Doctoral theses at NTNU, 2020:404

Anker Stubberud

Digital Technology Biofeedback for the Prophylaxis of Pediatric Migraine

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

NTNU
Norwegian University of Science and Technology

Anker Stubberud

Digital Technology Biofeedback for the Prophylaxis of Pediatric Migraine

Thesis for the Degree of Philosophiae Doctor

Trondheim, December 2020

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

ISBN 978-82-326-5158-0 (printed ver.)
ISBN 978-82-326-5159-7 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:404

Printed by NTNU Grafisk senter

Digitalisert biofeedback som forebyggende behandling for migrene hos barn og ungdom

Smarttelefonen som behandling for migrene

I denne doktorgradsavhandlingen har jeg undersøkt om en smarttelefon-app kan brukes i forebyggende behandling av migrene hos barn og ungdom. Appen tar for seg et behandlingsprinsipp som heter *biofeedback*. Ved biofeedback trener brukeren på avslapping gjennom måling av kroppslige signaler. Ved å måle kroppslige signaler slik som muskelspenning og puls kan en lære seg hvilken atferd som fører til avslapping og forutgår migreneanfall, og på denne måten behandle migrene. Tradisjonell biofeedback krever stasjonært utstyr og en trent kliniker, noe som betyr at behandlingen er tidkrevende og lite tilgjengelig. Målet med denne avhandlingen var å se om ungdom på egenhånd kan behandle sin migrene gjennom biofeedback med bærbare sensorer koblet til en app.

I avhandlingen viser jeg at vanlige bærbare sensorer, slik som pulsklokker, er egnet til biofeedback. Videre beskriver jeg utviklingen av en biofeedback app for ungdom med migrene som bruker bærbare sensorer for å måle puls, temperatur og muskelspenning. Til slutt beskriver jeg en klinisk studie som viser at appen kan redusere hyppigheten av migrene hos barn og ungdom med omtrent en femtedel.

Navn kandidat: Anker Stubberud
Institutt: Institutt for nevromedisin og bevegelsesvitenskap
Veiledere: Mattias Linde, Erling Tronvik, Alexander Olsen og Trond Sand
Finansieringskilde: Norges Forskningsråd gjennom forskerlinjen NTNU

Ovennevnte avhandling er funnet verdig til å forsvares offentlig
for graden PhD i klinisk medisin
Disputas finner sted digitalt
torsdag 10. desember 2020, kl. 0830

Contents

Preface and Acknowledgements

This Doctoral Thesis represents my work at the Medical Student Research Programme and as a Ph.D. candidate. It was carried out at the Department of Neuromedicine and Movement Science, the Faculty of Medicine and Health Sciences, NTNU Norwegian University of Science and Technology. The Norwegian Research Council and The Faculty of Medicine and Health Sciences, NTNU, funded the project.

The thesis aims to investigate if it is possible to use a smartphone app to deliver biofeedback treatment to pediatric migraine sufferers and describes the development process of such an app entitled "Mi-Insight."

At the initiation of this project, I could not have imagined all the challenges I would face. I have gone from the meticulous work with reading and reviewing hundreds of papers for meta-analyses, to designing, planning and executing a study at the neurophysiologic laboratory, exploring software design and usability, venturing into the field of medical innovation and business development, and conducting a randomized clinical trial. All of this has been educational, interesting, and rewarding—but also difficult, frustrating, and sometimes seemed despairing and impossible. I appreciate all the experience I have gained and hope to take this into future clinical and academic work.

None of this would have been possible without supervisors, collaborators, and study participants.

Firstly, I would like to thank my main supervisors Mattias Linde and Erling Tronvik. Mattias is always available, always provides motivation, and gladly shares his knowledge. Erling stepped in as supervisor when Mattias moved to Sweden, and without him, this project would not have been where it is today. Thanks again to Erling, who facilitated my six-months stay at UCL Queen Square Institute of Neurology, London. Thanks to Manjit Matharu and Parashkev Nachev for their hospitality and for making my stay in London highly educational.

I would also like to thank my co-supervisors, Alexander Olsen and Trond Sand, who always provide invaluable feedback and support.

Thanks also to Gøril Bruvik Gravdahl for helping with contact with participants, Petter Moe Omland and the personnel at the neurophysiologic laboratory for help during the validation study, all the paper co-authors, NTNU Technology Transfer Office for enabling the establishment of Nordic Brain Tech AS, and all the study participants.

Last but not least, I would like to thank my family and friends for continuously showing their support and interest in my work.

Anker Stubberud

Molde, 2020

List of Contributions

Paper I: Stubberud A, Varkey E, McCrory DC, Pedersen SA, Linde M. Biofeedback as Prophylaxis for Pediatric Migraine: A Meta-analysis. *Pediatrics*. 2016;138(2)

Paper II: Stubberud A, Omland PM, Tronvik E, Olsen A, Sand T, Linde M. Wireless Surface Electromyography and Skin Temperature Sensors for Biofeedback Treatment of Headache: Validation Study with Stationary Control Equipment. *JMIR Biomedical Engineering*. 2018;3(1):e1.

Paper III: Stubberud A, Tronvik E, Olsen A, Gravdahl GB, Linde M. Mobile Health Biofeedback Intervention for Pediatric Migraine Sufferers: A Development and Usability study. *Headache: The Journal of Head and Face Pain.* 2020;60(5):889-901.

Manuscript IV: Stubberud A, Linde M, Brenner E, Heier M, Olsen A, Aamodt AH, Gravdahl GB, Tronvik E. Self-Administered Biofeedback Treatment App for Pediatric Migraine: A Randomized Pilot Study. *In press: Brain and Behavior*

Sammendrag på norsk – (Summary in Norwegian)

Introduksjon

Migrene hos barn og ungdom er vanlig, men dessverre finnes det få effektive forebyggende behandlinger. Biofeedback (BFB) er en av behandlingene som ser ut til å være effektiv, men er lite brukt i Norge, sannsynligvis fordi det krever spesialisert utstyr og trent helsepersonell for å sette opp behandlingen. Målet med dette prosjekter var å utvikle en ny smarttelefonbasert BFB-behandling for ungdom med migrene.

Metode

Først gjennomførte vi en systematisk oversiktsstudie med meta-analyser av eksisterende utprøvinger av BFB hos barn og ungdom med migrene. Deretter undersøkte vi om bærbare, trådløse sensorer var egnet for BFB ved å koble de til en app og sammenligne målinger med gullstandard utstyr. Videre prøvde vi ut appen i en brukervennlighetsstudie med 10 ungdom med migrene, mens vi utviklet en algoritme for å kombinere tre ulike BFB-modaliteter. Til slutt gjennomførte vi en randomisert sham-kontrollert pilotstudie for å vurdere behandlingseffekten av BFB-behandlingen.

Resultater

Den mest robuste meta-analysen viste at BFB reduserte ukentlig hodepinefrekvens med -1.97 (95% konfidensintervall (KI) -2.72 til -1.21) dager sammenlignet med ventelistekontroll. Utmerket til rimelig overensstemmelse (korrelasjonskoeffisient 0.81 til 0.58) ble vist for muskelspenningssensoren, mens utmerket overensstemmelse ble vist for temperatursensoren (korrelasjonskoeffisient=0.90; 95% KI 0.83-0.97). Vi laget en BFB-app som kombinerer feedback av muskelspenning, fingertemperatur og puls i en algoritme. Appen ble vurdert som brukervennlig og trygg. Ett tilfelle av hudutslett var den eneste bivirkningen som ble observert. Vi observerte en ikke-

signifikant reduksjon i hodepinefrekvens under uke 1-4 av behandlingen (2.92 dager, 95% konfidensintervall -1.00 ti, 6.84, p=0.145) og uke 5-8 av behandlingen (1.85 dager, 95% konfidensintervall -2.01 til 5.72, p=0.395).

Konklusjon

BFB ser ut til å være en effektiv forebyggende behandling for migrene hos barn og ungdom sammenlignet med ventelistekontroll, men evidensen er basert på få og små studier med en rekke metodologiske problemer. Bærbare, trådløse sensorer er egnet for BFB og det fremstår brukervennlig og gjennomførbart å bruke en app for BFB-behandling av migrene. BFB-appen førte til en liten reduksjon i hodepinefrekvens som hverken var signifikant eller bedre enn sham. Den begrensede behandlingseffekten kan muligens forklares av den minimalistiske behandlingen og lav etterlevelse.

Summary

Introduction

Pediatric migraine is common and disabling. Unfortunately, preventive treatment options are limited in terms of effectiveness, tolerability, and coverage. Biofeedback (BFB) has long been considered valid prophylaxis, but specific pooled analyses for evidence are absent. Meanwhile, wearable health monitoring sensor technology (WHMS) and mobile health represent new means for delivering BFB. This project's objective was to develop and evaluate a novel BFB intervention for pediatric migraine sufferers by self-administration through a smartphone with wearable sensors.

Methods

We carried out a systematic review with a comprehensive database search of existing literature to assess the evidence for using BFB in pediatric migraine. Studies meeting a set of predefined eligibility criteria were reviewed and meta-analyzed as appropriate. Mean differences (MD) and odds ratios (OR) with 95% confidence intervals (CI) were calculated. All eligible studies were assessed for risk of bias. Thereafter WHMS suited for muscle tension and finger temperature BFB were identified. The sensors were connected to a preliminary smartphone app and used by 20 healthy young volunteers in a validation study. Agreement with golden-standard equipment was calculated using Bland-Altman plots, intraclass correlation coefficients (ICC), and concordance correlation coefficients (CCC). The software was thereafter developed and improved in a usability study in 10 pediatric migraine sufferers. Three cycles of usability and feasibility testing, including a two-week home testing period, were completed. Changes in usability scoring and software implementations were analyzed statistically. In parallel, a BFB algorithm combining physiological parameters and several sham BFB alternatives were developed. Finally,

a prospective two-armed parallel, randomized, sham-controlled, double-blind trial was carried out. Recruitment proceeded unexpectedly slow and was terminated prematurely due to the SARS-CoV-2 pandemic. Twenty-three pediatric migraine sufferers were recruited, of which 16 were randomized to the BFB treatment app or a similar sham-version. The primary outcome was mean within-group change in headache frequency in the BFB group. A comparison of the change in headache frequency between the two groups was pre-specified as a secondary outcome. The change in headache days within the BFB group was analyzed with the Wilcoxon signed-rank test at weeks 1-4 and weeks 5-8 of the treatment period, while the change in headache frequency between groups was compared with a Mann-Whitney-U test.

Results

Five studies met the eligibility criteria for the systematic review. The most robust meta-analysis showed that BFB as part of a treatment package, reduced migraine frequency by -1.97 (95% CI -2.72 to -1.21) days per week compared to waiting-list control. Forty percent of the risk of bias assessments were deemed “low risk.” Data on adverse events were limited. Excellent agreement (ICC=0.81; 95% CI 0.57 to 0.92) to fair agreement (ICC=0.58; 95% CI 0.19 to 0.81) was found for the muscle tension sensor and excellent agreement (CCC=0.90; 95% CI 0.83-0.97) was found for the temperature sensor when compared to golden-standard equipment. A BFB algorithm combining muscle tension, finger temperature, and heart rate to give individualized feedback was developed. The app received consistently high usability scores and was evaluated by participants as tolerable and safe. In the pilot study, sixteen participants were randomized (biofeedback n=12, sham n=4) and analyzed. In the BFB group, a non-significant reduction in headache frequency was observed at weeks 1-4 (2.92 days, 95% CI -1.00 to 6.84, p=0.145) and weeks 5-8 (1.85 days, 95%

CI -2.01 to 5.72, p=0.395). The BFB group experienced a median of one fewer headache days/month vs. sham that did not reach significance (95% CI -4.0 to 9.0, p=0.830). The only adverse event observed was a case of mild skin rash.

Conclusion

BFB seems to be effective in reducing the frequency of migraine in the pediatric population when compared to a waiting-list control. Despite these positive findings, the evidence is based on a few small studies, involving a series of methodological issues that hampered proper meta-analyses. It remains uncertain if BFB provides any therapeutic gain of clinical significance. WHMS are suited for monitoring physiological parameters that are of interest in a BFB setting. Our findings also indicated that a mHealth app coupled with WHMS is feasible for delivering BFB. The BFB treatment app led to only a small reduction in headache frequency that was not significant nor superior over sham. The highly minimalistic nature of the BFB treatment app, combined with limited adherence, likely resulted in a smaller treatment effect than expected. Further research with a revised and improved version of the BFB treatment app is warranted.

Acronyms and Abbreviations

°C	Degrees Celsius
App	Smartphone application
BFB	Biofeedback
CBT	Cognitive behavioral therapy
CCC	Concordance correlation coefficient
CGRP	Calcitonin gene-related peptide
CI	Confidence interval
ECG	Electrocardiogram
eHealth	Electronic Health
Hz	Hertz
ICC	Intraclass correlation coefficient
ICHD	International Classification of Headache Disorders
IHS	International Headache Society
ITT	Intention-to-treat
LED	Light emitting diode
LOA	Limits of agreement
LOCF	Latest observation carried forward
MD	Mean difference
mHealth	Mobile Health
mITT	Modified intention to treat
MOH	Medication overuse headache
MVC	Maximal voluntary contraction
MVP	Minimal viable product
NNTB	Numbers needed to treat to benefit

NSAIDs	Non-steroidal anti-inflammatory drugs
RMS	Root mean square
OR	Odds ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SEMG	Surface electromyography
VC25	Voluntary contraction at 25% force
VC50	Voluntary contraction at 50% force
WHMS	Wearable health monitoring sensors

1 Introduction

1.1 Migraine diagnostic criteria and epidemiology

Migraine is a heterogeneous neurological disorder, of which the dominant feature is recurring severe headaches accompanied either by nausea or vomiting, or photo- and phonophobia. The headache is sometimes preceded or accompanied by transient, reversible focal neurological deficits, termed aura. Diagnostic criteria and classification are defined in the International Classification of Headache Disorders 3 (ICHD-3)[1] by the International Headache Society (IHS). The classification illustrates the heterogeneity of the different migraine syndromes and indicates that there is a most complex underlying neurobiology and pathophysiology. According to the Global Burden of Disease Study, migraine is the number one cause of neurologic disability and the most disabling condition worldwide among young adults (under 50 years of age).[2, 3] Worldwide, over one billion individuals had a headache disorder in 2016 further elucidating the vast population burden.[2, 4]

1.2 Migraine pathophysiology

Migraine is more than just a severe headache—and should instead be considered a complex neurological disorder with an altered "brain state."[5] Migraine is characterized by several phases, including a prodromal phase, sometimes an aura phase, the headache, and a postdromal phase. According to the current understanding, the hypothalamus and brainstem are central in the prodrome; the cortex is responsible for the symptoms seen during the aura; and activation of the trigeminovascular system plays a key role in the pain.[6] Generally speaking, these recurring phases are believed to be the overall result of malfunctioning sensory processing constituted on the basis of genetic and environmental factors.[7] Genome-wide association studies have found

several migraine-related genes,[8, 9] and it appears that migraine sufferers have an inherited neuronal hyperexcitability[10] making them more susceptible to have migraine attacks.[11, 12]

The headache pain involves the trigeminovascular system where neurons from the trigeminal ganglion and upper cervical dorsal roots project to cerebral vessels and the meninges.[13] The upper cervical roots and trigeminal nerve converge at the trigeminal nucleus caudalis, where nerve fibers ascend to higher centers for pain modulation.[14] This convergence may explain the pain distribution with ipsilateral forehead, back head, and cervical pain. During a migraine attack, activation of the trigeminovascular system leads to the release of vasoactive neuropeptides, including calcitonin gene-related peptide (CGRP).[15] The release of these peptides causes neurogenic inflammation and a lasting nociceptive stimulation of trigeminal nerve terminals which are thought to prolong and intensify migraine pain.[15-17] Moreover, there seems to be sensitization of both peripheral afferent neurons and central second order and higher neurons in migraine sufferers.[18] This sensitization may explain clinical features of the migraine such as motion sensitivity, hyperalgesia and allodynia.

