024).

24. Chen, Y., Chen, T. & Hou, R. Locus coeruleus in the pathogenesis of Alzheimer's disease: A systematic review. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **8**, 1–12 (2022).

25. Koops, E. A. *et al.* Elevated locus coeruleus metabolism provides resilience against cognitive decline in preclinical Alzheimer's disease. *Alzheimer's and Dementia* 1–14 (2024) doi:10.1002/alz.14385.

26. Granholm, E. L. *et al.* Pupillary Responses as a Biomarker of Early Risk for Alzheimer's Disease. *Journal of Alzheimer's Disease* **56**, 1419–1428 (2017).

27. Kremen, W. S. *et al.* Pupillary dilation responses as a midlife indicator of risk for Alzheimer's disease: association with Alzheimer's disease polygenic risk. *Neurobiol Aging* **83**, 114–121 (2019).

28. Peinkhofer, C., Knudsen, G. M., Moretti, R. & Kondziella, D. Cortical modulation of pupillary function: systematic review. *PeerJ* **7**, e6882 (2019).

29. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270–9 (2011).

30. McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging and the Alzheimer's Association workgroup. *Alzheimer's and Dementia* **7**, 263–9 (2011).

31. ICD-10. (2020).

32. Gramkow, M. H. *et al.* Prognostic value of complementary biomarkers of neurodegeneration in a mixed memory clinic cohort. *PeerJ* **2020**, e9498 (2020).

33. Kelbsch, C. *et al.* Standards in pupillography. *Front Neurol* **10**, (2019).

34. Hughes, C. P., Berg, L., Danziger, W., Coben, L. A. & Martin, R. L. A New Clinical Scale for the Staging of Dementia. *British Journal of Psychiatry* **140**, 566–572 (1982).

35. Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–198 (1975).

36. Guedj, E. *et al.* EANM procedure guidelines for brain PET imaging using [18F]FDG, version 3. *Eur J Nucl Med Mol Imaging* **49**, 632 (2021).

37. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.r-project.org/ (2021).

38. Mathôt, S. Pupillometry: Psychology, Physiology, and Function. *J Cogn* **1**, 16 (2018).

39. Singh, K. *et al.* Functional connectome of arousal and motor brainstem nuclei in living humans by 7 Tesla resting-state fMRI. *Neuroimage* **249**, 118865 (2022).

40. Lechowski, L. *et al.* Persistent apathy in Alzheimer's disease as an independent factor of rapid functional decline: the REAL longitudinal cohort study. *Int J Geriatr Psychiatry* **24**, 341–346 (2009).

41. Huey, E. D., Lee, S., Cheran, G., Grafman, J. & Devanand, D. P. Brain Regions Involved in Arousal and Reward Processing are Associated with Apathy in Alzheimer's Disease and Frontotemporal Dementia. *Journal of Alzheimer's Disease* **55**, 551–558 (2017).

42. Fotiou, D., Kaltsatou, A., Tsiptsios, D. & Nakou, M. Evaluation of the cholinergic hypothesis in Alzheimer's disease with neuropsychological methods. *Aging Clin Exp Res* **27**, 727–733 (2015).

43. Gramkow, M. H., Clemmensen, F. K., Waldemar, G., Hasselbalch, S. G. & Frederiksen, K. S. Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort. *J Neurol Sci* **456**, 122856 (2024).

44. Fotiou, D. F. *et al.* Effect of age on pupillary light reflex: Evaluation of pupil mobility for clinical practice and research. *Electromyogr Clin Neurophysiol* **47**, 11–22 (2007).

45. Austin, P. C. & Steyerberg, E. W. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res* **26**, 796 (2014).

46. Landau, S. M. *et al.* Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging* **32**, 1207–1218 (2011).

47. Franco-Marina, F. *et al.* The Mini-mental State Examination revisited: ceiling and floor effects after score adjustment for educational level in an aging Mexican population. *Int Psychogeriatr* **22**, 72–81 (2010).

48. Bakdash, J. Z. & Marusich, L. R. Repeated measures correlation. *Front Psychol* **8**, 1–27 (2017).

49. Williams, M. M., Storandt, M., Roe, C. M. & Morris, J. C. Progression of Alzheimer disease as measured by Clinical Dementia Rating sum of boxes scores. *Alzheimers Dement* **9**, S39 (2012).

50. Arevalo-Rodriguez, I. *et al.* Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* **2021**, CD010783 (2021).

51. Piano, M., Nilforooshan, R. & Evans, S. Binocular vision, visual function, and pupil dynamics in people living with dementia and their relation to the rate of cognitive decline and structural changes within the brain: Protocol for an observational study. *JMIR Res Protoc* **9**, (2020).

Figures

Figure 1. Correlogram showing Spearman's ρ for significant correlations, NS = not significant.