The headache pain is sometimes preceded or accompanied by an aura, where the individual experiences one or more transient and completely reversible focal neurological symptoms. These symptoms may be positive, such as lines, shapes, and objects in the visual field; and negative, such as loss of vision, speech, somatosensation, or motor function. A cortical spreading depression is regarded as the neurophysiologic mechanism causing the aura and is characterized as neuronal and glial depolarization spreading across the cortex at 2-3 mm/min.[19]

Until the early 1980s, the vascular autonomic theory of migraine activation dominated the literature.[20] This theory hypothesized that excessive release of noradrenaline triggers intracranial vasoconstriction, which in turn results in a reflex release of vasodilators. The vasodilators lead to depolarization of primary nociceptive neurons in intracranial vessels and thereby pain. Although the vascular theory has been abandoned in favor of the above described neurogenic theory, many concepts of autonomic nervous system dysfunction are still valid in migraine pathophysiology. The emerging trend from several functional studies on sympathetic and parasympathetic function is that migraine sufferers seem to have a relative sympathetic impairment in the interictal period with paradoxical sympathetic hypersensitivity during the migraine attack.[20] Moreover, cranial autonomic parasympathetic symptoms such as lacrimation and rhinorrhea are common during migraine attacks and the number of symptoms seem to increase in frequency with the severity of migraine.[21] These parasympathetic symptoms are likely a consequence of an intense trigeminal activation in severe migraines and the trigeminal autonomic reflex.[22] Interestingly, cranial autonomic parasympathetic symptoms seem to be even more frequent in the pediatric migraine population.[23, 24]

1.3 Migraine in children and adolescents

The diagnostic criteria for migraine in children and adolescents are mostly the same as for adults. However, the definition of attack duration has varied through different versions of the classification system.[25, 26] The current diagnostic classification[1] defines that attacks in individuals below age 18 may be as short as two hours. In addition, the headaches in children and adolescents are more often bilateral, occipital headaches are rare, and photo- and phonophobia may be inferred by the behavior.

A review summarizing 64 studies from 32 countries and including 227,249 children and adolescents, estimated the mean prevalence of migraine to be 9.1%,[27] while another review estimated the six-month to lifetime prevalence to be 7.7%.[28] Both of these estimates are somewhat lower than the adult population, which seems reasonable as the prevalence of migraine increases into the mid-twenties before slowly decreasing.[29-31] Before puberty, the ratio of boys to girls is about 1:1, but during and after puberty more girls than boys are affected.[29] On the other hand, a Norwegian study estimated that over a third of school-aged children and adolescents might suffer from migraine when including probable migraine.[32]

1.4 Treatment of pediatric migraine

1.4.1 General measures

The first step to treating pediatric migraine should be applying general measures, such as educating the child and family, applying lifestyle measures, and initiating the use of a headache diary. Lifestyle measures include good sleep hygiene, regular and adequate meal and fluid intake, regular exercise, and avoidance of trigger factors. The lifestyle measures often restrain common precipitating factors such as stress, poor sleep habits, irregular meals, limited fluid intake, odors, and foods. Together with such measures, a headache diary is helpful to identify both deleterious factors and triggers and thus possibly avoid attacks.[33] Besides, the diary gives a useful overview of the disease burden and possible treatment effects and aids the clinician in further management.

1.4.2 Abortive treatment

When migraine symptoms develop, the use of abortive medications should be considered. Over-the-counter analgesics are effective in treating migraine attacks, and

NSAIDs seem to be more useful than paracetamol.[34] These drugs are more effective if given early in the course of an attack and are often successful in treating mild to moderate attacks. For moderate to severe attacks, triptans should be considered.[35] Triptans are serotonin (5-HT) agonists with an affinity for the receptor subtypes 5-$HT_{1B/1D}$ that constrict intracranial vessels, inhibit nociceptive transmission in the trigeminovascular system of the brainstem, and inhibit the release of vasoactive neuropeptides such as CGRP.[36, 37] In addition, the combination of triptans and NSAIDs seems to be more effective than placebo.[38] Triptans may also be administered nasally or subcutaneously in cases of nausea and gastric stasis.[39, 40] In Norway, there is no accepted indication to prescribe triptans to children under the age of 12. However, a recent Cochrane review concludes that there is moderate evidence for using sumatriptan and rizatriptan in younger children,[34] which should be considered when other measures are inadequate.

Frequent administration of abortive drugs may lead to medication overuse headache (MOH). The child and family should be educated about the risk of MOH, and abortive medication should be limited to no more than two days per week.[41] When migraine attacks are frequent or long-lasting, cause significant disability, reduced quality of life, abortive therapy have failed, or there is a risk of MOH, prophylactic treatment should be considered.

1.4.3 Pharmacological prophylaxis

Propranolol, amitriptyline, topiramate, cinnarizine, and flunarizine are commonly used prophylactic medications for children and adolescents with migraine,[42] but the evidence for all of these are limited.[43] A systematic review of pharmacological prophylaxis of migraine in children and adolescents found topiramate to be effective,

whereas propranolol and flunarizine were ineffective.[43] The practice guideline recommendations from the American Academy of Neurology and American Headache Society from 2019 also concluded that the majority of prophylactic medications for pediatric migraine fail to demonstrate superiority over placebo.[44] Only propranolol, topiramate, and cinnarizine were found to possibly be more effective than placebo.[44] While propranolol may be effective, the drug may prove problematic in patients with asthma, diabetes, or depression. Moreover, the recent double-blind CHAMP-trial,[45] randomizing over 300 children and adolescents, found no advantage of neither amitriptyline nor topiramate over placebo. On the contrary, the active drugs were associated with a higher degree of adverse events (AE).[45] Still, amitriptyline is considered by some experts as a first-line choice, which may partly be based on its beneficial once-daily dosing.[46] Finally, both the practice guidelines from 2019[40] and a recent meta-analysis concluded that the effect of flunarizine is not significant.[43] Together, this shows that options for prophylactic treatment of pediatric migraine are limited and that there is a great need for new therapeutic options.

1.4.4 Non-pharmacological prophylaxis

Non-pharmacologic treatments seem to be valid options for prophylactic treatment of pediatric migraine. Several systematic reviews with meta-analyses,[47-49] including a Cochrane review updated to 2018,[50] indicate that psychological treatments effectively treat chronic pain conditions in children and adolescents. Among these, the behavioral treatment biofeedback (BFB), was found to be effective in pediatric headache prophylaxis.[50] Moreover, behavioral interventions seem to be beneficial adjuvants to pharmacologic prophylaxis.[51-53] Unfortunately, several studies included in the meta-analyses mentioned above had a high risk of bias, and mostly waiting-list controls leading to low-quality grade of evidence and uncertainties regarding the therapeutic

gain.[47-50] In addition, headache disorders and different interventions were merged, which gives diagnostic uncertainties and population heterogeneity.[54] Prior to the initiation of this thesis, there were no pooled analyses specific for BFB in pediatric migraine.

1.5 Biofeedback

Throughout the 1960s -70s, both animal and human studies demonstrated that control over many physiologic parameters could be learned through feedback.[55] BFB utilizes technologies and equipment that monitor physiological processes that usually are considered to be involuntary and modulated without conscious awareness. These physiological processes are converted to a signal, usually visual or auditory, and presented to the individual (Figure 1). With this feedback, users can learn to control specific physiological processes. BFB can be applied in a wide range of areas, of which headache treatment seems to be one of the most fruitful.[56] Finger temperature, head-neck muscle tension, and heart rate are physiologic parameters that are usually considered the most effective for BFB treatment of headaches.[57-59]

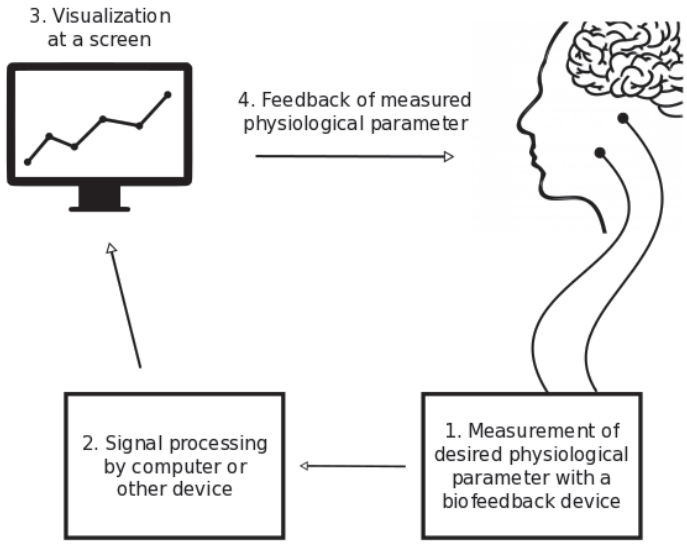


Figure 1. *A biofeedback device is used to measure the desired physiological parameter. The device is attached to the body, and the signal is transmitted to a computer for signal processing before visualization at a screen. This allows the user to easily interpret a "feedback" of his or her physiological responses. Traditionally this is a stationary and bulky setup that requires specialized equipment.*

BFB is today available in both clinic- as well as home-based (minimal therapist-contact) formats, and there is no clear evidence for the superiority of either in pediatric pain.[47, 60, 61] However, delivery of BFB in a clinic-setting requires the presence of a trained therapist and suited equipment to monitor the desired physiological parameter, which is costly and time-consuming. Even though BFB seems to be effective as migraine prophylaxis, entirely home-based and therapist-independent, self-administration of BFB specifically for pediatric migraine remains to be investigated.

1.5.1 Peripheral skin temperature

For BFB purposes, peripheral skin temperature is usually measured using a thermistor. This is a thermally sensitive resistor that reacts with precise changes in resistance proportional to small body temperature changes. Finger temperature depends mostly on the supply of warm arterial blood in the capillaries, but blood can also be shunted in deeper layers of the skin which dissipate less heat.[62] The finger temperature is correlated with stress and sympathetic-parasympathetic tone.[63] Thus, various stimuli triggering the sympathetic nervous system induce decreased capillary microcirculation in peripheral skin and thereby cold fingers.[64] The association between headache improvement, stress, autonomic nervous system activity, and finger temperature is not fully understood, but early studies stemming from the vascular pathophysiologic theory of migraine have found that self-regulation of skin temperature is correlated to changes in cerebral blood flow,[65] and predict headache outcome.[66]

1.5.2 Heart rate

Heart rate is commonly used as a BFB parameter, usually measured as a photoplethysmogram (PPG). The PPG was described as early as in 1938 by Hertzman,[67] and is simply a visualization of the blood volume change in transilluminated tissue caused by the passage of blood.[68] The devices used for photoplethysmography consist of a light-emitting diode (LED) and a photodetector.[69] Blood absorbs more light than the surrounding tissue. Especially green light has a great absorptivity in both oxyhemoglobin and deoxyhemoglobin and is thus suited for heart rate measurement as the blood pulses through the skin.[70, 71] The PPG is applied in various tools such as wristband heart rate monitors and oximeters.[72] In most

wristband heart rate monitors, the photodetector senses reflection of the light that is scattered back from tissue, which enables placement of the LED and photodetector near each other. The reflection detection is prone to disturbances by motion artifacts, pressure, and hampered skin contact during activity.[69] Therefore, such wristband heart rate monitors are not as accurate as the golden-standard electrocardiogram (ECG).[73, 74] As an example, the Fitbit Charge HR™ (Fitbit Inc.), Apple Watch™ (Apple Inc.), Mio Alpha™ (Mio Global (Physical Enterprises)), and Basis Peak™ (Basis Science Inc.) are shown to have variable accuracy compared to the ECG, with the greatest accuracy at rest, and diminishing accuracy with increasing exercise intensity,[74] and are thus limited to less demanding heart rate measurements.[75]

1.5.3 Muscle tension

To utilize muscle tension as a BFB parameter, one must measure the surface electromyographic (SEMG) voltage. The SEMG signal source is the motor unit action potential, occurring upon depolarization of muscle fibers. The depolarization creates an electrical signal that can be picked up by electrodes attached to the skin. The pickup of the electric signal at skin surface is influenced by the impedance in the electrode-skin interface, which is decided by a series of factors such as dry, cornified skin or oily skin. Thus it is desirable to prepare and optimize electrode-skin contact by different means such as cleaning and using electrode conductive gel before recording.[76] If the impedance is too high, this will distort the signal. By using three electrodes (one for active recording, another for the reference recording, and a third "patient ground" electrode), one may compare the voltage that is common to electrodes and eliminate this, a strategy referred to as the common-mode rejection. This allows for amplification of the difference in voltage, described as gain, in order to boost the signal strength and to utilize the SEMG signal in a clinical setting.[77] This

analogous amplified signal is then usually filtered to reduce unwanted signal sources such as motion artifacts and remaining power line interference.[76] It is common to use a notch filter to remove power line interference noise (in Norway at 50 Hz); a high-pass filter at 10-20 Hz to reduce motion artifacts; and a low-pass filter at 1000 Hz. The low-pass filter is set to at least double the highest signal frequency, which for SEMG is 500 Hz. This double frequency is defined as the Nyquist frequency and inhibits so-called *aliasing* in which an incorrect (alias) signal is picked up from the source signal.[76] The signal is then digitalized by an analog to digital converter to allow for visualization at a computer. Finally, the signal may be quantified by calculating the root mean square (RMS) or the area under the curve,[78] which allows for voltage quantification and analyses of muscle activity across devices and individuals.

1.5.4 Biofeedback physiology and mechanisms of effect

The mechanisms of effect for BFB treatment in headache are poorly understood, but it is likely that a large proportion of the treatment effect may be attributed to non-specific effects.[79] Nevertheless, I will provide some insights into physiological assumptions and potential mechanisms.

The central assumption about the mechanisms of effect of BFB is that bodily responses, often autonomic, traditionally believed to be unmodifiable, can indeed be modified by instrumental conditioning through feedback.[80] This instrumental conditioning (also known as operant conditioning) is a type of learning process where behavior is modified by reinforcement or punishment. In one experiment form 1973, researchers demonstrated that baboons could learn self-control over blood-pressure.[81] The baboons were strapped to a restraining chair, having their blood pressure

measured intraarterially, while feedback over diastolic blood pressure was given both as a light signal; and rewarded with food and punished with electrical shocks. The baboons were able to induce a lasting large-magnitude increase in blood pressure voluntarily. Later, several studies in humans also demonstrated that subjects could be trained to increase and decrease vasomotor responses, heart rate and rhythm, and galvanic skin response.[80, 82, 83]

The most perspicacious theory explains that learning and inducing a physiological change directly leads to biological adaptations that are beneficial for migraine. Generally speaking within this directly causal framework, researchers argue that BFB may induce long-term alterations in autonomic tones, muscle tension, and blood flow, reduce the excitability within central nervous system networks, and render individuals more resilient to effects of environmental stressors.[84, 85] Yet, this idea is partialy built on the vascular pathophysiological construct of migraine and provides insufficient explanations. It remains unclear whether there are any specific physiological changes induced by BFB that drive the improvements in headaches.[86] This is supported by several studies that have failed to find a correlation between improvement in BFB parameters and headache outcome, and a study demonstrating that the *inability* to raise hand temperature predicted treatment success.[87]

A more recent theory is that the vagus nerve might mediate the antinociceptive effect of BFB. This idea stems from evidence that vagus nerve stimulation as a migraine therapy modulates pain and improves headaches.[88] It is conceivable that the same mechanism of vagus nerve stimulation occurs during BFB—through voluntary control and modulation of parasympathetically controlled functions.[89] This theory is supported by studies demonstrating that heart-rate variability BFB induces vagal

afferent pathway activity.[89, 90] However, it should be noted that the observed vagal afferent activity is mainly valid for heart rate variability BFB and not necessarily BFB of other modalities.

Other theories of mechanisms of effect cover a wide range of explanations from the reduction of oxidative stress[91] to cognitive and biological mechanisms of distraction from the pain.[92]

1.6 Digital technology in migraine

1.6.1 Mobile health and wearable health monitoring sensors

In the current digital technology era, health services are increasingly often provided through the use of electronic devices, communication technology and informatics (eHealth). A subcategory of eHealth, labeled mobile health (mHealth), covers the use of smartphones, applications (apps), and wearable sensors for medical purposes.[93, 94] Wearable sensors, such as heart rate wristbands, let patients access real-time data from a broad range of physiological parameters at home.[95, 96] Technically, a sensor is the component of a system whose purpose is to detect events or changes in its environment, and this information is thereafter sent to other electronics for processing. However, in commercial terms, the word *sensor* is often used for the whole setup—i.e., including the microcontroller, display, and other functions. The latter definition will be used in this thesis. There is an increasing trend in the use of wearable sensors[96] within a wide range of medical fields, such as endocrinology (diabetes care), cardiology and neurology.[97-102] Unfortunately, the efficacy, acceptability, and credibility of mHealth is limited[103-106] and validation of new wearable sensors is still insufficient.[107-109]

1.6.2 Mobile health in migraine

While some aspects of mHealth in migraine have been explored, there is still a gap between commercially available solutions and scientifically validated and developed solutions.[103, 104] Within headache medicine, most of the available mHealth products are headache diaries.[110] However, electronic behavioral interventions for migraine seem acceptable and feasible, but efficacy measures are uncertain.[111] Clinical trials investigating the efficacy of mHealth-based classical behavioral therapies such as cognitive behavioral therapy (CBT), BFB, and relaxation are nearly non-existent.[112] Currently, no mHealth solution delivering BFB as prophylaxis specifically for pediatric migraine exists.

1.6.3 Development and usability of new mobile health therapies

To mitigate the limitations that mHealth is facing today, several researchers propose means for assessing mHealth quality and recommendations for mHealth development.[113-118] The importance of usability and functionality in the development of mHealth is especially emphasized. Usability is a term used to assess the ease of user interfaces, and also refers to the process of improving user experience during a design process. A commonly used design strategy is iterative and incremental development in which several rounds (or iterations) of usability testing are conducted while implementing changes in the software for each round.[119] In addition, researchers should adhere to guidelines and regulations; consider the potential market and target group; and ensure accountability and availability when developing new mHealth.[112]

Considering the potential effectiveness of BFB as a pediatric migraine prophylactic, combined with the paucity of mHealth in migraine therapy in this era of digital

technology—we embarked on a project to develop a new mHealth BFB intervention for pediatric migraine sufferers. The overall aim of this thesis is to describe the development process and research leading to a new mHealth BFB intervention for pediatric migraine sufferers entitled “Mi-Insight,” and a pilot clinical trial of this intervention.

2 Aims

Hypothesis:

Biofeedback is an effective, tolerable, and safe prophylaxis for pediatric migraine, and can be self-administered through a smartphone with wearable sensors.

Aims:

- Assess the pooled evidence for using biofeedback as migraine prophylaxis in the pediatric population through a systematic review and meta-analysis of existing studies (Paper I).

- Evaluate the validity of using wireless sensors to measure surface electromyography and peripheral skin temperature, in combination with a mobile phone application, as the basis for a self-administered biofeedback intervention (Paper II).

- Develop a biofeedback treatment app aimed at pediatric migraine sufferers while investigating the intervention's usability and feasibility (Paper III).

- Assess initial estimates of efficacy, safety, and tolerability of a biofeedback treatment app for pediatric migraine, while also investigating the suitability of a sham biofeedback treatment app (Manuscript IV).

3 Methods

3.1 Project overview and overview of intervention development

This section gives an overview of the project and the development process of the BFB intervention. The development process spans all the studies contributing to the thesis.

In the first phase, we set out to review the available literature to serve as a base for upcoming studies and establish a knowledge base for developing the BFB intervention. Through a systematic review and meta-analyses, we learned how BFB traditionally is delivered to pediatric migraine sufferers. In parallel, we identified WHMS suitable for monitoring physiological parameters of interest in BFB treatment—i.e., SEMG, peripheral skin temperature, and heart rate.[57, 58] Heart rate has previously been thoroughly studied for validation,[73, 74] and thus the main goal was to identify SEMG and skin temperature sensors, meeting a set of predefined criteria. Thereafter we set out to create a preliminary minimal viable product (MVP) of the app software. This preliminary version was programmed to serve as the substrate for validating the chosen sensors and the starting point for further software development. The preliminary MVP was to include the essentials for BFB training, a headache diary, and basic information and instructions.

Secondly, we recruited healthy volunteers to establish the validity of the chosen WHMS. The process was exploratory with a main aim to evaluate the agreement between the chosen WHMS and stationary golden-standard neurophysiological equipment following recommended guidelines for agreement studies.[120]

Thirdly, following validation of the sensors, we recruited pediatric migraine sufferers to use the app in an iterative and incremental fashion through three cycles of usability and software testing.[119] In between each cycle of testing, issues with the software

were addressed and improved. Once the final cycle of testing was complete, we described a BFB algorithm combining three physiological parameters. We also used the collected data to design several options for sham-BFB.

Lastly, after usability testing and development, a final version of the app was created to be employed in clinical trials. A prospective, randomized, sham-controlled, double-blind, pilot trial was conducted to give initial estimates of the efficacy, safety, and tolerability of the intervention in pediatric migraine sufferers.

3.2 Systematic review with meta-analysis

In order to review the evidence base for BFB as a prophylactic intervention for pediatric migraine sufferers we carried out a systematic review with meta-analyses according to the standards of the Cochrane Collaboration[121] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[122]

3.2.1 Eligibility criteria for study inclusion

Included studies were required to be prospective randomized controlled trials investigating BFB as a prophylactic treatment for episodic migraine in children or adolescents. Participants were children and adolescents up to the age of 18 suffering from episodic migraine. We did not require the use of a specific set of diagnostic criteria (e.g., IHS Classification Committee 1988[25] or ICHD-II 2004[26]), but the diagnosis had to be based on a least some of the distinctive migraine features defined by the IHS (ICHD-3beta[123] at the time of the study). Studies were eligible if at least one arm represented BFB treatment, and when some degree of behavioral treatment was delivered alongside BFB during the same session, or BFB was the only difference between the intervention and comparison groups. Eligible comparison groups were active treatment with documented effectiveness; non-pharmacological

therapies with documented effectiveness; waiting-list control; or treatment as usual. Migraine frequency was chosen as the primary outcome of interest. Pre-specified secondary outcomes to be extracted were: responder rate equal to or greater than 50%, headache intensity, attack duration, disability, quality of life, doses of acute medication, and AEs.[54]

3.2.2 Search methods and study selection

A medical librarian performed the literature search as recommended for systematic reviews.[124] The searched databases included MEDLINE, EMBASE, CENTRAL, CINAHL, and PsychINFO. The search was updated on November 23, 2015. It involved a combination of thesaurus and free-text terms optimized to cover randomized controlled trials where patients under the age of 18 had received BFB treatment as a prophylaxis for migraine. The lists of references in all reviews encountered on the subject were hand-searched in order to capture potentially relevant studies not detected in the electronic search. The search results were screened to identify eligible studies. In cases where papers could not be excluded based on information in the title and abstract, full texts were obtained and screened.

3.2.3 Data extraction and management

Characteristics of each included study and information on BFB treatment and any additional treatment were reviewed. Raw outcome data were extracted from the studies for meta-analyses. We primarily sought the number of participants, means, and standard deviations. In such cases where this could not be obtained directly from the paper, the data were calculated in-house from the information provided in the paper. We attempted to standardize the unit of time over which outcomes were measured.

3.2.4 Risk of bias assessment in included studies

Four categories of bias were considered: selection bias with regard to random sequence generation and allocation concealment; detection bias with regard to blinding of outcome assessors; attrition bias, i.e. selective occurrence and biased handling of protocol deviations and losses to follow-up; and reporting bias determined by differences between pre-specified measures and reported outcomes. Performance bias was not assessed due to the difficulty of blinding participants and personnel when delivering BFB treatment. The risk of each bias was graded as being "low," "high," or "unclear." The latter was chosen when the information provided in the paper was insufficient to determine the risk.

3.3 Sensor validation

In order to assess the validity of the sensors with the preliminary MVP software, we conducted a sensor validation study with young adult healthy volunteers.

3.3.1 Participants and equipment

Based on a sample size calculation for agreement studies by Bonett,[125] we recruited 20 healthy volunteers (Appendix 1). Participants were recruited as a convenience sample by actively seeking out young individuals from the local research and student community.

The NeckSensor™ (EXPAIN AS, Oslo, Norway) was used as the wearable sensor to measure muscle tension. For wireless measurement of temperature, we selected the PASPORT Skin/Surface Temperature Probe, PS-2131 combined with PASPORT Temperature sensor, PS-2125 and AirLink, PS-3200 (Pasco Inc., Roseville, CA, USA). Both of the sensors transmitted signals via Bluetooth® Smart/4.0. As

stationary equipment, we used an ADInstruments Inc. (Dunedin, New Zealand) setup with 5-Lead wires attached to silver/silver-chloride electrodes and fed through a BioAmp to a PowerLab 8/35 for SEMG and ECG recording and an ADI Skin temperature probe for temperature recording.

3.3.2 Experimental procedure

Participants were seated in a recliner at a 90-degree angle in the neurophysiological laboratory. The two NeckSensor^{TM} electrodes were placed over the upper fibers of the right trapezius muscle midway along the line between the spinous process C7 and the acromion.[78, 126] Because simultaneous registrations of SEMG signals from the exact same location with different sets of surface electrodes are not possible, one set of electrodes from the stationary equipment was placed 2 cm cranially of the NeckSensor, and one set was placed 2 cm caudally. The inter-electrode distance was 4 cm. The "patient ground" electrode for the stationary equipment was placed over the spinous process C7 (Figure 1, Paper II). The skin beneath the stationary electrodes was washed with alcohol swabs. The two skin temperature sensors were attached, without touching each other, to the volar pad of the distal phalange on the second finger with sticky tape, with the stationary sensor placed radially of the two.

Initially, each participant was asked to relax for 5 min to allow the skin temperature to increase during relaxation. The relaxation was achieved by asking the participant to "do nothing" and sit still in the recliner. This served to give a baseline (relaxed) muscle tension measurement. Recording of relaxed trapezius muscle tension (baseline) was made during the last 30 seconds of relaxation. Thereafter, the temperature sensors were detached to allow the measurement of room temperature for the remainder of the procedure. Subsequently, the participant was instructed to

complete a series of exercises to activate the upper fibers of the trapezius muscle. Arbitrary angle isometric maximal voluntary contraction (MVC), through shoulder elevation, was completed in three repetitions, each lasting for 6 seconds.[78, 127-129] The SEMG and the force were simultaneously registered. The force was recorded by a dynamometer (Manual Muscle Tester, Lafayette Instruments, USA) attached to a fixed sling placed over the acromion. Subsequently, the participant was asked to complete similar sets of contractions at 50% (VC50), and 25% (VC25) of maximal contraction guided by a sound signal from the dynamometer elicited at the corresponding set force. Finally, the participant was asked to complete four repetitions of static contractions (15 s each) performed by abducting both shoulders to a 90-degree angle and holding against gravity.[78] After completing the exercises, the participant was asked to answer a 5-item user evaluation questionnaire to assess the practicality of use and the safety of the sensors.

3.3.3 Data management

The NeckSensor uses a 12-bit ADC resolution sampled at 1,024 Hz with a third-order 10–480 Hz active bandpass filter. The PowerLab sampled the SEMG signals at 2,000 Hz with a fourth-order Bessel lowpass filter at 500 Hz and a first-order high pass filter at 10 Hz. In addition, a 50-Hz notch analog filter was applied.[130] All stationary recordings were evaluated visually for the presence of ECG artifacts. The stationary readings were averaged over the two sets of electrodes. Stationary and wireless readings were then RMS rectified, and the RMS values were analyzed as the mean for each muscle contraction exercise. For the temperature measurements, we calculated the difference from the start to the end of relaxation, and the difference between the temperature at the end of relaxation and the room temperature.

3.4 Software development and usability

To further develop the software and intervention as a whole—and assess its usability, feasibility, and safety—we conducted a prospective open-label iterative and incremental development and usability study at St. Olavs University Hospital in Trondheim, Norway, from September 2017 to June 2018. In the first part of the study, we programmed and developed an app coupled to wearable sensors for measuring muscle tension, finger temperature, and heart rate for delivering BFB treatment. The app was based on the preliminary MVP from the validation study. Thereafter, we recruited ten adolescents aged 12-18 and diagnosed with migraine according to ICHD-3 beta[123] to complete three usability testing cycles. After the usability testing, the data collected were used to develop an algorithm for processing and combining multimodal physiological data for BFB, and evaluate alternatives for sham-BFB.

3.4.1 Biofeedback setup

The setup consisted of three sensors measuring muscle tension, finger temperature, and heart rate, connected to an iPhone® with Bluetooth® (Appendix 2, Photos 1-3). To measures muscle tension and finger temperature, we used the same sensors as in the validation study, i.e. NeckSensor™ and PASPORT Skin/Surface Temperature Thermistor Probe, PS-2131. However, the Thermistor was soldered onto a NeckSensor™ for the final usability cycle. In addition, the MIO Fuse™ (Mio Global, Physical Enterprises) wristband was used to measure heart rate over the left wrist.

3.4.2 Usability evaluation

Usability evaluation and BFB app development consisted of three iterative cycles. Each cycle included the following steps: (1) app programming and design; (2)

intervention review by a neurologist, neuropsychologist, computer engineer, and medical student; and (3) usability testing by adolescent migraine sufferers.

The two first usability cycles were held as one-hour sessions in a consultation room at St. Olavs University Hospital, Trondheim. During the first cycle, the participants were first given an introduction and rationale of the treatment, and basic instructions on how to use the app. Thereafter they were asked to start the app, set up the equipment, and complete a BFB session of 10 minutes duration. Participants were not trained or instructed in relaxation or stress management techniques. In the second cycle, the participants completed three cycles of 5 minutes duration with 20 minutes of rest between each session. The final cycle was conducted as a two-week use of the app in a home setting. The participants were given sensors to work with their iPhone® at home, downloaded the app from a webpage link, and were asked to complete daily BFB sessions of 10 minutes duration with daily headache diary entries. After each usability cycle, the participants were given a comprehensive, structured and age-appropriate user evaluation. During the two first sessions, one of the investigators was present to assist the completion of the evaluation. The experiences and findings from the review of the intervention and usability testing from each cycle were used to implement changes for the next iteration of testing. Descriptive analyses of changes to the app interface and development were summarized by a simple thematic analysis.

3.4.3 Biofeedback algorithm development

The BFB algorithm was designed to give a compound feedback signal based on all three input parameters, i.e., muscle tension, finger temperature, and heart rate. Two settings of the algorithm were individually adjusted to each user to optimize feedback. Firstly, the default upper and lower measurement limits for the three physiologic

parameters were defined based on normalizing graphs of participant data. A factor was then defined to adjust the upper and the lower limits between sessions based on the participant's performance. From these upper and lower individual physiological limits, a 0–100 score for each parameter was created. Secondly, we defined an internal weighting factor for combining the three parameter-scores. The weighting was implemented to ensure that a lack of improvement in one parameter during a session and the absence of a decreasing score would still result in a moderate positive combined score. These variable factors were decided based on the usability evaluation and confirmed as suitable using a regression analysis after the final iteration.

We also developed a set of sham algorithms by manipulating the raw data. The sham algorithms were visually and statistically analyzed to evaluate if they produced sufficient disruption between the physiological data and feedback, while, importantly, still retaining masking and motivation for the user.