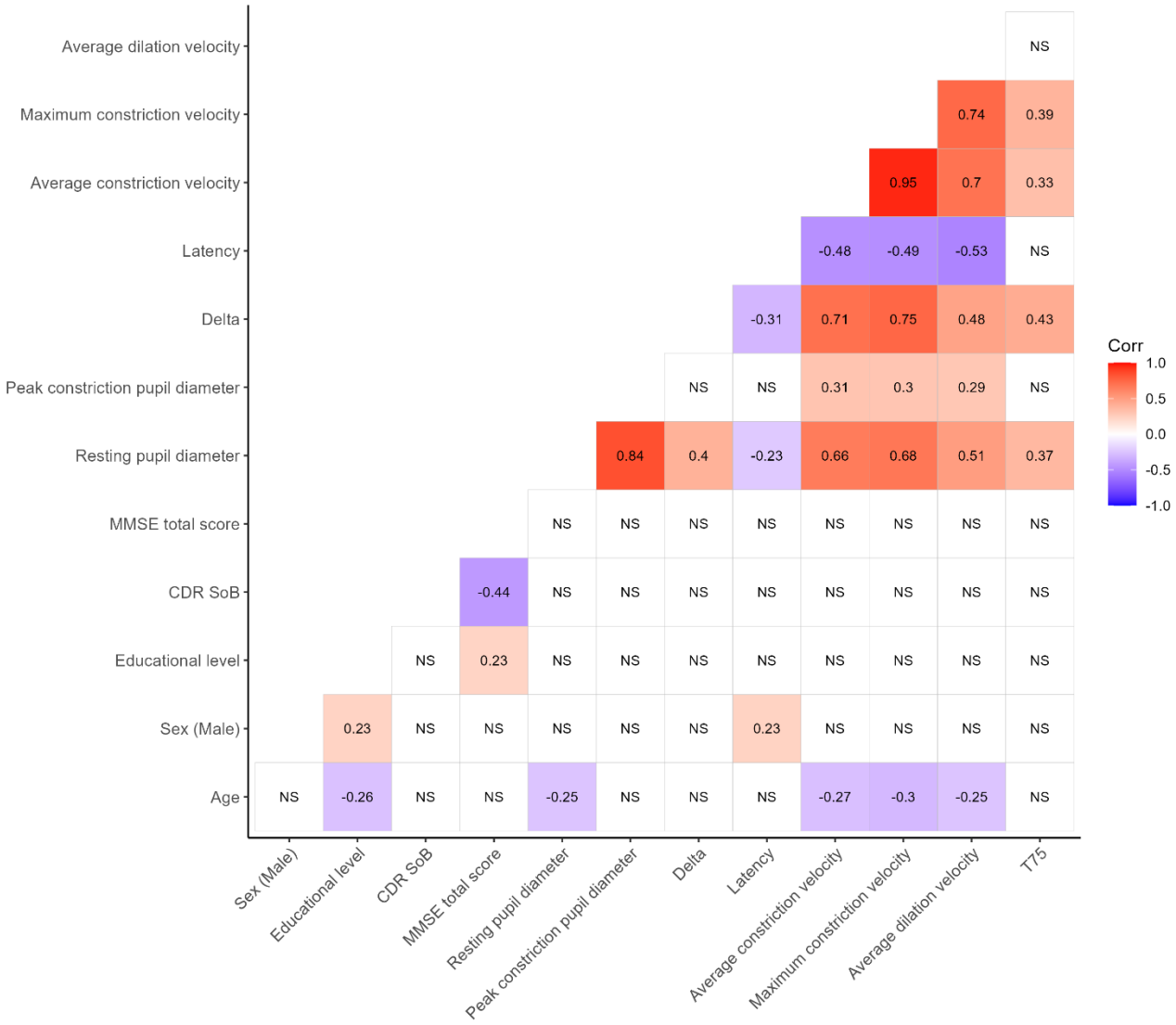


Figure 2. Short-term dynamic changes in pupillary parameters from baseline-3-month FU. P-values reflect the significance of differences between Progression and Stable groups. The dashed line represents the short-term variability that could be considered due to random fluctuations as based on a previous study[43].

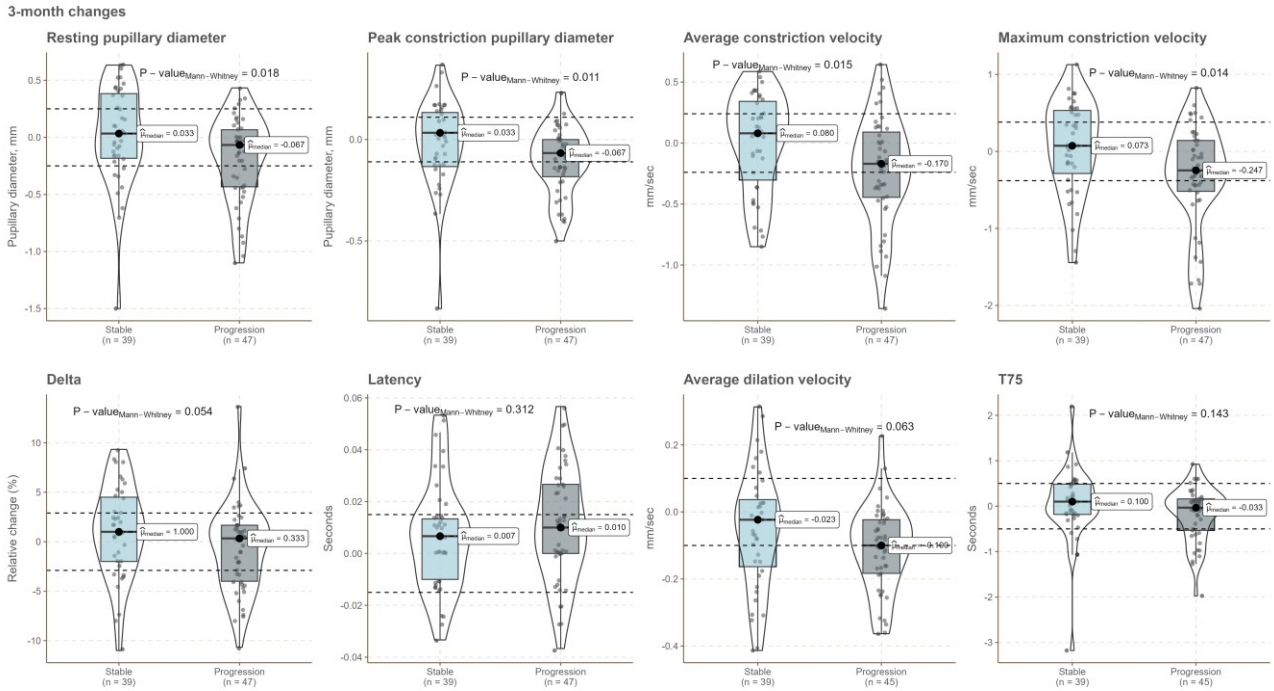


Figure 3. qLRP parameters measured at baseline stratified according to disease severity at baseline. One patient had moderate dementia at baseline (not shown in plot). MCI: mild cognitive impairment.

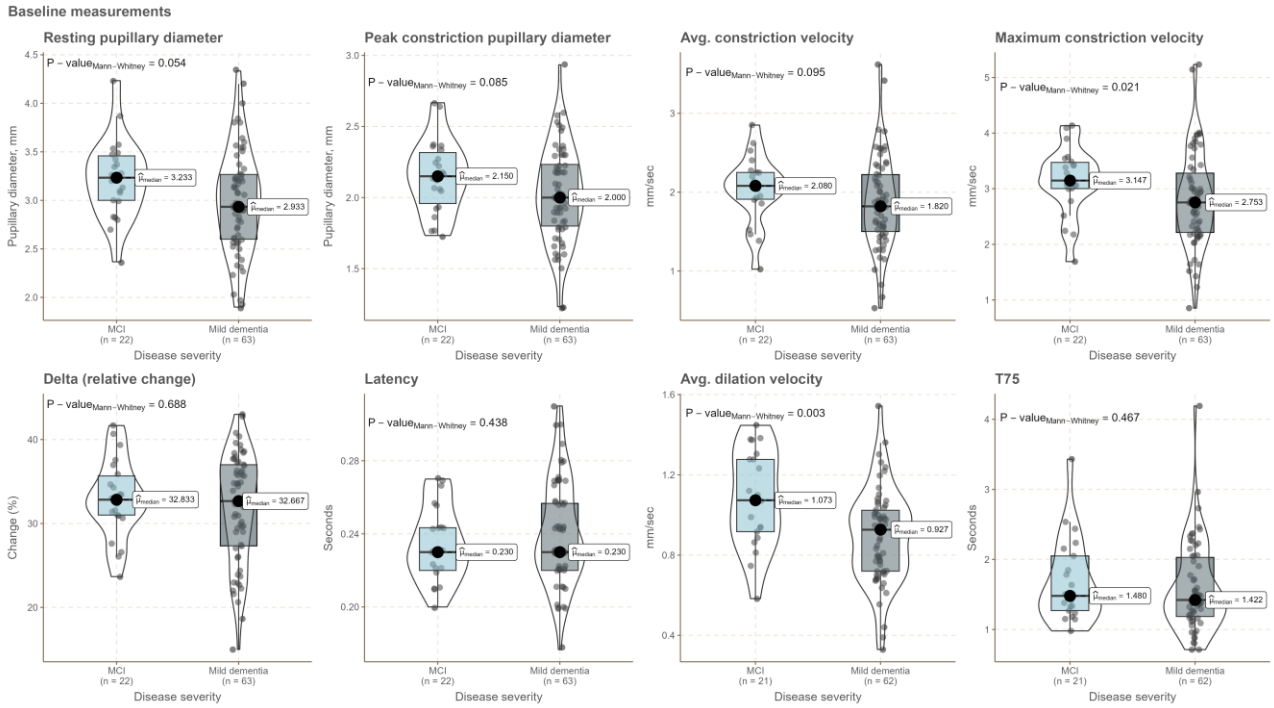


Figure 4. Short-term dynamic changes in qLRP parameters stratified according to disease severity. As only one patient had moderate dementia at baseline, this patient was omitted from the plot. MCI = Mild cognitive impairment.

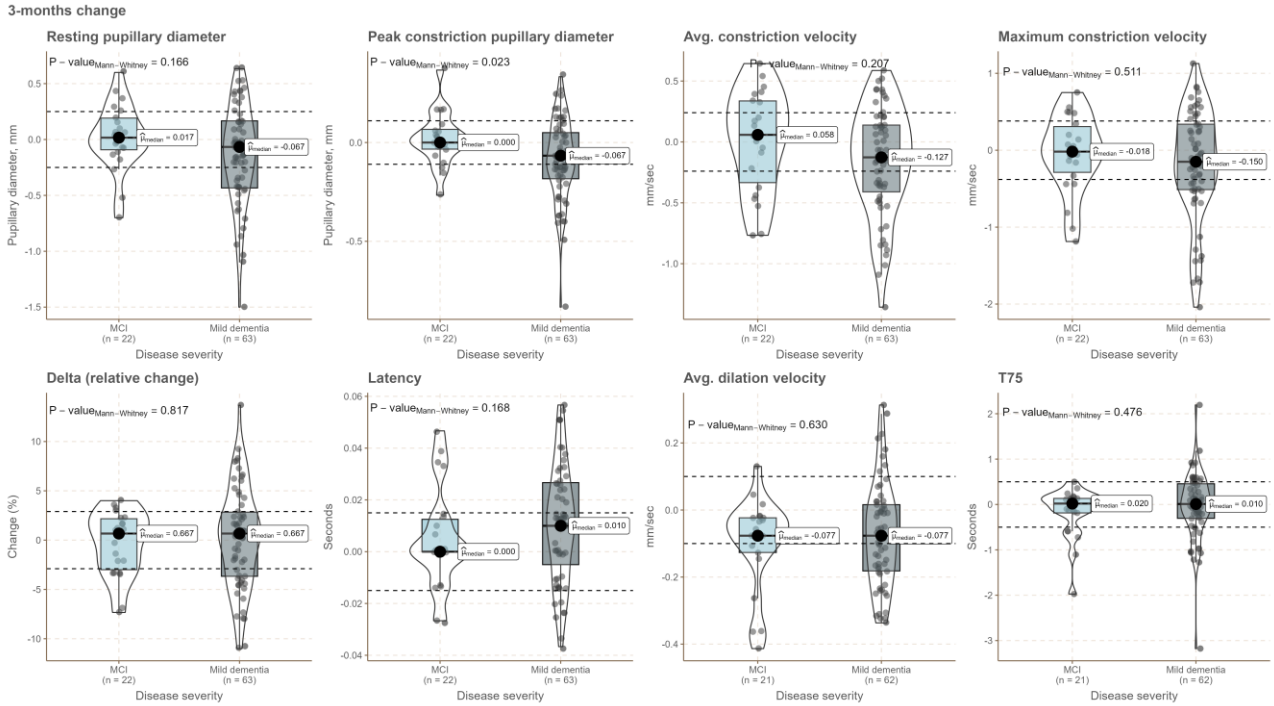


Figure 5. Model performance of individual short-term dynamic qLRP changes shown as receiver operating characteristics (ROC) curves for predicting clinically evaluated progression. The 95 % confidence interval (95 % CI) is calculated with the DeLong method.