3.4.4 Data management

The duration (hours) of daily smartphone use, general experience with apps, and experience with wearable sensors were averaged over the three cycles for each participant. Usability evaluations were scored on a 5-point Likert interval scale, ranging from 1-"completely disagree" to 5-"completely agree." These scores were averaged over each domain for all participants. We used the principle of last observed value carried forward (LOCF) for missing data from dropouts in the usability analyses. We also analyzed complete data to serve as a comparison to the imputed data. Baseline feedback score and change in feedback score (i.e., the change from the start to the end of a session) for surface electromyographic voltage, skin temperature, and heart rate were registered for all completed sessions. Combined unweighted

"raw" scores were created using an equal 33.3% weighting for each of the three physiologic parameter scores, while BFB algorithm weighted change values were calculated using the above-described BFB algorithm. We used only complete data for analyses of physiological measurements without imputing data.

3.5 Intervention efficacy, safety and tolerability

To make initial efficacy estimates of the intervention, we conducted a prospective, 3:1 ratio randomized, sham-controlled, double-blind, pilot study at St. Olavs Hospital, Trondheim, Norway; and Oslo University Hospital, Oslo, Norway with planned enrollment from January 2019 to June 2020.

3.5.1 Study design and participants

The study comprised a four-week baseline period, followed by an eight-week intervention period with either the BFB treatment app or a sham BFB app. Twenty-three adolescents aged 12-18 and diagnosed with migraine according to the ICHD-3[131] were recruited. Eligible participants met with a consultant neurologist or pediatrician with headache expertise to confirm the migraine diagnosis. During baseline, participants were instructed to daily register maximal headache intensity, average headache intensity, functioning in daily activities, and abortive drug consumption on a paper headache diary. After a minimum 28-day baseline period, participants were randomly assigned to one of the two intervention groups by a computer-generated block-randomization list. In each block of four, participants had a 75% chance of being allocated to the BFB group and a 25% chance of being allocated to the sham group. Participants were asked to download the app and enter a 5-digit number to unlock the app. The 5-digit number was drawn by the enrolling physician sequentially from a list of 40 numbers. One random in every four numbers resulted in

downloading a sham-version of the app while the other three numbers resulted in downloading the proper BFB app. Both versions of the app looked alike, and no pattern in the 5-digit number or the randomization list could reveal which version of the app was given—this ensured blinding of participants, care providers, and investigators. Blinding of outcome assessors was not possible due to the 3:1 block-randomization. Breaking of the randomization was made after follow-up of the last participant, when the software developers revealed if the 5-digit number corresponded to the BFB or sham version of the app. During treatment, participants were encouraged to complete daily headache diary entries (the same questions as in the paper diary) and BFB sessions within the app. Participants were also encouraged to contact investigators with inquiries on how to use the equipment, report errors or shortcomings regarding both hardware and software, and take notes of any AEs and report these to the researchers. Finally, participants met with one of the researchers at the end of the two-month intervention period for evaluation, AEs questioning, and to return the equipment.

3.5.2 Interventions

The active treatment arm comprised a self-administered treatment app, including BFB training, instructions for self-delivery, and a headache diary. The app gave a push-reminder to complete a headache diary entry and a BFB session of 10 minutes duration daily (Appendix 2, Photo 4). The headache diary entry had to be completed to start a BFB session. Prior to commencing treatment, participants were given basic information on the rationale behind BFB treatment and instructed how to use the equipment and software and complete a BFB session. Sham BFB was made by adding sine-curve fluctuations to the correct feedback signal, as described in Paper III.

3.5.3 Outcomes

The primary outcome was the change in the frequency of headache days from baseline to the end of treatment. Secondary outcomes were responder rate (more than 50% reduction in headache frequency); change in maximal and average pain intensity recorded on an ordinal 4-point scale (0=no headache, 3=severe headache); change in functioning in daily activities recorded on an ordinal 4-point scale (0=no problems with daily activities, 3=severe problems with daily activities); change in the number of days with abortive drug consumption; and AEs.

Headache-related functioning in daily activities and average pain intensity was not pre-specified in the protocol and was included in the headache diary prior to enrollment as per trial guideline recommendations.[54] While the pre-specified and primary objective of this pilot study was to observe the change in outcomes *within* the BFB group only, we also conducted post-hoc comparative analyses of outcomes between the two groups. We also conducted a second post-hoc response rate analysis, changing the response threshold to 30% or greater reduction in headache frequency. Finally, we included a post-hoc analysis of mean change in BFB physiological parameters from the start to the end of sessions.

3.5.4 Safety and tolerability

The intervention's safety and tolerability were assessed in the validation study, the development and usability study, and the pilot study. In the validation study and development and usability study, AEs were assessed in structured questionnaires (Table 1, Paper II; and Appendix 3). At the end of the pilot study, participants were explicitly asked for skin reactions, nausea, and dizziness. Any additional AEs were

also recorded. All AEs in the pilot study were recorded with a physician-judged degree of seriousness and causality.

3.6 Statistical methods

3.6.1 Statistical methods for meta-analyses

For continuous outcomes, we calculated the summary mean difference (MD) with 95% confidence intervals (CI) using an inverse variance fixed-effects model. We calculated the summary odds ratio (OR) with 95% CI with a fixed-effects model for dichotomous outcomes. Owing to the low number of participants in each meta-analysis, the Mantel-Haenszel method was used for calculating dichotomous outcomes. We also calculated the number needed to treat to benefit (NNTB) based on an assumed control risk calculated from the responder rate in the control groups. Statistical heterogeneity was also calculated for each meta-analysis to evaluate the variability of intervention effects across the included studies.

3.6.2 Statistical methods for sensor validation

The means and standard deviations for the RMS values during trapezius muscle exercises and the chosen data temperature points were calculated. Systematic differences between stationary and wireless equipment were assessed with the Wilcoxon signed-rank test. MD and limits of agreement (LOA), together with Bland-Altman plots, were used as descriptive tools.[132] We calculated the intra-class correlation coefficient (ICC) with a two-way, mixed-effects consistency of agreement model. Coefficients for both individual and average agreement were presented. In addition, we calculated Lin's concordance correlation coefficient (CCC).[133-135] For the ICC and CCC analyses, the data were first transformed to meet assumptions for a two-way analysis of variance model. The data were transformed by calculating the

natural logarithm after adding 0.1 as a constant to adjust for values being close to zero. The ICC values were interpreted as suggested by Cicchetti *et al.*[136]—0.00-0.40 = unacceptable/poor; 0.41-0.60 = fair; 0.61-0.75 = good; and 0.75–1.00 = excellent.

3.6.3 Statistical methods for usability metrics and algorithm development

Data were reported as means, standard deviations (SD), medians, and interquartile ranges (IQR). Usability scores were compared between cycles with a two-tailed Wilcoxon signed-rank test and summarized with medians and IQR. We calculated the Pearson correlation coefficient to assess the association between the combined unweighted scores and BFB-algorithm scores and described the association using a two-tailed linear regression analysis. The regression analysis was applied to evaluate if the BFB-algorithm would provide a non-random and systematic improvement in feedback scores. All normality assumptions were checked by visual inspection of histograms. P-values < 0.05 were considered statistically significant.

3.6.4 Statistical methods for efficacy estimates

A priori, we planned to conduct an intention to treat (ITT) analysis of all randomized patients comparing baseline data to the last 28 days (weeks 5-8) of treatment. However, because several participants completed no BFB sessions during weeks 5-8 (and thus did not receive treatment and had no headache diary entries) and to avoid imputing data, we conducted a modified ITT (mITT) analysis. To be included in the mITT analysis, participants were required to have complete at least 7 of the planned 28 headache diary entries in weeks 5-8. Because all participants had completed at least seven BFB sessions and headache diary entries during weeks 1-4, we also included an analysis comparing baseline to weeks 1-4. We used only available data in the analyses with no imputation of data.

Within-group changes were analyzed with a two-tailed Wilcoxon signed-rank test and summarized with the MD with 95% CIs. A two-tailed Mann-Whitney-U test was used to compare changes in outcomes between the two groups, and median effect estimates with 95% CIs were produced with the Hodges-Lehman estimator. Finally, we analyzed for systematic differences in the physiological measurements between the start and end of BFB sessions with a two-tailed paired t-test summarized with MDs with 95% CIs. Normality assumptions were based on visual inspection of histograms. P-values were evaluated at the 0.05 significance level.

4 Synopsis of Main Results

This section presents the main results relating to the overall aims of the thesis. Detailed results for each study may be found in the respective papers.

4.1 Evidence for biofeedback as a pediatric migraine prophylaxis

The electronic database search of the systematic review yielded 908 records. Through the study selection process (Paper I, Figure 1), five clinical trials [52, 137-140] met all the eligibility criteria and were included in analyses. Characteristics of the included studies in summary are found in Table 2 of Paper I. Four studies qualified for comparisons of BFB versus waiting-list control.[137, 138, 140, 141] In all four studies, hand-warming BFB, with an additional behavioral therapy delivered during the same session, was compared to a waiting-list control. BFB reduced the weekly migraine frequency (MD= -1.97; 95% CI -2.72 to -1.21; $p<0.001$), compared with a waiting-list control (Paper I, Figure 3). Participants treated with BFB showed a significantly higher ($p<0.001$) proportion of responders to treatment at the end of treatment compared with waiting-list control (OR=27.71; 95% CI 6.66 to 115.35) (Paper I, Figure 4). The NNTB was 2. BFB demonstrated no adjuvant effect when combined with other behavioral treatment; neither did it have significant advantages over active treatment. Only 40% of bias judgments were deemed as low risk. Figure 2 in Paper I provide details of the risk of bias assessment for all included studies.

4.2 Sensor validation

A total of 20 healthy participants were recruited and completed the experimental sensor validation procedure. Of these, 12 were male, and the mean age was 24.7 ± 2.7 years (range 18–29 years). Table 3 in Paper II gives detailed information on agreement indices. Excellent to fair agreement was found for the SEMG sensor. The

ICC for the average of three repetitions during four different target levels ranged from 0.58 (95% CI 0.19 to 0.81) during static hold to 0.81 (95% CI 0.57 to 0.92) during MVC. The wireless sensor showed consistency in muscle tension change during moderate muscle activity (Paper II, Figure 4), and we observed no ECG artifacts in the SEMG recordings. Excellent agreement was found for the temperature sensor regarding the increase in temperature (CCC=0.90; 95% CI 0.83-0.97). A similar rise and fall in temperature for both equipment sets during the experimental procedure were seen (Paper II, Figure 6).

4.3 Usability and algorithm development

Ten participants with a mean age of 15 ± 1.6 years (range 13-17) were included in the usability study. Seven were boys. The numbers of participants in the three cycles were nine, seven and five, respectively. Five participants completed all usability cycles, and five of ten participants dropped out (50% attrition rate). A total of 72 BFB sessions were completed throughout the study. The average daily hours of smartphone usage was 3.7 ± 1.6 hours. The median value familiarity with smartphone apps was 4 (much familiarity), while the median value familiarity with wearable sensors was 1 (very little familiarity). Usability scores were consistently high but without significant difference between cycles for any of the main usability domains (Paper III, Figure 1). A Pearson's product-moment correlation analysis established a strong positive correlation between the change in unweighted scores and BFB-algorithm scores, $r(40) = 0.85$, $p<0.001$, with the following regression equation: BFB-algorithm scores = 7.41 + 0.85 x (unweighted score), $p<0.001$ (Paper III, Figure 4). Four different sham algorithms were evaluated (Paper III, Table 2), where a sham algorithm adding random sine wave fluctuations to the raw feedback signal was deemed the most suited (Figure 5, Paper III).

4.4 Efficacy, safety and tolerability

Twenty-three participants were recruited in the pilot study, 18 from St. Olavs University Hospital, and five from Oslo University Hospital. Seven participants dropped out during baseline or were excluded, and 16 patients were randomized (biofeedback n=12, sham n=4; Figure 1, Manuscript IV). Within the biofeedback group, a not statistically significant mean reduction in headache frequency of 2.92 days/month (95% CI -1.00 to 6.84, p=0.145) was observed during weeks 1-4. A not statistically significant mean reduction in headache frequency of 1.85 days/month (95% CI -2.01 to 5.72, p=0.395) was observed during weeks 5-8. No statistically significant changes in maximal headache intensity, average headache intensity, headache-related daily functioning, and abortive drug consumption were observed within the BFB group (Table 2, Manuscript IV). No statistically significant difference in change in headache frequency between the two groups was observed during weeks 1-4 (0.5 headache days/month, 95% CI -9.0 to 16.0, p=1.000), and weeks 5-8 (-1.0 headache days/month, 95% CI -9.0 to 4.0, p=0.760).

During the validation study, all participants regarded the use of wireless sensors as "safe" (n = 2) or "very safe" (n = 18). In contrast, two of the 20 participants reported undesirable harmful effects, both of them stating that the removal of the electrodes attached to the *stationary* equipment was unpleasant. In the usability study evaluations, twelve out of 20 ratings of intervention discomfort were rated as "very little discomfort," while the remaining eight were rated as "little discomfort." Out of 20 ratings of sensor discomfort, 14 were rated as "very little discomfort," five were rated as "little discomfort," and one was rated as "very much discomfort." During the pilot study, a single AE was reported by a participant experiencing a mild skin rash

related to the SEMG electrode patch. The rash lasted for a week. None of the other pre-specified AEs were reported during the pilot study.

5 Discussion

5.1 Principal findings

Our systematic review indicates that BFB is an effective intervention for pediatric migraine. The most robust finding is the meta-analysis showing that BFB can reduce the frequency of migraine when compared to a waiting-list control. Unfortunately, methodological issues such as incomplete reporting of data and risk of bias hampered the meta-analyses and decreased our confidence in the estimates. Furthermore, we have proven that WHMS for muscle tension and temperature are suited for BFB purposes. Compared to stationary equipment, the wireless temperature sensor had an almost perfect agreement regarding the change in finger temperature during relaxation, and the wireless SEMG sensor had a fair to an excellent agreement for measuring tension in the trapezius muscle. Next, we have developed a new mHealth BFB intervention for young migraine sufferers fit for self-administration showing acceptable usability scores, but limited adherence. The intervention includes a BFB-algorithm developed to give an individualized compound feedback signal based on three physiological parameters, usually considered effective in migraine prophylaxis. Finally, the intervention was evaluated in a pilot clinical trial. Limited adherence remained an issue in the pilot trial. The trial suffered from attrition, difficulties in the recruitment process, and prematurely terminated data collection due to the SARS-CoV-2 pandemic. No statistically significant reduction in headache frequency in the active treatment group or superiority over sham was observed. Still, several patients experienced a meaningful reduction in headache frequency, and the intervention was nearly free of AEs.

5.2 Biofeedback as a pediatric migraine prophylaxis

The systematic review with meta-analyses presented in this thesis is the first to attempt to estimate the pooled intervention effect specifically for BFB treatment among children and adolescents with migraine. Despite the positive findings, the number of identified studies and participants was small, and a series of methodological issues hampered the meta-analyses. Prominently, most trials used waiting-list control as comparison groups, offering uncertain estimates of specific treatment effects. Moreover, low risk of bias was found in just 40% of the scores, the remaining being deemed unclear or high. The large proportion of high and unclear risk of bias assessments further decreases the confidence in our estimates. Our findings are nonetheless in accordance with the evidence promoting the use of behavioral treatments and BFB as migraine prophylaxis.[56, 57, 59, 142] Based on this evidence, some researchers, when comparing BFB to other pediatric migraine prophylactics, argue that BFB is the better alternative as pharmacological prophylactics generally fail to show a meaningful effect.[42, 43, 45] However, this statement is based on assumptions of problematic methodologies and scientific conclusions.