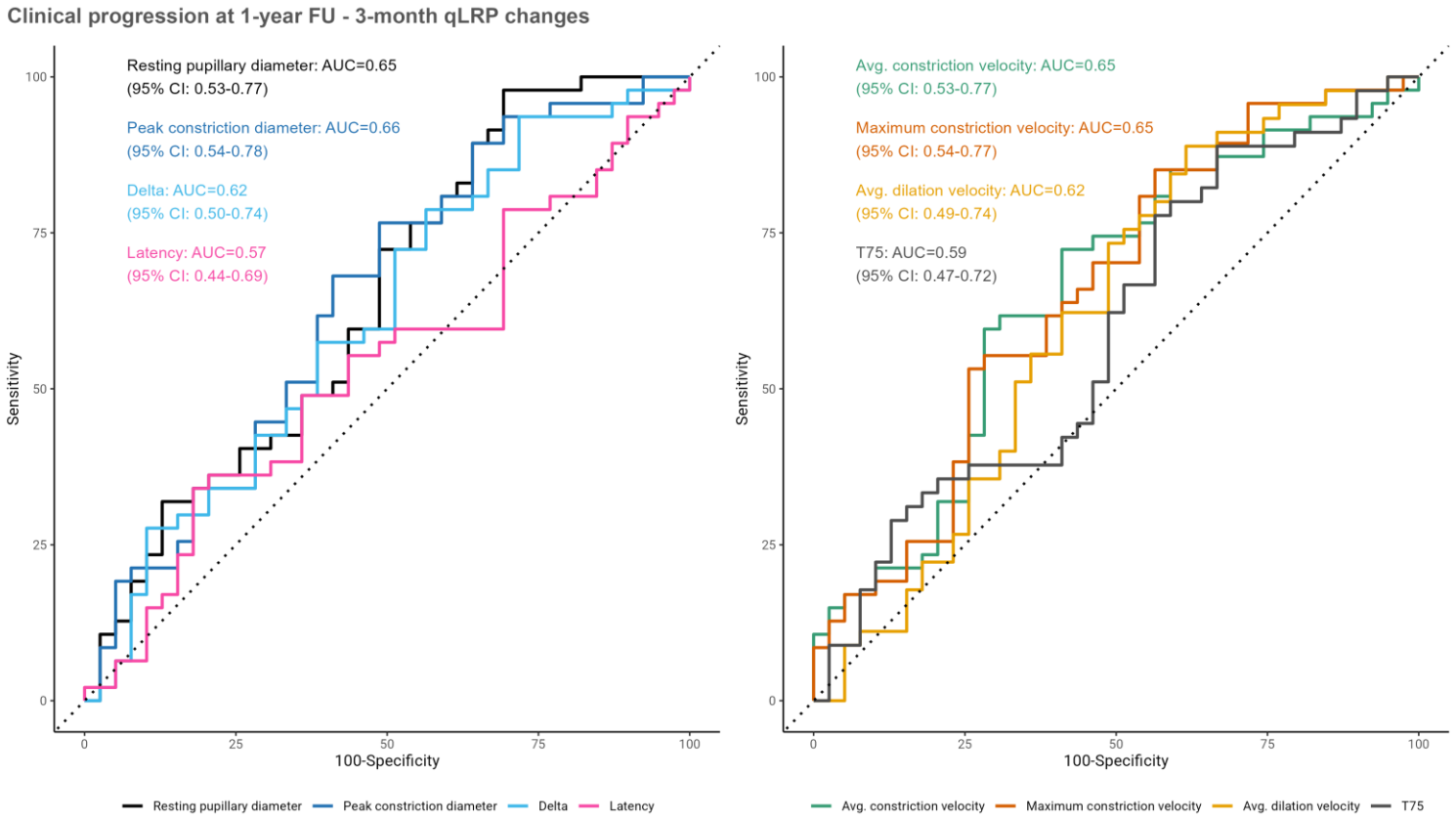
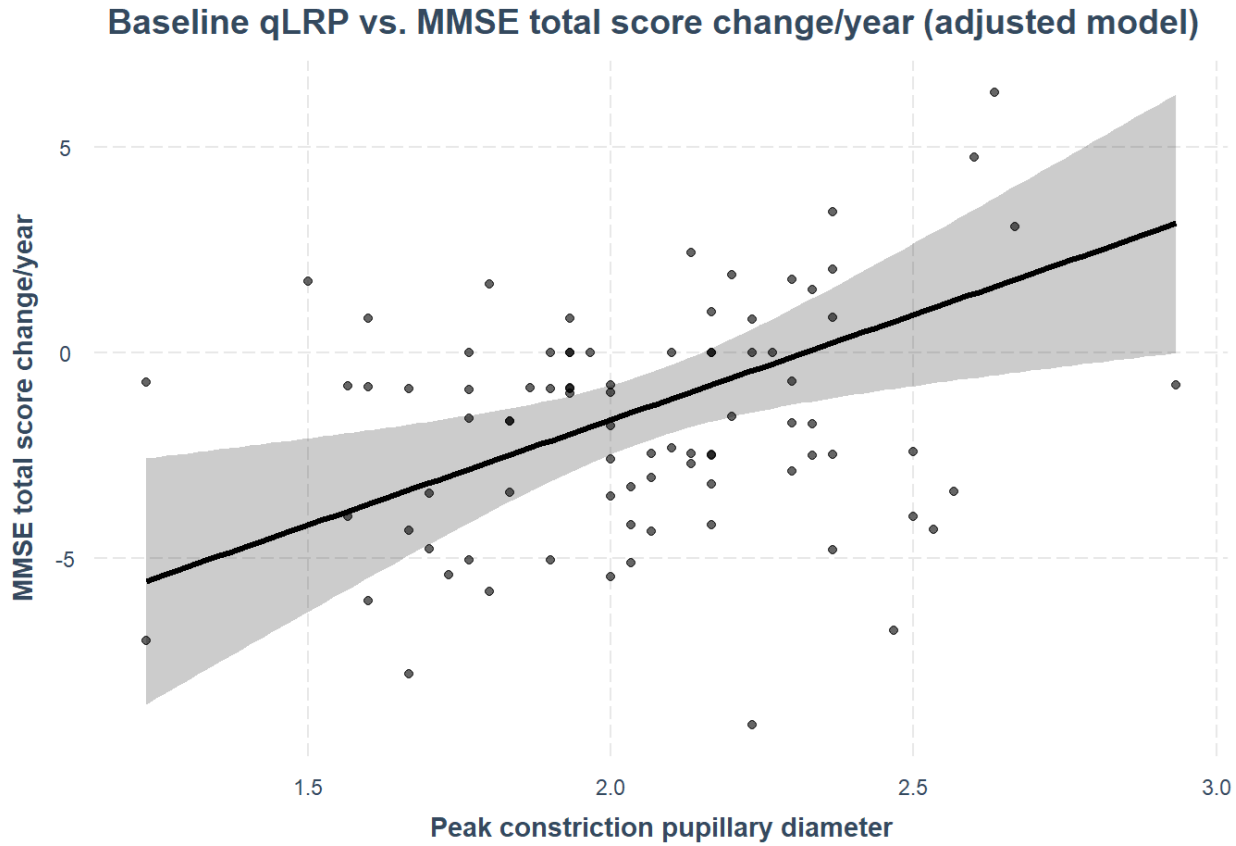
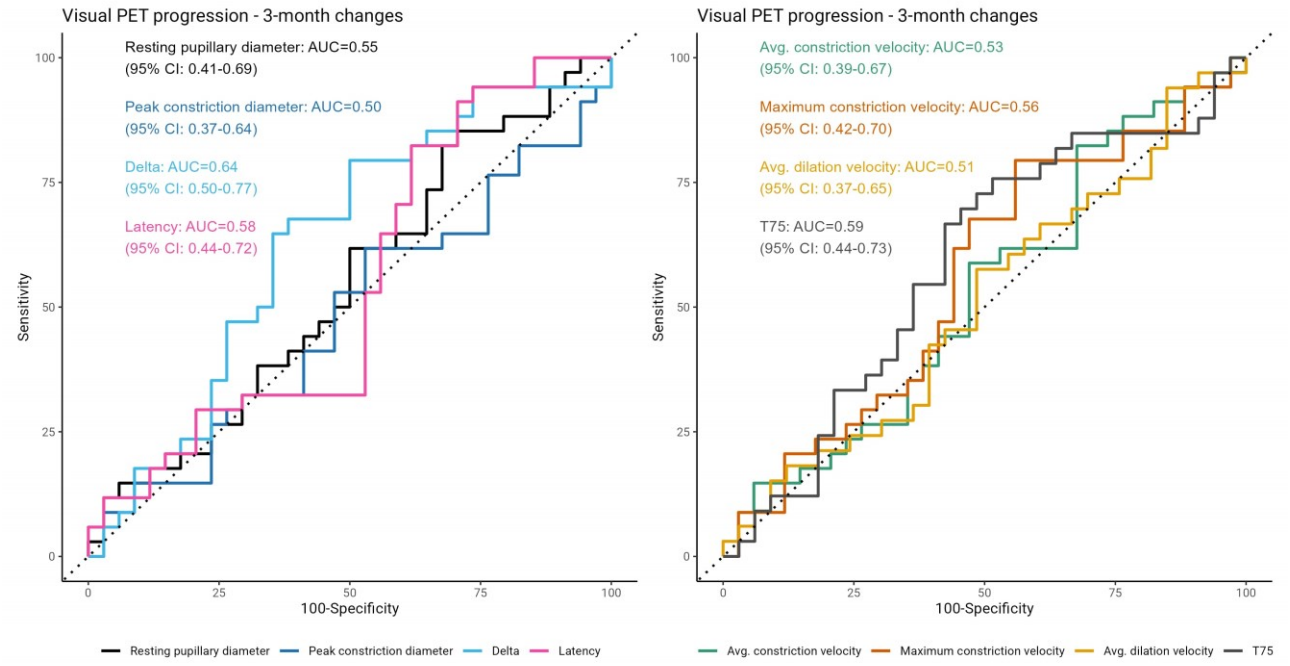


Figure 6. Partial regression plot showing the association between baseline peak constriction diameter and MMSE annualized Δ. The adjustment was for age, sex and resting pupillary diameter.



710 **Figure 7.** Receiver operating characteristics (ROC) curves for the prediction of visual [^{18}F]FDG-
711 PET progression for individual qLRP parameters.



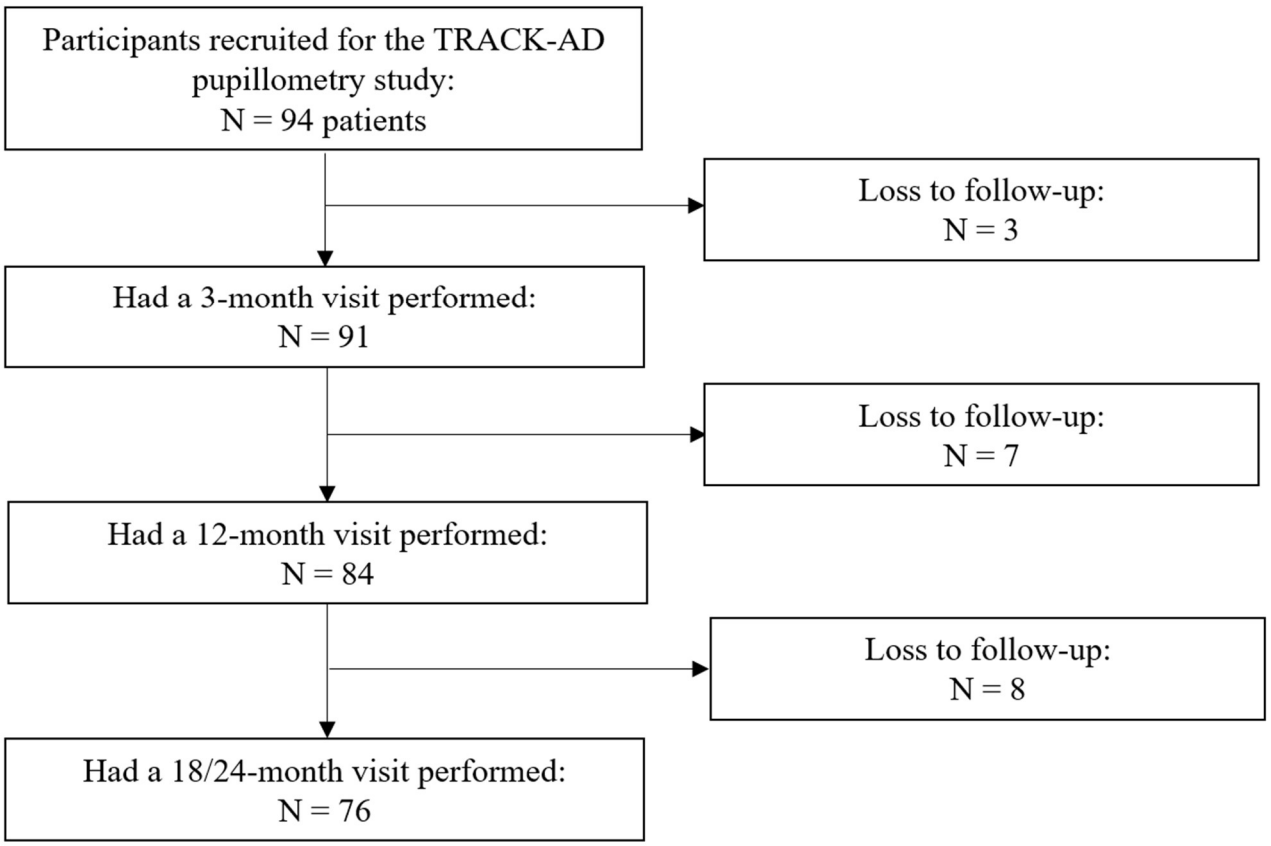
712

Supplementary Material

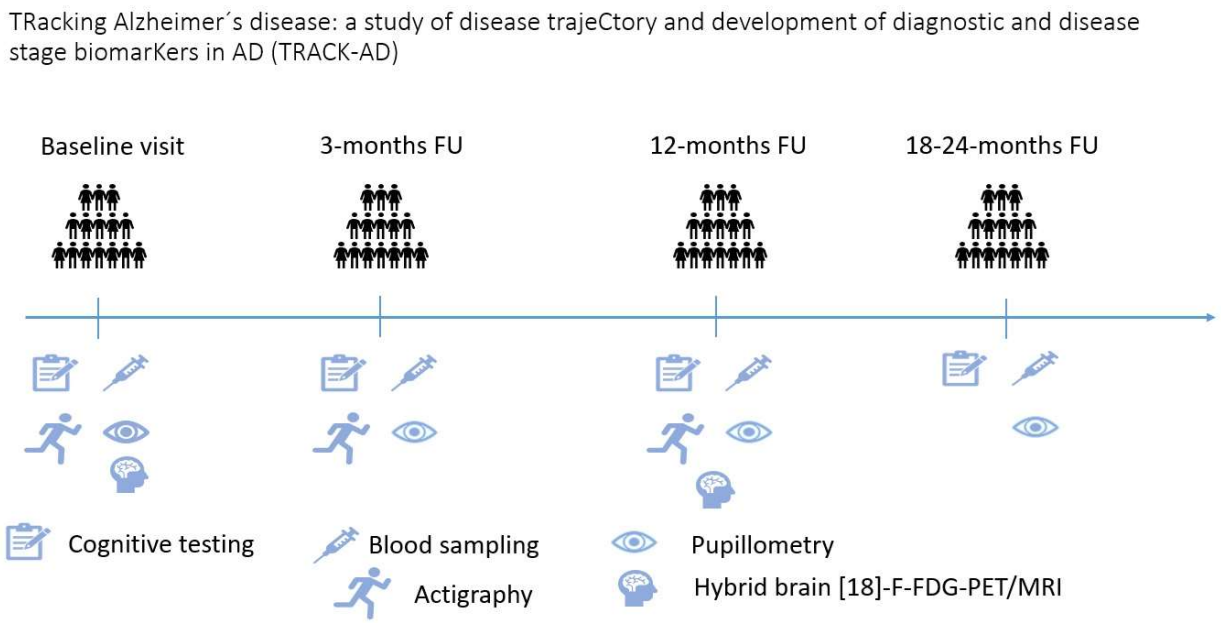
Related to *Prognostic value of light reflex pupillometry in Alzheimer's disease – a longitudinal cohort study*