Methodological issues encountered when investigating pediatric populations and non-pharmacological interventions may, in part, explain the impressive treatment effects often observed in trials of BFB. Non-specific effects, including the placebo response, contributes to a large proportion of the treatment effect in any migraine therapy.[54] The placebo effect is believed to account for up to 35% of the treatment effect in adults and up to 50% of the treatment effect in children and adolescents.[143] Moreover, the placebo response is even higher in non-pharmacological compared to drug trials.[144] Unfortunately, this high placebo response often makes it challenging to show

statistical and clinical superiority of any verum intervention, both pharmacological and non-pharmacological.[144, 145] Besides the placebo response, phenomena such as the natural fluctuations in the migraine disorder and regression to the mean, i.e., amelioration due to the passage of time, impact the treatment response seen in trials of migraine.[146] Together with the placebo response, these factors often explain the treatment effect observed in trials where superiority over baseline or waiting-list control is established,[54] which indeed is true for many trials of BFB. Trials of pharmacotherapy, on the other hand, can effectively be double-blinded and placebo-controlled. Thus, with sufficient power, such trials enable the demonstration of a specific treatment effect extending beyond the placebo response and non-specific effects—i.e., they enable the demonstration of a therapeutic gain. Such double-blind, placebo-controlled designs are difficult, maybe even impossible, for behavioral interventions—meaning that the evidence for BFB, including our findings, are likely to suffer from the above-mentioned methodological imperfections.

The omnipresent issues of non-specific effects and lack of double-blinded placebo controls in BFB trials bring us to two crucial questions: Is there truly a therapeutic gain from BFB treatment of headache, and if so, how can we best quantify it?

In the early years of BFB research, several studies were conducted to quantify the therapeutic gain. Because there are no obvious BFB-placebos, sham comparisons are often used. In a trial from 1978, the control group received sham in the form of a "positive" skin temperature feedback signal independent of the true temperature measurements but rather controlled by the investigator.[147] The control group experienced similar headache improvements as participants in the "true" BFB treatment group, suggesting that non-specific effects account for a large proportion of

the treatment response. Another study from 1981 found no difference in hand-warming vs. hand-cooling, further indicating that the BFB *per se* is unimportant.[148] Kewman observed the same in a study in 1980,[149] and again, this was true for one of the studies included in our meta-analysis.[141] In addition, BFB is traditionally administered as a treatment package—a heterogeneous composition of therapies[150, 151]—making it even more challenging to assess if the BFB in itself produces the effect. The fact that we did not observe an adjuvant effect of BFB in the meta-analysis, further supports the notion that BFB *per se* produces no effect. Nevertheless, there is a possibility that the lack of power in the "typical" small-sized sham-controlled BFB study might not be able to detect the actual therapeutic gain. Notably, some higher-powered trials indeed suggest a positive adjuvant effect of BFB,[51-53] and a meta-analysis of sham-controlled trials summarized that BFB for migraine in *adults* is possibly "as good as," or slightly superior to sham.[58] Together, these ambiguous findings illustrate that we, at present, cannot justly ascertain if there is a therapeutic gain of BFB in migraine.

However, to accurately quantify the potential therapeutic gain, one would need to compare BFB to a control intervention that perfectly mimics placebo and allows for effective double-blinding. The shams mentioned above, such as therapist-controlled feedback and hand-cooling, provide some degree of placebo control but remain insufficient as they are often single-blinded or because small inconsistencies in the intervention might reveal allocation. The same is true for the shams we evaluated in Paper III and Manuscript IV. In the pilot study, three-fourths of sham participants discovered that they were using sham, indicating that treatment allocation was unmasked. Moreover, the absence of a difference in headache frequency reduction between the two groups in the pilot study further suggests that the observed treatment

effect may be attributed to placebo and regression to the mean. To finely dissect the treatment effect components, one should conduct a sufficiently powered three-armed study with verum, placebo/sham, and waiting-list. Such a three-armed study would be able to demonstrate the specific effect/therapeutic gain (verum minus placebo), the placebo effect (placebo minus waiting-list), and the remaining non-specific effects, such as regression to the mean (effect in no treatment arm).[79, 152]

To sum up, based on available evidence and with the methodological considerations in mind, evidence for a therapeutic gain—at least to a clinically meaningful degree—in BFB treatment of *adult* migraine appears weak. Whether the same is true for pediatric migraine remains to be investigated. Meanwhile, we can safely say that BFB for pediatric migraine is better than doing nothing, improves migraine burden, and is free from adverse effects—we cannot tell if it is solely a placebo effect.

5.3 Feasibility of using wearable sensors for biofeedback

The validation study aimed to provide a proof-of-concept for using a smartphone and WHMS for BFB purposes. We chose to investigate the validity of temperature and SMEG sensors because our meta-analytic findings and previous studies indicate that these are especially effective in children and adolescents.[153, 154] In addition, several studies have been conducted indicating that heart rate sensors give satisfactory signals.[73, 74, 93] Although the SEMG sensor did not demonstrate excellent agreement in all analyses, several factors indicate that perfect absolute agreement is not a prerequisite for muscle tension BFB. Studies suggest that users most likely will not be able to decrease their muscle tension throughout the entire duration of a BFB session,[155] and that the feedback itself is more important than lowering the muscle tension.[156] As discussed above, the same is even true for temperature feedback, where

a change in temperature does not necessarily predict headache improvement.[87, 148] Taken together with our findings that all participants in the validation study had similar and consistent changes in muscle tension through the sets of exercises, one can argue that detecting a change is more important than absolute values. Based on our validation and previous findings of heart rate sensor validity, we suggest that the most common BFB parameters, i.e., finger temperature, muscle tension, and heart rate, may be measured with WHMS. Even though the use of different temperature and SEMG sensors would not yield identical results, our approach seems to have provided a proof-of-concept.

5.4 Development of Mi-Insight

Through the development process of Mi-Insight, we have attempted to overcome the challenges that often make new mHealth fall short of the mark.[112-118] Our development process was conducted according to relevant guidelines and recommendations: Firstly, prior to initiation of development we set out to review the evidence base for BFB in pediatric migraine adhering to current standards,[121, 122] and to gain experience with traditional BFB setup. Thereafter we made a validity and feasibility assessment[120] of using wearable sensors for monitoring BFB parameters of interest. Finally, we used an iterative and incremental design to test and assess the intervention's usability and gain valuable experience prior to efficacy trials. Both the validation study and the usability study aided us in several design decisions to improve the development of the app. Especially throughout the usability study, we aimed to assess and improve the understandability, learnability, operability, and attractiveness of the app, which is essential to obtaining satisfactory adherence.[157] We believe that this rigorous approach may yield results that are more fruitful and suited for clinical trials, and may be considered as similar to phase I-II development of new

drug treatments.[158] Within other medical fields, similar studies have been carried out to assess feasibility, usability, and draw initial efficacy measures of mHealth and WHMS applications.[98, 101, 159] All of these studies have detected and addressed several issues regarding the feasibility and usability of the applications. This shows that such studies aid in improving the development of mHealth applications.

5.5 Mi-Insight as a prophylactic pediatric migraine intervention

Several factors make Mi-Insight unique as a novel BFB intervention. Firstly, the intervention allows for the widespread and inexpensive administration of BFB. The ease of access and therapist-independence may result in increased population coverage, and we have estimated an annual consumer cost reduction of at least 64% compared to traditional BFB. Secondly, it uses and combines three BFB parameters, whereas traditional treatment uses one parameter.[57, 58] Based on the three parameters, it uses an algorithm to combine the scores and make individualized feedback. This algorithm was intended to overcome challenges that might arise from using several feedback parameters. For example, if a user improves one feedback parameter, and worsens another, the algorithm will fade out the latter throughout a session. This function was implemented with the belief that such a feedback will potentially embrace potential users and reduce, or even omit, the necessity of a therapist to aid during the session. Thirdly, the application displayed good usability and functionality, which, together with the algorithm, is supposed to make it more therapist-independent. As pointed out earlier, one of the main challenges for widespread application of BFB seems to be its time- and cost-consuming nature, something we have attempted to overcome. Despite the observed good usability and functionality, treatment adherence remained a problem in the usability and pilot studies. Addressing the issue of limited adherence is of paramount importance in future iterations and

studies of the intervention. Finally, the application includes a quick and easy to complete headache diary, with questions adhering to guidelines making it easy to use for research.[54]

Although Mi-Insight was designed to be easily accessible, therapist-independent, and easy to use, the highly minimalistic intervention likely resulted in a suboptimal treatment effect. When comparing the pilot trial's effect estimates to the meta-analytical findings of paper I, Mi-Insight comes out as clearly inferior. Behavioral treatment effects are typically in the range of 35-50% reduction in headache frequency,[160] whereas Mi-Insight produced an approximate 20% reduction. Several factors may contribute to understanding why we observed a limited treatment effect that was not statistically significant.

Firstly, the nature of the BFB intervention used in the pilot study was quite different from the traditional BFB. As evident in the systematic review, BFB treatment is usually administered as a treatment package with regular therapist-contact sessions and combined with adjunctive behavioral therapies such as relaxation and stress management. The therapist aids the user to achieve the "correct" self-control, and the treatment package promotes several of the non-specific effects seen with BFB, such as expectancy, conditioning, and regular contact and procedural repetitions.[79] In the pilot study, participants were given a very minimalistic intervention, only consisting of a brief introduction to the concept of BFB and brief instructions on how to use the equipment and perform a session. Thereafter, learning self-control was entirely based on operant conditioning from the feedback instruments. Participants appeared to quickly learn to increase temperature and lower muscle tension within biofeedback sessions. However, there was no evident improvement across sessions, and we also

observed a paradoxical increase in heart rate within sessions. A real-world therapist could potentially have helped to modulate the self-control towards the assumed "correct" state, which is hypothesized to predict positive outcomes.[66] Moreover, the absence of therapist contact and adjunctive therapies may have led to a reduction in the non-specific effects, further explaining the limited treatment effect.[79]At any rate, self-control is only a surrogate outcome and does not necessarily translate into an improvement in migraine burden, even though it is hypothesized that a higher degree of achieved self-control predicts greater headache reductions.[66]

Secondly, the adherence rate to BFB treatment remained low in the pilot study, potentially reducing treatment effects. A systematic review found that the typical adherence to behavioral interventions among children was 52% to 86%.[161] These figures are higher than what we observed, especially in weeks 5-8 of treatment. There are no clear estimates of how much adherence influences treatment outcome, but lower adherence is believed to undermine behavioral intervention's efficacy.[162] A study of app-based progressive muscle relaxation as a prophylactic treatment for migraine in adults found that high adherence users (defined as two or more sessions per week) had a significantly greater reduction in headache frequency than low adherence users,[163] supporting our results from the pilot study. The adherence observed with Mi-Insight was likely low both because the users perceived the daily routine of mandatory diary entries and BFB sessions as unengaging; and because the limited therapist contact did not promote adherence.

Finally, the issues with the use of sham-control and the identification of therapeutic gains in studies of BFB re-emerges in light of the pilot study. The sham group experienced a reduction in headache frequency quite similar to the BFB group,

suggesting that the improvement in all clinical outcomes is caused by placebo and regression to the mean, and supporting the notion that there is no significant therapeutic gain.[57] Still, this argument should be considered carefully, as the power in the pilot study was insufficient to detect a therapeutic gain. Even if there is no true specific effect of the intervention we have created, one could argue that it has an advantage over pharmacological prophylaxis (recall the uncertainty around therapeutic gains also for pharmacological prophylaxes), simply in its absence of AEs. This would be in line with the reasoning about similar therapies (with the same methodological issues as found in studies of BFB), such as CBT, that has been suggested as a first-line treatment for pediatric migraine.[164] On the other hand, this argument could be considered unethical, as the patient is required to invest time and money into an "ineffective" treatment, regardless of the absence of AEs.

5.6 Future studies and iterations of app-based biofeedback treatment

Despite the negative findings of the pilot trial, we believe there is a rationale for continued research. All the unique factors discussed in the previous section—including the potential for widespread and inexpensive use, the novel combination of BFB modalities, apparent good usability and functionality, and research-centered design—warrants continued research. Moreover, the fact that the highly minimalistic nature of the intervention produced approximately 20% reduction in headache frequency should be considered remarkable regardless of the attributable proportions of the effect. Held together with the highly beneficial AE profile, the intervention has the potential to be clinically non-inferior to the most commonly used prophylactic treatments for pediatric migraine.

Several measures should be considered for future iterations and studies of app-based BFB treatment. The intervention should include more comprehensive instructions, guidance during BFB sessions, and even adjunctive therapies such as relaxation. Such features should be intelligently implemented into the app to ensure therapist-independence and may facilitate the treatment packages used in traditional BFB. In addition, measures should be taken to keep adherence high through regular reminders, motivation, and gamification.[165] Next, the use of a sham comparator group should be carefully considered. As previously discussed, it is difficult to create a BFB sham that accurately mimics proper placebo effects. It might be advisable to abandon the hunt for the *therapeutic gain* in BFB treatment of pediatric migraine as it is likely small. Almost all pediatric migraine prophylactics fail to demonstrate convincing evidence of effectiveness,[43] and a more rewarding approach might be to demonstrate non-inferiority compared to the most commonly used prophylactic medications. Evidence of non-inferiority would construct a clinically strong argument for using BFB—though lacking the scientific rigor of placebo-superiority—simply because the AE profile is undoubtedly beneficial while the treatment is at least "as good as" the currently used treatment options. Yet, we would still be unsure if either is more than mere placebo and non-specific effects. In any event, the study should be sufficiently powered to detect small treatment effects or non-inferiority. Another potential improvement of the app and diary would be to include recommendations on lifestyle habits[166] and registration of potential triggers and premonitory symptoms.[167] These could further be combined with external data and physiological measurements and, through machine learning, aid in predicting migraine attacks to facilitate lifestyle measures and pre-emptive treatment strategies.[168, 169]

5.7 Limitations

5.7.1 Pooled analyses

A limitation that often hampers pooled analyses is the clinical heterogeneity of interventions.[121] In the comparison of BFB with waiting-list control, we lumped together the somewhat heterogeneous intervention packages, assuming that the analyses might give information on the intervention effect of BFB in pediatric migraine—again raising the question of what part of the observed package effect may be attributed to the BFB. The analyses are also limited by the fact that children and adolescents were regarded as one group,[170] while biological and psychological differences between these age groups could hamper interpretation of the results. In addition, one must always keep in mind that meta-analyses are no better than the trials put into them, and considering the potential bias in the included studies, our findings must be interpreted with similar caution.

5.7.2 Physiological measurements

Several factors have limited the preciseness of the physiological measurements. Because stationary and wireless equipment not could be placed on the same spot over the trapezius muscle, EMG crosstalk may have occurred, and muscle contraction exercises performed by untrained participants may additionally have resulted in movement artifacts, and suboptimal and varying performance.[77] This combined with variations in individual human anatomic properties[78] may have limited the precision of our measurements and contributed to a larger degree of inter-individual differences, thus lower SEMG agreement. Likewise, the placement of the two temperature sensors next to each other on the finger might have led to differences in

measurements, but the analyses did indeed show excellent agreement for the temperature sensors.

5.7.3 Usability testing

Several factors make us reluctant to draw firm conclusions based on the usability study. The questionnaire was based on common surveys and recommendations for mHealth app assessments but had not undergone formal validation prior to the usability study. The questionnaire is also highly susceptible to response bias,[171] i.e., participants responding inaccurately to survey questions. A category of response bias termed acquiescence bias, in which participants automatically endorse statements to please the interviewer, may especially have been present.[172] This could, in part, explain the high usability scorings in the initial cycle and the trend towards lower usability scorings in the home-testing cycle. Moreover, the two first rounds of the evaluation were conducted in a controlled environment that is not fully representative of the intended usage, and the home testing session was made over 14 days instead of the recommended 28-day period for headache assessments.[54, 173] Together, this adds some uncertainty concerning the adherence to the intervention. Thirdly, the study had a moderate sample size and suffered from attrition. This may represent poor usability and decrease the confidence in our findings.