Supplementary Figures

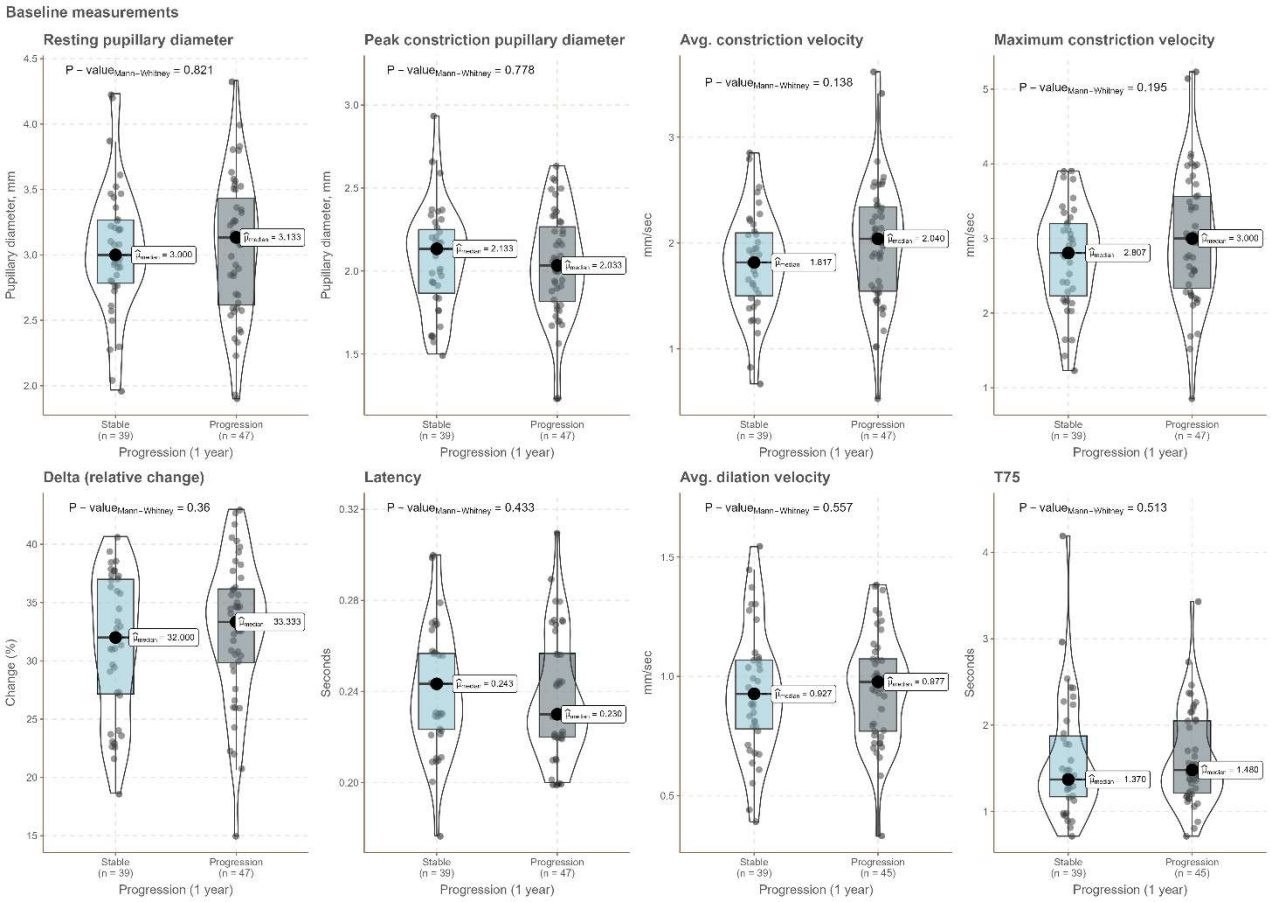
Supplementary Figure 1. Flow chart for inclusion and follow-up.