5.7.4 Sample sizes

Another general limitation of this thesis is the limited sample sizes of the studies. Traditionally, a sample of 15–20 participants is considered sufficient for validation studies.[174] However, the use of more precise calculations of sample sizes has been suggested.[175] Therefore, we used a confidence interval estimation model suggested by Bonett[125] to determine the minimum sample size required prior to recruiting

participants. Due to the inter-individual variation in our findings, the analyses would possibly have benefited from having a larger sample size because we did not obtain the pre-defined confidence intervals for all analyses. In addition, one may argue that a convenience sample is unsuited, which is indeed true when making inferences about effect size estimates. But for other cases, such as validation, this does not necessarily hold.[176]

On the other hand, a sample size of about five participants has been deemed sufficient to uncover the majority of usability problems.[177, 178] Nonetheless, recent studies have emphasized the need for larger sample sizes and customized sample sizes to individual studies.[179, 180] In light of these studies, we chose a sample size of ten persons stratified across the adolescent age range to ensure uncovering of essential usability problems while also receiving evaluations from the whole heterogeneous age spectrum. Unfortunately, the study suffered from attrition and thereby a limited sample size in the final usability cycle and LOCF analyses.

The pilot trial also suffered greatly from a limited sample size. This clearly reduced the precision of our estimates and limited interpretability of clinical outcomes both in the BFB and sham groups. In addition to the small sample size, the study suffered from attrition and missing data. Several participants were excluded or declined to participate, and the overall adherence was low, with resultant missing data, which further decrease confidence in our estimates.

6 Conclusion

BFB, when delivered together with other behavioral therapies, seems to be effective in reducing the frequency of migraine in the pediatric population compared to waiting-list control. This evidence is based on a few small studies, in which a series of methodological issues hampered the meta-analyses. The treatment effect observed with BFB appears to be mainly non-specific, and it is uncertain if there is any therapeutic gain of clinical significance.

In spite of the uncertainties around the therapeutic gain and specific effects of BFB, this thesis provides some insights into the feasibility, safety, and efficacy of a self-administered therapist-independent BFB app. WHMS are suited for monitoring physiological parameters that are of interest in a BFB setting, and our findings indicate that a mHealth app coupled with WHMS is feasible and usable for delivering BFB treatment. Efficacy measures for such a self-administered and therapist-independent mHealth BFB intervention remain uncertain, and our underpowered and methodologically limited pilot trial failed to demonstrate a convincing specific treatment effect. On the positive side, the intervention was nearly free of AEs, and the findings support further research. Future iterations of the intervention should include a more comprehensive intervention and ensure increased adherence through means such as gamification. Futures studies of the intervention should strongly consider using an active comparison group and be powered to detect the likely small but potentially meaningful treatment effects.

7 References

1. Headache Classification Committee of the International Headache Society (IHS) The International Classification of Headache Disorders, 3rd edition. *Cephalalgia* 2018; 38: 1-211. DOI: 10.1177/0333102417738202.
2. Vos T, Allen C, Arora M, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* 2016; 388: 1545-1602.
3. Stovner LJ, Nichols E, Steiner TJ, et al. Headache in the Global Burden of Disease (GBD) Studies. *Societal Impact of Headache*. Springer, 2019, pp.105-125.
4. Stovner L, Hagen K, Jensen R, et al. The global burden of headache: a documentation of headache prevalence and disability worldwide. *Cephalalgia* 2007; 27: 193-210. 2007/03/27. DOI: 10.1111/j.1468-2982.2007.01288.x.
5. Charles A. Migraine: a brain state. *Current opinion in neurology* 2013; 26: 235-239.
6. Dodick DW. A phase‐by‐phase review of migraine pathophysiology. *Headache: The Journal of Head and Face Pain* 2018; 58: 4-16.
7. Goadsby PJ, Holland PR, Martins-Oliveira M, et al. Pathophysiology of migraine: a disorder of sensory processing. *Physiological reviews* 2017; 97: 553-622.
8. Chasman DI, Schürks M, Anttila V, et al. Genome-wide association study reveals three susceptibility loci for common migraine in the general population. *Nature genetics* 2011; 43: 695-698.
9. Freilinger T, Anttila V, de Vries B, et al. Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nature genetics* 2012; 44: 777-782.
10. Scheffer M, van den Berg A and Ferrari MD. Migraine strikes as neuronal excitability reaches a tipping point. *PloS one* 2013; 8: e72514.
11. Lance JW and Anthony M. Some clinical aspects of migraine: a prospective survey of 500 patients. *Archives of Neurology* 1966; 15: 356-361.
12. Honkasalo ML, Kaprio J, Winter T, et al. Migraine and concomitant symptoms among 8167 adult twin pairs. *Headache: The Journal of Head and Face Pain* 1995; 35: 70-78.
13. Arbab MA, Wiklund L and Svendgaard NA. Origin and distribution of cerebral vascular innervation from superior cervical, trigeminal and spinal ganglia investigated with retrograde and anterograde WGA-HRP tracing in the rat. *Neuroscience* 1986; 19: 695-708. 1986/11/01. DOI: 10.1016/0306-4522(86)90293-9.
14. Bartsch T and Goadsby PJ. Increased responses in trigeminocervical nociceptive neurons to cervical input after stimulation of the dura mater. *Brain* 2003; 126: 1801-1813. 2003/06/25. DOI: 10.1093/brain/awg190.
15. Goadsby P, Edvinsson L and Ekman R. Release of vasoactive peptides in the extracerebral circulation of humans and the cat during activation of the trigeminovascular system. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 1988; 23: 193-196.
16. Messlinger K, Fischer MJ and Lennerz JK. Neuropeptide effects in the trigeminal system: pathophysiology and clinical relevance in migraine. *The Keio journal of medicine* 2011; 60: 82-89.
17. Lassen LH, Haderslev PA, Jacobsen VB, et al. CGRP may play a causative role in migraine. *Cephalalgia* 2002; 22: 54-61. 2002/05/08. DOI: 10.1046/j.1468-2982.2002.00310.x.

18. Strassman AM, Raymond SA and Burstein R. Sensitization of meningeal sensory neurons and the origin of headaches. *Nature* 1996; 384: 560-564. 1996/12/12. DOI: 10.1038/384560a0.
19. Lauritzen M. Pathophysiology of the migraine aura: the spreading depression theory. *Brain* 1994; 117: 199-210.
20. Miglis MG. Migraine and Autonomic Dysfunction: Which Is the Horse and Which Is the Jockey? *Current pain and headache reports* 2018; 22: 19-19. DOI: 10.1007/s11916-018-0671-y.
21. Barbanti P, Aurilia C, Dall'Armi V, et al. The phenotype of migraine with unilateral cranial autonomic symptoms documents increased peripheral and central trigeminal sensitization. A case series of 757 patients. *Cephalalgia* 2016; 36: 1334-1340. DOI: 10.1177/0333102416630579.
22. Möller M and May A. The unique role of the trigeminal autonomic reflex and its modulation in primary headache disorders. *Current opinion in neurology* 2019; 32: 438-442.
23. Gelfand AA, Reider AC and Goadsby PJ. Cranial autonomic symptoms in pediatric migraine are the rule, not the exception. *Neurology* 2013; 81: 431-436. 2013/06/28. DOI: 10.1212/WNL.0b013e31829d872a.
24. Raieli V, Giordano G, Spitaleri C, et al. Migraine and cranial autonomic symptoms in children and adolescents: a clinical study. *Journal of child neurology* 2015; 30: 182-186. 2014/05/19. DOI: 10.1177/0883073814535494.
25. Headache. Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. *Cephalalgia* 1988; 8 Suppl 7: 1-96. 1988/01/01.
26. The International Classification of Headache Disorders: 2nd edition. *Cephalalgia* 2004; 24 Suppl 1: 9-160. 2004/02/26.
27. Wober-Bingol C. Epidemiology of migraine and headache in children and adolescents. *Curr Pain Headache Rep* 2013; 17: 341. 2013/05/24. DOI: 10.1007/s11916-013-0341-z.
28. Abu‐Arafeh I, Razak S, Sivaraman B, et al. Prevalence of headache and migraine in children and adolescents: a systematic review of population‐based studies. *Developmental medicine & child neurology* 2010; 52: 1088-1097.
29. Victor TW, Hu X, Campbell JC, et al. Migraine prevalence by age and sex in the United States: a life-span study. *Cephalalgia* 2010; 30: 1065-1072. 2010/08/18. DOI: 10.1177/0333102409355601.
30. Tarasco V, Grasso G, Versace A, et al. Epidemiological and clinical features of migraine in the pediatric population of Northern Italy. *Cephalalgia* 2016; 36: 510-517. 2015/08/06. DOI: 10.1177/0333102415598758.
31. Sillanpaa M and Saarinen MM. Long term outcome of childhood onset headache: A prospective community study. *Cephalalgia* 2018; 38: 1159-1166. 2017/08/23. DOI: 10.1177/0333102417727536.
32. Krogh AB, Larsson B and Linde M. Prevalence and disability of headache among Norwegian adolescents: A cross-sectional school-based study. *Cephalalgia* 2015. DOI: 10.1177/0333102415573512.
33. Metsahonkala L, Sillanpaa M and Tuominen J. Headache diary in the diagnosis of childhood migraine. *Headache* 1997; 37: 240-244. 1997/04/01.
34. Richer L, Billinghurst L, Linsdell MA, et al. Drugs for the acute treatment of migraine in children and adolescents. *Cochrane Database Syst Rev* 2016; 4: Cd005220. 2016/04/20. DOI: 10.1002/14651858.CD005220.pub2.
35. Eiland LS and Hunt MO. The use of triptans for pediatric migraines. *Pediatric Drugs* 2010; 12: 379-389.

36. Goadsby PJ and Edvinsson L. The trigeminovascular system and migraine: studies characterizing cerebrovascular and neuropeptide changes seen in humans and cats. *Annals of neurology* 1993; 33: 48-56.
37. Tepper SJ, Rapoport AM and Sheftell FD. Mechanisms of action of the 5-HT1B/1D receptor agonists. *Archives of neurology* 2002; 59: 1084-1088.
38. Derosier FJ, Lewis D, Hershey AD, et al. Randomized trial of sumatriptan and naproxen sodium combination in adolescent migraine. *Pediatrics* 2012: peds. 2011-2455.
39. Damen L, Bruijn JK, Verhagen AP, et al. Symptomatic treatment of migraine in children: a systematic review of medication trials. *Pediatrics* 2005; 116: e295-302. 2005/08/03. DOI: 10.1542/peds.2004-2742.
40. Lewis D, Ashwal S, Hershey Ao, et al. Practice Parameter: Pharmacological treatment of migraine headache in children and adolescents Report of the American Academy of Neurology Quality Standards Subcommittee and the Practice Committee of the Child Neurology Society. *Neurology* 2004; 63: 2215-2224.
41. Chiappedi M and Balottin U. Medication overuse headache in children and adolescents. *Current pain and headache reports* 2014; 18: 404.
42. Shamliyan TA, Kane RL, Ramakrishnan R, et al. Episodic migraines in children: limited evidence on preventive pharmacological treatments. *Journal of child neurology* 2013; 28: 1320-1341. 2013/06/12. DOI: 10.1177/0883073813488659.
43. El-Chammas K, Keyes J, Thompson N, et al. Pharmacologic treatment of pediatric headaches: a meta-analysis. *JAMA pediatrics* 2013; 167: 250-258. 2013/01/30. DOI: 10.1001/jamapediatrics.2013.508.
44. Oskoui M, Pringsheim T, Billinghurst L, et al. Practice guideline update summary: Pharmacologic treatment for pediatric migraine prevention: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the American Headache Society. *Neurology* 2019; 93: 500-509. 2019/08/16. DOI: 10.1212/wnl.0000000000008105.
45. Powers SW, Coffey CS, Chamberlin LA, et al. Trial of Amitriptyline, Topiramate, and Placebo for Pediatric Migraine. *The New England journal of medicine* 2017; 376: 115-124. 2016/10/28. DOI: 10.1056/NEJMoa1610384.
46. O'Brien HL, Kabbouche MA and Hershey AD. Treating pediatric migraine: an expert opinion. *Expert opinion on pharmacotherapy* 2012; 13: 959-966. 2012/04/17. DOI: 10.1517/14656566.2012.677434.
47. Palermo TM, Eccleston C, Lewandowski AS, et al. Randomized controlled trials of psychological therapies for management of chronic pain in children and adolescents: an updated meta-analytic review. *Pain* 2010; 148: 387-397. DOI: 10.1016/j.pain.2009.10.004.
48. Trautmann E, Lackschewitz H and Kroner-Herwig B. Psychological treatment of recurrent headache in children and adolescents--a meta-analysis. *Cephalalgia* 2006; 26: 1411-1426. DOI: 10.1111/j.1468-2982.2006.01226.x.
49. Fisher E, Heathcote L, Palermo TM, et al. Systematic review and meta-analysis of psychological therapies for children with chronic pain. *Journal of pediatric psychology* 2014; 39: 763-782.
50. Fisher E, Law E, Dudeney J, et al. Psychological therapies for the management of chronic and recurrent pain in children and adolescents. *Cochrane Database of Systematic Reviews* 2018.
51. Powers SW, Kashikar-Zuck SM, Allen JR, et al. Cognitive behavioral therapy plus amitriptyline for chronic migraine in children and adolescents: a randomized

clinical trial. *Jama* 2013; 310: 2622-2630. 2013/12/26. DOI: 10.1001/jama.2013.282533.
52. Sartory G, Müller B, Metsch J, et al. A comparison of psychological and pharmacological treatment of pediatric migraine. *Behaviour research and therapy 36(12)*(1998).
53. Holroyd KA, Cottrell CK, O'Donnell FJ, et al. Effect of preventive (β blocker) treatment, behavioural migraine management, or their combination on outcomes of optimised acute treatment in frequent migraine: randomised controlled trial. *BMJ (Clinical research ed)* 2010; 341: c4871.
54. Tfelt-Hansen P, Pascual J, Ramadan N, et al. Guidelines for controlled trials of drugs in migraine: third edition. A guide for investigators. *Cephalalgia* 2012; 32: 6-38.
55. Shapiro D and Schwartz GE. Biofeedback and visceral learning: clinical applications. *Seminars in psychiatry* 1972; 4: 171-184. 1972/05/01.
56. Kondo K, Noonan KM, Freeman M, et al. Efficacy of biofeedback for medical conditions: An evidence map. *Journal of general internal medicine* 2019: 1-11.
57. Nestoriuc Y and Martin A. Efficacy of biofeedback for migraine: a meta-analysis. *Pain* 2007; 128: 111-127. DOI: 10.1016/j.pain.2006.09.007.
58. Nestoriuc Y, Martin A, Rief W, et al. Biofeedback treatment for headache disorders: a comprehensive efficacy review. *Appl Psychophysiol Biofeedback* 2008; 33: 125-140. 2008/08/30. DOI: 10.1007/s10484-008-9060-3.
59. Penzien DB, Irby MB, Smitherman TA, et al. Well-Established and Empirically Supported Behavioral Treatments for Migraine. *Curr Pain Headache Rep* 2015; 19: 34. 2015/06/13. DOI: 10.1007/s11916-015-0500-5.
60. Hermann C, Blanchard EB and Flor H. Biofeedback treatment for pediatric migraine: Prediction of treatment outcome. *Journal of Consulting and Clinical Psychology* 1997; 65: 611-616. DOI: 10.1037/0022-006X.65.4.611.
61. Burke EJ and Andrasik F. Home- vs. clinic-based biofeedback treatment for pediatric migraine: Results of treatment through one-year follow-up. *Headache: The Journal of Head and Face Pain* 1989; 29: 434-440.
62. Hales J. Skin arteriovenous anastomoses, their control and role in thermoregulation. 1985.
63. Lin HP, Lin HY, Lin WL, et al. Effects of stress, depression, and their interaction on heart rate, skin conductance, finger temperature, and respiratory rate: sympathetic‐parasympathetic hypothesis of stress and depression. *Journal of clinical psychology* 2011; 67: 1080-1091.
64. Kistler A, Mariauzouls C and von Berlepsch K. Fingertip temperature as an indicator for sympathetic responses. *International Journal of Psychophysiology* 1998; 29: 35-41.
65. Mathew RJ, Largen JW, Dobbins K, et al. Biofeedback control of skin temperature and cerebral blood flow in migraine. *Headache: The Journal of Head and Face Pain* 1980; 20: 19-28.
66. Lisspers J, Öst L-G and Skagerberg B. Clinical effects of biofeedback treatment in migraine: the relation to achieved self-control and pretreatment predictors. *Cognitive Behaviour Therapy* 1992; 21: 171-190.
67. Hertzman AB. The blood supply of various skin areas as estimated by the photoelectric plethysmograph. *American Journal of Physiology-Legacy Content* 1938; 124: 328-340.
68. Murray WB and Foster PA. The peripheral pulse wave: information overlooked. *Journal of clinical monitoring* 1996; 12: 365-377.