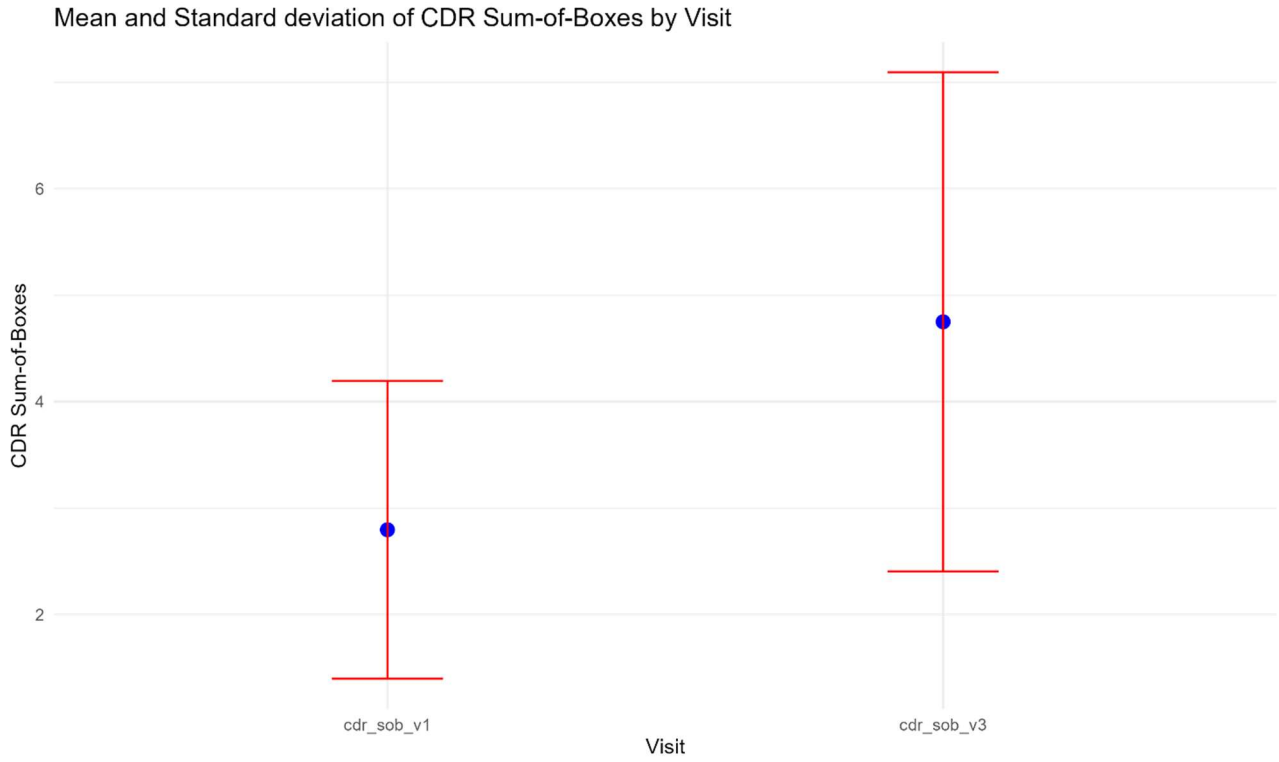
Supplementary Figure 2. Study overview



Supplementary Figure 3. Baseline visit measurements of qLRP and progression status at the 1-year follow-up visit.



Supplementary Figure 4. Mean and standard deviation for the Clinical Dementia Rating Sum-of-Boxes total score for each visit.



Supplementary Tables

Supplementary Table 1. qLRP metrics and a non-exhaustive list of the physiological interpretation of each marker as well brain areas associated with the marker.

Light reflex metric	**Physiological interpretation and brain areas associated with the marker**
Resting pupillary diameter	Indicator of autonomic balance[1], cognitive strain[2,3] and arousal[4]. Brain areas: LC[3], DRn[5], EWn[6–8], hypothalamus, insula, frontal eye field, colliculus superior.
Peak constriction diameter	Should be interpreted together with resting pupillary diameter, resulting in the delta value (see below). Brain areas: EWn, pretectal olivary nucleus
Delta, relative pupillary change	Reflects overall responsiveness of the pupil. Can be modulated by the frontal eye field during saccadic eye movements. Brain areas: EWn, frontal eye field (Brodmann area 8), ciliary ganglion
Latency	Represents overall nerve conductivity of the involved reflex arc, as delays could indicate demyelination[9]. Brain areas: optic nerve, chiasm and tract, pretectal brain area.
Constriction velocity	Indicator of EWn function, as the constriction mainly arises, when the neurons of this nucleus fires[1]. Brain areas: EWn, ciliary ganglion
Maximum constriction velocity	As for *constriction velocity*, but could indicate synchronicity in firing, i.e. robustness, of the EWn.
Average dilation velocity	Indicator of sympathetic recovery dynamics[1,10]. Brain areas: Hypothalamus, sympathetic spinal pathways, superior cervical ganglion.
T75	Same as for average dilation velocity, although a less reliable metric according to[11].

Abbreviations: EWn: Edinger-Westphal nucleus, LC: Locus coeruleus, DRn: Dorsal raphe nucleus.

Supplementary Table 2. Cohort characteristics at 1-year follow-up.

	Decline (N=47)	Stable (N=39)	Total (N=86)	p value
Disease severity				< 0.001[1]
N missing	1	0	1	
MCI	1 (2.2%)	10 (25.6%)	11 (12.9%)	
Mild	18 (39.1%)	28 (71.8%)	46 (54.1%)	
Moderate	26 (56.5%)	1 (2.6%)	27 (31.8%)	
Severe	1 (2.2%)	0 (0.0%)	1 (1.2%)	
MMSE total score				< 0.001[2]
N missing	1	0	1	
Mean (SD)	22.261 (3.242)	24.718 (3.379)	23.388 (3.509)	
Range	16.000 - 30.000	17.000 - 30.000	16.000 - 30.000	
CDR Sum-of-Boxes				< 0.001[3]
N missing	1	1	2	
Mean (SD)	5.891 (2.397)	3.368 (1.329)	4.750 (2.345)	
Range	2.000 - 11.000	0.000 - 8.000	0.000 - 11.000	

1. Pearson's Chi-squared test
2. Linear Model ANOVA
3. Kruskal-Wallis rank sum test

Clinically evaluated progression

Supplementary Table 3. Clinically evaluated progression (progressor vs. stable at 1-year follow-up) and qLRP (baseline measurements).

	Odds ratio	95% CI	p-value
Baseline pupil diameter (mm) (baseline)	1.05	[0.45, 2.42]	>0.9
Peak constriction pupil diameter (mm) (baseline)	0.72	[0.18, 2.79]	0.6
Delta, relative pupillary change (percent (baseline))	1.04	[0.97, 1.12]	0.3
Latency (s) (baseline)	0.01	-	0.6
Average constriction velocity (mm/s) (baseline)	1.84	[0.83, 4.35]	0.14
Maximum constriction velocity (mm/s) (baseline)	1.50	[0.87, 2.69]	0.2
Average dilation velocity (mm/s) (baseline)	1.43	[0.24, 8.72]	0.7
T75 (s) (baseline)	1.08	[0.54, 2.24]	0.8

Clinical Dementia Rating

Supplementary Table 4. CDR SoB (1-year annualized change) and qLRP variables (baseline measurements).

	Beta	95% CI	p-value
Baseline pupil diameter (mm) (baseline)	-0.19	[-0.57, 0.19]	0.3
Peak constriction pupil diameter (mm) (baseline)	0.00	[-0.62, 0.62]	>0.9
Delta, relative pupillary change (percent (baseline))	-0.03	[-0.06, 0.00]	0.081
Latency (s) (baseline)	-0.52	[-8.0, 7.0]	0.9
Average constriction velocity (mm/s) (baseline)	-0.12	[-0.48, 0.23]	0.5
Maximum constriction velocity (mm/s) (baseline)	-0.14	[-0.38, 0.11]	0.3
Average dilation velocity (mm/s) (baseline)	-0.58	[-1.4, 0.23]	0.2
T75 (s) (baseline)	-0.17	[-0.50, 0.16]	0.3

Supplementary Table 5. CDR SoB (1-year annualized change, log-transformed) and qLRP variables (3-month changes). Beta values are shown for 1 unit decreases in qLRP variables.