69. Tamura T, Maeda Y, Sekine M, et al. Wearable photoplethysmographic sensors—past and present. *Electronics* 2014; 3: 282-302.
70. Meada Y, Sekine M and Tamura T. The advantage of green reflected photoplethysmograph. *J Med Syst* 2011; 35: 829-834.
71. Cui W, Ostrander LE and Lee BY. In vivo reflectance of blood and tissue as a function of light wavelength. *IEEE transactions on biomedical engineering* 1990; 37: 632-639.
72. Shelley K and Shelley S. Pulse oximeter waveform: photoelectric plethysmography. *Clinical Monitoring, Carol Lake, R Hines, and C Blitt, Eds: WB Saunders Company* 2001: 420-428.
73. Gillinov S, Etiwy M, Wang R, et al. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Medicine and science in sports and exercise* 2017; 49: 1697-1703.
74. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. *Jama cardiology* 2017; 2: 104-106.
75. Cheung CC, Krahn AD and Andrade JG. The Emerging Role of Wearable Technologies in Arrhythmia Detection. *Canadian Journal of Cardiology* 2018.
76. Clancy EA, Morin EL and Merletti R. Sampling, noise-reduction and amplitude estimation issues in surface electromyography. *Journal of electromyography and kinesiology* 2002; 12: 1-16.
77. Merletti R and Di Torino P. Standards for reporting EMG data. *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology* 1999; 9: 3-4.
78. Criswell E. *Cram's introduction to surface electromyography*. 2nd edition ed. Sudbury, Ma: Jones & Bartlett Publishers, 2010.
79. Autret A, Valade D and Debiais S. Placebo and other psychological interactions in headache treatment. *The journal of headache and pain* 2012; 13: 191-198. 2012/03/01. DOI: 10.1007/s10194-012-0422-0.
80. Miller NE. Biofeedback and visceral learning. *Annual review of psychology* 1978; 29: 373-404.
81. Harris AH, Gilliam WJ, Findley JD, et al. Instrumental conditioning of large-magnitude, daily, 12-hour blood pressure elevations in the baboon. *Science* 1973; 182: 175-177.
82. Hayduk AW. Increasing hand efficiency at cold temperatures by training hand vasodilation with a classical conditioning-biofeedback overlap design. *Biofeedback and self-regulation* 1980; 5: 307-326. DOI: 10.1007/bf00999807.
83. Bild R and Adams HE. Modification of migraine headaches by cephalic blood volume pulse and EMG biofeedback. *Journal of Consulting and Clinical Psychology* 1980; 48: 51.
84. Lehrer P and Eddie D. Dynamic processes in regulation and some implications for biofeedback and biobehavioral interventions. *Appl Psychophysiol Biofeedback* 2013; 38: 143-155. DOI: 10.1007/s10484-013-9217-6.
85. Siniatchkin M, Hierundar A, Kropp P, et al. Self-regulation of slow cortical potentials in children with migraine: An exploratory study. *Applied Psychophysiology Biofeedback* 2000; 25: 13-32. DOI: 10.1023/A:1009581321624.
86. Hermann C and Blanchard EB. Biofeedback in the treatment of headache and other childhood pain. *Applied Psychophysiology and Biofeedback* 2002; 27: 143-162.
87. Morrill B and Blanchard EB. Two studies of the potential mechanisms of action in the thermal biofeedback treatment of vascular headache. *Headache: The Journal of Head and Face Pain* 1989; 29: 169-176.

88. Yuan H and Silberstein SD. Vagus Nerve and Vagus Nerve Stimulation, a Comprehensive Review: Part III. *Headache* 2016; 56: 479-490. 2015/09/15. DOI: 10.1111/head.12649.
89. Lehrer PM and Gevirtz R. Heart rate variability biofeedback: how and why does it work? *Frontiers in psychology* 2014; 5: 756. 2014/08/08. DOI: 10.3389/fpsyg.2014.00756.
90. Huang C, Gevirtz R, Onton J, et al. Investigation of vagal afferent functioning using the heartbeat event related potential. *International Journal of Psychophysiology* 2018; 131: 113-123.
91. Ciancarelli I, Tozzi-Ciancarelli M, Spacca G, et al. Relationship between biofeedback and oxidative stress in patients with chronic migraine. *Cephalalgia* 2007; 27: 1136-1141.
92. Birnie KA, Chambers CT and Spellman CM. Mechanisms of distraction in acute pain perception and modulation. *Pain* 2017; 158: 1012-1013. 2017/05/18. DOI: 10.1097/j.pain.0000000000000913.
93. Kay M, Santos J and Takane M. mHealth: New horizons for health through mobile technologies. *World Health Organization* 2011; 3: 66-71.
94. Pantelopoulos A and Bourbakis NG. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 2010; 40: 1-12. DOI: 10.1109/TSMCC.2009.2032660.
95. Dobkin BH and Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. *Neurorehabilitation and neural repair* 2011; 25: 788-798.
96. Dunn J, Runge R and Snyder M. Wearables and the medical revolution. *Per Med* 2018; 15: 429-448. 2018/09/28. DOI: 10.2217/pme-2018-0044.
97. Cafazzo JA, Casselman M, Hamming N, et al. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. *Journal of medical Internet research* 2012; 14: e70.
98. Uddin AA, Morita PP, Tallevi K, et al. Development of a Wearable Cardiac Monitoring System for Behavioral Neurocardiac Training: A Usability Study. *JMIR Mhealth Uhealth* 2016; 4: e45. 2016/04/24. DOI: 10.2196/mhealth.5288.
99. Brunelli D, Farella E, Rocchi L, et al. Bio-feedback system for rehabilitation based on a wireless body area network. In: *Pervasive Computing and Communications Workshops, 2006 PerCom Workshops 2006 Fourth Annual IEEE International Conference on* 2006, pp.5 pp.-531. IEEE.
100. Mirelman A, Herman T, Nicolai S, et al. Audio-biofeedback training for posture and balance in patients with Parkinson's disease. *Journal of neuroengineering and rehabilitation* 2011; 8: 35.
101. Ginis P, Nieuwboer A, Dorfman M, et al. Feasibility and effects of home-based smartphone-delivered automated feedback training for gait in people with Parkinson's disease: A pilot randomized controlled trial. *Parkinsonism & related disorders* 2016; 22: 28-34.
102. Imani S, Bandodkar AJ, Mohan AV, et al. A wearable chemical-electrophysiological hybrid biosensing system for real-time health and fitness monitoring. *Nature communications* 2016; 7.
103. De la Vega R and Miró J. mHealth: a strategic field without a solid scientific soul: a systematic review of pain-related apps. *PloS one* 2014; 9: e101312.

104. Lalloo C, Jibb LA, Rivera J, et al. “There’sa pain app for that”: Review of patient-targeted smartphone applications for pain management. *The Clinical journal of pain* 2015; 31: 557-563.
105. Huckvale K, Morrison C, Ouyang J, et al. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. *BMC medicine* 2015; 13: 58.
106. Neubeck L, Lowres N, Benjamin EJ, et al. The mobile revolution—using smartphone apps to prevent cardiovascular disease. *Nature Reviews Cardiology* 2015; 12: 350.
107. Wen D, Zhang X, Liu X, et al. Evaluating the Consistency of Current Mainstream Wearable Devices in Health Monitoring: A Comparison Under Free-Living Conditions. *Journal of medical Internet research* 2017; 19.
108. Garbarino M, Lai M, Bender D, et al. Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In: *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on* 2014, pp.39-42. IEEE.
109. Ollander S, Godin C, Campagne A, et al. A comparison of wearable and stationary sensors for stress detection. In: *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on* 2016, pp.004362-004366. IEEE.
110. Mosadeghi-Nik M, Askari MS and Fatehi F. Mobile health (mHealth) for headache disorders: A review of the evidence base. *Journal of Telemedicine and Telecare* 2016; 22: 472-477.
111. Minen MT, Torous J, Raynowska J, et al. Electronic behavioral interventions for headache: a systematic review. *The journal of headache and pain* 2016; 17: 51. 2016/05/11. DOI: 10.1186/s10194-016-0608-y.
112. Stubberud A and Linde M. Digital technology and mobile health in behavioral migraine therapy: a narrative review. *Current pain and headache reports* 2018; 22: 66.
113. Labrique AB, Vasudevan L, Kochi E, et al. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. *Global health: science and practice* 2013; 1: 160-171.
114. BinDhim NF, Hawkey A and Trevena L. A systematic review of quality assessment methods for smartphone health apps. *Telemedicine and e-Health* 2015; 21: 97-104.
115. McKay FH, Cheng C, Wright A, et al. Evaluating mobile phone applications for health behaviour change: A systematic review. *Journal of telemedicine and telecare* 2016: 1357633X16673538.
116. Grundy QH, Wang Z and Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. *American journal of preventive medicine* 2016; 51: 1051-1059.
117. Eysenbach G and Group C-E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal of medical Internet research* 2011; 13.
118. Chatzipavlou IA, Christoforidou SA and Vlachopoulou M. A recommended guideline for the development of mHealth Apps. *mHealth* 2016; 2.
119. Larman C and Basili VR. Iterative and incremental developments. a brief history. *Computer* 2003; 36: 47-56. DOI: 10.1109/MC.2003.1204375.
120. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *International Journal of Nursing Studies* 2011; 48: 661-671. DOI: 10.1016/j.ijnurstu.2011.01.016.

121. Higgins JPT and Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.
122. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 2009; 151: 264-269.
123. Headache. The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia* 2013; 33: 629-808. 2013/06/19. DOI: 10.1177/0333102413485658.
124. Rethlefsen ML, Murad MH and Livingston EH. Engaging medical librarians to improve the quality of review articles. *Jama* 2014; 312: 999-1000. 2014/09/10. DOI: 10.1001/jama.2014.9263.
125. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in medicine* 2002; 21: 1331-1335.
126. Hermens HJ, Freriks B, Disselhorst-Klug C, et al. Development of recommendations for SEMG sensors and sensor placement procedures. *Journal of Electromyography and Kinesiology* 2000; 10: 361-374. DOI: 10.1016/S1050-6411(00)00027-4.
127. Mathiassen SE, Winkel J and Hagg GM. Normalization of surface EMG amplitude from the upper trapezius muscle in ergonomic studies - A review. *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology* 1995; 5: 197-226. 1995/12/01.
128. Burden A. How should we normalize electromyograms obtained from healthy participants? What we have learned from over 25 years of research. *Journal of Electromyography and Kinesiology* 2010; 20: 1023-1035. DOI: 10.1016/j.jelekin.2010.07.004.
129. Knutson LM, Soderberg GL, Ballantyne BT, et al. A study of various normalization procedures for within day electromyographic data. *Journal of Electromyography and Kinesiology* 1994; 4: 47-59.
130. Chowdhury RH, Reaz MB, Ali MA, et al. Surface electromyography signal processing and classification techniques. *Sensors (Basel, Switzerland)* 2013; 13: 12431-12466. 2013/09/21. DOI: 10.3390/s130912431.
131. Headache Classification Subcommittee of The International Headache Society. The International Classification of Headache Disorders 3rd edition. *Cephalalgia* 2018; 38: 1-211.
132. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)* 1986; 1: 307-310. 1986/02/08. DOI: 10.1016/S0140-6736(86)90837-8.
133. Shrout PE and Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 1979; 86: 420.
134. Watson PF and Petrie A. Method agreement analysis: A review of correct methodology. *Theriogenology* 2010; 73: 1167-1179. DOI: 10.1016/j.theriogenology.2010.01.003.
135. Barnhart HX, Haber MJ and Lin LI. An overview on assessing agreement with continuous measurements. *Journal of biopharmaceutical statistics* 2007; 17: 529-569. 2007/07/07. DOI: 10.1080/10543400701376480.
136. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment* 1994; 6: 284. DOI: 10.1037/1040-3590.6.4.284.

137. Labbé EL and Williamson DA. Treatment of childhood migraine using autogenic feedback training. *Journal of consulting and clinical psychology 52(6)*(1984).
138. Labbé EE. Treatment of childhood migraine with autogenic training and skin temperature biofeedback: a component analysis. *Headache 35(1)*(1995).
139. Scharff L, Marcus DA and Masek BJ. A controlled study of minimal-contact thermal biofeedback treatment in children with migraine. *Journal of pediatric psychology 27(2)*(2002).
140. Fentress DW, Masek BJ, Mehegan JE, et al. Biofeedback and relaxation-response training in the treatment of pediatric migraine. *Developmental medicine and child neurology* 1986; 28: 139-146. 1986/04/01.
141. Scharff L, Marcus DA and Masek BJ. A controlled study of minimal-contact thermal biofeedback treatment in children with migraine. *Journal of pediatric psychology* 2002; 27: 109-119. Clinical Trial; Comparative Study; Randomized Controlled Trial; Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, P.H.S.
142. Steiner TJ, Paemeleire K, Jensen R, et al. European principles of management of common headache disorders in primary care. *The journal of headache and pain* 2007; 8 Suppl 1: S3-47. 2007/12/06. DOI: 10.1007/s10194-007-0366-y.
143. Lewis DW, Winner P and Wasiewski W. The placebo responder rate in children and adolescents. *Headache* 2005; 45: 232-239. 2005/04/20. DOI: 10.1111/j.1526-4610.2005.05050.x.
144. Diener H, Schorn C, Bingel U, et al. The importance of placebo in headache research. *Cephalalgia* 2008; 28: 1003-1011.
145. Evers S, Marziniak M, Frese A, et al. Placebo efficacy in childhood and adolescence migraine: an analysis of double-blind and placebo-controlled studies. *Cephalalgia* 2009; 29: 436-444. 2009/01/28. DOI: 10.1111/j.1468-2982.2008.01752.x.
146. Diener H, Dodick D, Aurora S, et al. OnabotulinumtoxinA for treatment of chronic migraine: results from the double-blind, randomized, placebo-controlled phase of the PREEMPT 2 trial. *Cephalalgia* 2010; 30: 804-814.
147. Mullinix JM, Norton BJ, Hack S, et al. Skin temperature biofeedback and migraine. *Headache* 1978; 17: 242-244. 1978/01/01. DOI: 10.1111/j.1526-4610.1978.hed1706242.x.
148. Largen JW, Mathew RJ, Dobbins K, et al. Specific and non-specific effects of skin temperature control in migraine management. *Headache* 1981; 21: 36-44. 1981/03/01. DOI: 10.1111/j.1526-4610.1981.hed2102036.x.
149. Kewman D and Roberts AH. Skin temperature biofeedback and migraine headaches. *Biofeedback and Self-Regulation* 1980; 5: 327-345.
150. Schwartz MS and Andrasik F. Headache. In: Schwartz MS and Andrasik F (eds) *Biofeedback A Practicioners Guide*. New York, NY: Guilford Press, 2003.
151. Schwartz NM and Schwartz MS. Definitions of Biofeedback and Applied Psychophysiology. In: Schwartz MS and Andrasik F (eds) *Biofeedback*. New York, NY: The Guilford Press, 2003.
152. Krogsboll LT, Hrobjartsson A and Gotzsche PC. Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no treatment, placebo and active intervention. *BMC Med Res Methodol* 2009; 9: 1. 2009/01/07. DOI: 10.1186/1471-2288-9-1.