	Beta	95% CI	p-value
Baseline pupil diameter (mm) (3-month change)	0.34	[-0.81, 0.13]	0.2
Peak constriction pupil diameter (mm) (3-month change)	0.74	[-1.8, 0.28]	0.2
Delta relative pupillary change (percent (3-month change))	0.02	[-0.06, 0.03]	0.5
Latency (s) (3-month change)	-3.8	[-5.5, 13]	0.4
Average constriction velocity (mm/s) (3-month change)	0.42	[-0.87, 0.02]	0.061
Maximum constriction velocity (mm/s) (3-month change)	0.23	[-0.52, 0.07]	0.13
Average dilation velocity (mm/s) (3-month change)	0.32	[-1.6, 0.99]	0.6
T75 (s) (3-month change)	0.04	[-0.24, 0.33]	0.8

Mini Mental-State Examination

Supplementary Table 6. MMSE annualized change (1-year) and qLRP variables (3-month changes).

	Beta	95% CI	p-value
Baseline pupil diameter (mm) (3-month change)	1.1	[-0.35, 2.5]	0.14
Peak constriction pupil diameter (mm) (3-month change)	1.7	[-1.4, 4.7]	0.3
Latency (s) (3-month change)	-8.0	[-35, 19]	0.6
Delta, relative pupillary change (percent (3-month change))	0.09	[-0.03, 0.22]	0.14
Average constriction velocity (mm/s) (3-month change)	1.2	[-0.13, 2.5]	0.076
Maximum constriction velocity (mm/s) (3-month change)	0.53	[-0.35, 1.4]	0.2
Average dilation velocity (mm/s) (3-month change)	-2.3	[-6.0, 1.4]	0.2
T75 (s) (3-month change)	0.79	[-0.001 1.6]	0.054

Visual [^{18}F]-FDG-PET Progression

Supplementary Table 7. Visual PET progression and qLRP variables (baseline).

	Odds ratio	95% CI	p-value
Baseline pupil diameter (mm) (baseline)	1.18	[0.39, 3.63]	0.8
Peak constriction pupil diameter (mm) (baseline)	0.64	[0.10, 3.99]	0.6
Delta, relative pupillary change (percent (baseline))	1.05	[0.97, 1.14]	0.2
Latency (s) (baseline)	0.08	-	0.8
Average constriction velocity (mm/s) (baseline)	1.42	[0.57, 3.69]	0.5
Maximum constriction velocity (mm/s) (baseline)	1.41	[0.76, 2.70]	0.3
Average dilation velocity (mm/s) (baseline)	2.08	[0.29, 16.2]	0.5
T75 (s) (baseline)	1.24	[0.53, 3.00]	0.6

- the 95 % CI for latency was very broad owing to the small variance in the parameter.

Supplementary Table 8. Visual PET progression and qLRP variables (3-month change).

	Odds ratio	95% CI	p-value
Baseline pupil diameter (mm) (3-month change)	1.77	[0.51, 6.46]	0.4
Peak constriction pupil diameter (mm) (3-month change)	1.29	[0.08, 20.6]	0.9
Delta, relative pupillary change (percent (3-month change))	1.08	[0.98, 1.21]	0.15
Latency (s) (3-month change)	-	-	0.2
Average constriction velocity (mm/s) (3-month change)	1.22	[0.42, 3.64]	0.7
Maximum constriction velocity (mm/s) (3-month change)	1.33	[0.65, 2.8]	0.4
Average dilation velocity (mm/s) (3-month change)	1.52	[0.06, 40.8]	0.8
T75 (s) (3-month change)	1.41	[0.65, 3.27]	0.4

- The odds ratio for latency was very high due to modeling difficulty. The reciprocal odds are shown meaning that the odds ratio reflects the increase in risk of progression resulting from a 1 unit decrease in the predictor variable.

Confounder analysis

Supplementary Table 9. Results of the logistic regression models presented in Table 2 with additional adjustment for presence of mild eye disease (model 2) and use of pupil medication (Model 3). The possible confounders are added to the fully adjusted model (model 1).

	Model 1 (full model w/o eye disease and medication)		Model 2 (Mild eye disease)		Model 3 (Pupil medication)	
	Odds ratio [95 % CI]	**p-value**	**Odds ratio [95 % CI]**	**p-value**	**Odds ratio [95 % CI]**	**p-value**
Resting pupil diameter (3-month Δ)	4.28 [1.24, 16.9]	**0.028**[a]	4.7 [1.3,19.4]	**0.02**	3.7 [1.04, 14.8]	0.0504

References

1. McDougal DH & Gamlin PD. Autonomic control of the eye. *Compr Physiol* **5**, 439–73 (2015).
2. Steinhauer, S. R., Condray, R. & Kasparek, A. Cognitive modulation of midbrain function: Task-induced reduction of the pupillary light reflex. *International Journal of Psychophysiology* **39**, 21–30 (2000).
3. Huang, R. & Clewett, D. The Locus Coeruleus: Where Cognitive and Emotional Processing Meet the Eye. *Modern Pupillometry* 3–75 (2024) doi:10.1007/978-3-031-54896-3_1.
4. Cazettes, F., Reato, D., Morais, J. P., Renart, A. & Mainen, Z. F. Phasic Activation of Dorsal Raphe Serotonergic Neurons Increases Pupil Size. *Current Biology* **31**, 192-197.e4 (2021).
5. Grinberg, L. T. *et al.* The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer̈s disease. A precocious onset? *Neuropathol Appl Neurobiol* **35**, 406–416 (2009).
6. Scinto, L. F. M. *et al.* Selective cell loss in Edinger-Westphal in asymptomatic elders and Alzheimer's patients. *Neurobiol Aging* **22**, 729–736 (2001).
7. Scinto, L. F. M. *et al.* Focal pathology in the Edinger-Westphal nucleus explains pupillary hypersensitivity in Alzheimer's disease. *Acta Neuropathol* **97**, 557–564 (1999).
8. Mavroudis, I. A. *et al.* Dendritic and spinal alterations of neurons from Edinger-Westphal nucleus in Alzheimer's disease. *Folia Neuropathol* **52**, 197–204 (2014).
9. Bergamin, O. & Kardon, R. H. Latency of the pupil light reflex: Sample rate, stimulus intensity, and variation in normal subjects. *Invest Ophthalmol Vis Sci* **44**, 1546–1554 (2003).

10. Ferencova, N., Visnovcova, Z., Olexova, L. B. & Tonhajzerova, I. Eye Pupil – A Window into Central Autonomic Regulation via Emotional/Cognitive Processing. *Physiol Res* **70**, S669 (2021).

11. Gramkow, M. H., Clemmensen, F. K., Waldemar, G., Hasselbalch, S. G. & Frederiksen, K. S. Test-retest reliability and short-term variability of quantitative light reflex pupillometry in a mixed memory clinic cohort. *J Neurol Sci* **456**, 122856 (2024).

UNIVERSITY OF COPENHAGEN

FACULTY OF HEALTH AND MEDICAL SCIENCES