153. Sarafino EP and Goehring P. Age comparisons in acquiring biofeedback control and success in reducing headache pain. *Annals of Behavioral Medicine* 2000; 22: 10-16.
154. Stubberud A, Varkey E, McCrory DC, et al. Biofeedback as prophylaxis for pediatric migraine: a meta-analysis. *Pediatrics* 2016; 138 2016/07/28. DOI: 10.1542/peds.2016-0675.
155. Rausa M, Palomba D, Cevoli S, et al. Biofeedback in the prophylactic treatment of medication overuse headache: a pilot randomized controlled trial. *The journal of headache and pain* 2016; 17: 87.
156. Rains JC. Change mechanisms in EMG biofeedback training: cognitive changes underlying improvements in tension headache. *Headache: The Journal of Head and Face Pain* 2008; 48: 735-736.
157. Zapata BC, Fernández-Alemán JL, Idri A, et al. Empirical studies on usability of mHealth apps: a systematic literature review. *Journal of Medical Systems* 2015; 39: 1. journal article. DOI: 10.1007/s10916-014-0182-2.
158. Schmidt B. Proof of principle studies. *Epilepsy research* 2006; 68: 48-52.
159. Huguet A, McGrath PJ, Wheaton M, et al. Testing the feasibility and psychometric properties of a mobile diary (myWHI) in adolescents and young adults with headaches. *JMIR Mhealth Uhealth* 2015; 3: e39. 2015/05/10. DOI: 10.2196/mhealth.3879.
160. Penzien DB, Rains JC and Andrasik F. Behavioral management of recurrent headache: three decades of experience and empiricism. *Applied psychophysiology and biofeedback* 2002; 27: 163-181.
161. Ramsey RR, Ryan JL, Hershey AD, et al. Treatment adherence in patients with headache: a systematic review. *Headache: The Journal of Head and Face Pain* 2014; 54: 795-816.
162. Gewirtz A and Minen M. Adherence to behavioral therapy for migraine: knowledge to date, mechanisms for assessing adherence, and methods for improving adherence. *Current pain and headache reports* 2019; 23: 3.
163. Minen MT, Adhikari S, Seng EK, et al. Smartphone-based migraine behavioral therapy: a single-arm study with assessment of mental health predictors. *NPJ digital medicine* 2019; 2: 1-9.
164. Kroon Van Diest AM and Powers SW. Cognitive Behavioral Therapy for Pediatric Headache and Migraine: Why to Prescribe and What New Research Is Critical for Advancing Integrated Biobehavioral Care. *Headache* 2019; 59: 289-297. 2018/11/18. DOI: 10.1111/head.13438.
165. Pramana G, Parmanto B, Lomas J, et al. Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. *JMIR serious games* 2018; 6: e9.
166. Kroon Van Diest AM, Ramsey R, Aylward B, et al. Adherence to Biobehavioral Recommendations in Pediatric Migraine as Measured by Electronic Monitoring: The Adherence in Migraine (AIM) Study. *Headache* 2016; 56: 1137-1146. 2016/05/12. DOI: 10.1111/head.12836.
167. Park J-W, Chu MK, Kim J-M, et al. Analysis of trigger factors in episodic migraineurs using a smartphone headache diary applications. *PloS one* 2016; 11: e0149577.
168. Pagán J, Zapater M and Ayala JL. Power transmission and workload balancing policies in eHealth mobile cloud computing scenarios. *Future Generation Computer Systems* 2018; 78: 587-601.

169. Turner DP, Lebowitz AD, Chtay I, et al. Forecasting Migraine Attacks and the Utility of Identifying Triggers. *Curr Pain Headache Rep* 2018; 22: 62. 2018/07/18. DOI: 10.1007/s11916-018-0715-3.
170. Age limits and adolescents. *Paediatrics & Child Health* 2003; 8: 577.
171. Furnham A. Response bias, social desirability and dissimulation. *Personality and individual differences* 1986; 7: 385-400.
172. Knowles ES and Nathan KT. Acquiescent responding in self-reports: Cognitive style or social concern? *Journal of Research in Personality* 1997; 31: 293-301.
173. Penzien DB, Andrasik F, Freidenberg BM, et al. Guidelines for trials of behavioral treatments for recurrent headache, first edition: American Headache Society Behavioral Clinical Trials Workgroup. *Headache* 2005; 45 Suppl 2: S110-132. 2005/06/01. DOI: 10.1111/j.1526-4610.2005.4502004.x.
174. Fleiss JL. *Design and Analysis of Clinical Experiments*. Wiley, 2011.
175. Walter S, Eliasziw M and Donner A. Sample size and optimal designs for reliability studies. *Statistics in medicine* 1998; 17: 101-110.
176. Johanson GA and Brooks GP. Initial scale development: sample size for pilot studies. *Educational and Psychological Measurement* 2010; 70: 394-400.
177. Schmettow M. Sample size in usability studies. *Communications of the ACM* 2012; 55: 64-70.
178. Nielsen J and Landauer TK. A mathematical model of the finding of usability problems. In: *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* 1993, pp.206-213. ACM.
179. Hwang W and Salvendy G. Number of people required for usability evaluation: the 10±2 rule. *Communications of the ACM* 2010; 53: 130-133.
180. Borsci S, Londei A and Federici S. The Bootstrap Discovery Behaviour (BDB): a new outlook on usability evaluation. *Cognitive processing* 2011; 12: 23-31.

8 Appendices

8.1 Appendix 1: Sample size calculation for validation study

Based on a model for sample size determination in reliability studies by Bonett,[125] we calculated the following sample size:

$$n = \frac{8z_{\alpha/2}^2\{(1-\tilde{\rho})^2(1+(k-1)\tilde{\rho})^2\}}{\{k(k-1)w^2\}} + 1$$

Assuming good agreement ($\tilde{\rho}$ = 0.8) between stationary and wireless equipment, a sample of n = 13 and two fixed observers (k) are sufficient to achieve a 95% confidence interval with width w = 0.4 (with z-value corresponding to a significance level at α = 0.05). This ensures a lower confidence limit that indicates reliable agreement[136]. The model also suggests adding $5\tilde{\rho}$ samples for increased accuracy, resulting in a total of 18 participants. Thus, we set out to recruit 20 healthy volunteers in order to account for potential dropouts.

8.2 Appendix 2: Photos of sensors in use and app screenshots

Photo 1: The muscle tension sensor in use. Placed over the upper trapezius muscle fibers midway between the acromion and spinous process of C7 and attached to the skin by a sticky electrode with conductive gel.

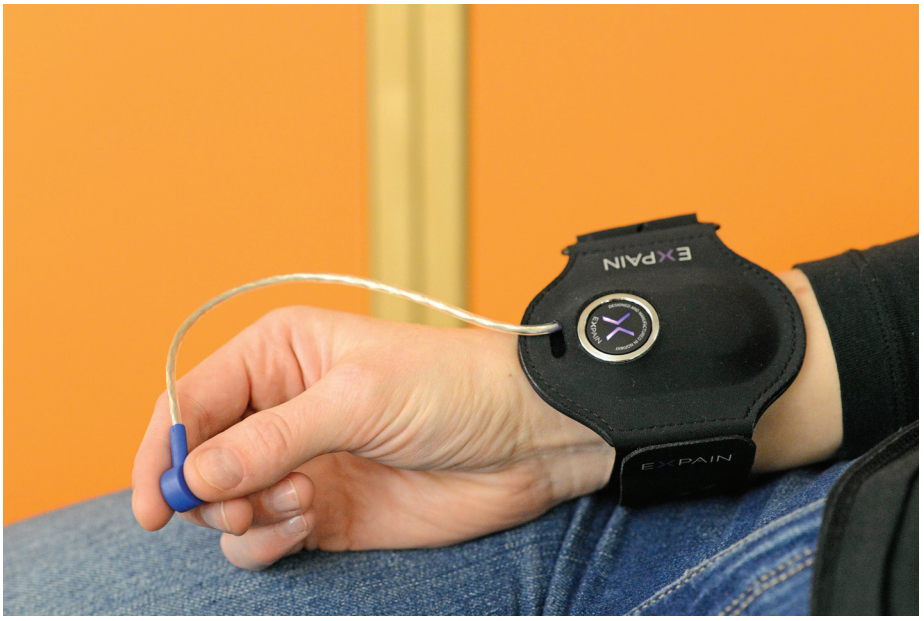


Photo 2: The temperature sensor in use. The thermistor is held between the thumb and index finger of the right hand. The thermistor is soldered onto a NeckSensor™, to use this for signal processing and Bluetooth® transmission.

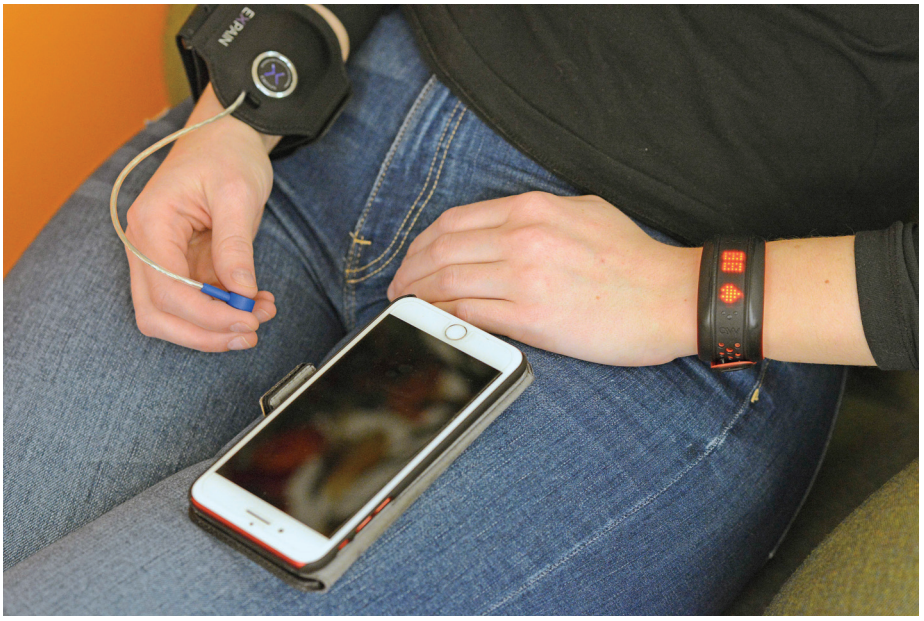


Photo 3: The temperature sensor and the heart rate wristband in use. The thermistor is held between the thumb and index finger of the right hand. The heart rate wristband sensor is worn as a watch around the left wrist and measures heart rate on the dorsal aspect of the forearm.

Photo 4. Mi-Insight app screenshots.

8.3 Appendix 3: Usability evaluation questionnaire

Usability questionnaire biofeedback app

Circle the alternative that best answers the question.

Background

Age: _____ years

Gender: *Male* *Female*

ID-code in the app: _____

General

How many hours do you use your smartphone daily?: _____

How much have you used apps on your smartphone?

1. Very little **2**. Little **3**. Some **4.** Much **5**. Very much

How much have you used wearable devices to measure functions in you body (for example pulse or blood sugar?

Very little **2**. Little **3**. Some **4.** Much **5**. Very much

How many hours do you work out during a week?: _____

How many hours did you sleep last night?: _____

How was the quality of your sleep last night?

1. Very poor **2**. Poor **3**. Normal **4.** Good **5**. Very good

How sleepy do you feel today?

1. Very sleepy **2**. Sleepy **3**. Neutral **4.** Not sleepy **5**. Not sleepy at all

How tired do you feel today?

1. Very tired **2**. Tired **3**. Neutral **4.** Not tired **5**. Not tired at all

How tense do you feel today?

1. Very tense **2**. Tense **3**. Neutral **4.** Not tense **5**. Not tense at all

Engaging

I enjoyed using the app together with the sensors.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

The contents of the app were interesting.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

I would like to use the app again.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

I would recommend the app to others.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

Functionality

It was easy setting up and connecting the sensors.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

The sensors were easy to use.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

The app was easy to use.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

The app responded quickly.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

Navigating between screens and functions was easy.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

The app was unneccesarily complicated.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

Design

I liked the looks of the app.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

Information

The app had enough information to be easy to use.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

I felt I could achieve the goals presented in the app (for example se changes in muscle tension and temperature, and using the headache diary).

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

Biofeedback

The instructions i received during relaxation were useful.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

I felt that changes in feedback reflected my muscle tension.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

The duration of the session was to long.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

I liked that temperature and muscle tension were combined as one feedback.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

I liked how the feedback was presented on the screen.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

How would you like the feedback to be presented on the screen?

_____
_____
_____
_____

Safety

I experienced discomfort using the app.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

I experienced discomfort using the sensors.

1. Completely disagree **2.** Disagree **3.** Unsure **4.** Agree **5.** Completely agree

If you experienced discomfort please describe:

General feedback

What two things about the app did you enjoy the most?

What two things about the app did you not like?

Describe with your own words your feelings when using the app:

How can the app be improved to be more user friendly and fun to use?

Do you think this app could be used for other purposes?

Described with your own words what you want an app for migraine/headache should include:

Paper I

Biofeedback as Prophylaxis for Pediatric Migraine: A Meta-Analysis

Anker Stubberud[1], Stud.Med, Emma Varkey[2], PhD, Douglas C. McCrory[3], MD, PhD, Sindre Andre Pedersen[4], PhD and Mattias Linde[1,5], MD, PhD

[1] Department of Neuroscience, NTNU Norwegian University of Science and Technology, Trondheim, Norway
[2] Department of Health and Rehabilitation/Physiotherapy, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[3] Duke Evidence Synthesis Group, Duke Clinical Research Institute; Department of Medicine, Duke University School of Medicine; and Center for Health Services Research in Primary Care, Durham Veterans Affairs Medical Center, Durham, North Carolina
[4] Section for Medicine, NTNU University Library, NTNU Norwegian University of Science and Technology, Trondheim, Norway
[5] Norwegian Advisory Unit on Headaches, St. Olavs Hospital, Trondheim, Norway

Published in Pediatrics 2016; 138(2), p.e20160675.
DOI: https://doi.org/10.1542/peds.2016-0675

Corresponding Author: Anker Stubberud, Department of Neuroscience, Norwegian University of Science and Technology, Norway. Telephone number: 004745229174 Email: ankers@stud.ntnu.no

Short title: Biofeedback as Prophylaxis for Pediatric Migraine

Funding source: Funding for this project was provided by the Medical Student Research Programme at the Norwegian University for Science and Technology, NTNU (project number: 70367015).

Financial disclosure: The authors have no financial relationships relevant to this article to disclose.

Conflict of interest: The authors declare that there are no conflicts of interest.

Abbreviations: IHS = International Headache Society. MD = mean difference. CI = confidence interval. OR = odds ratio. NNTB = number needed to treat to benefit.

Contributors statement

Anker Stubberud co-ordinated the study, screened search results, retrieved papers, screened retrieved papers against eligibility criteria, appraised quality of papers, extracted data from papers, managed data and entered it into RevMan, analyzed and interpreted data, wrote the review, made critical revision of the manuscript for important intellectual content, and approved the final manuscript as submitted.

Emma Varkey screened search results, screened retrieved papers against eligibility criteria, appraised quality of papers, extracted data from papers, provided a clinical perspective, made critical revision of the manuscript for important intellectual content, and approved the final manuscript as submitted.

Sindre Andre Pedersen undertook the systematic literature search, screened search results, made critical revision of the manuscript for important intellectual content, and approved the final manuscript as submitted.

Douglas C. McCrory analyzed and interpreted data, made critical revision of the manuscript for important intellectual content, and approved the final manuscript as submitted.

Mattias Linde conceptualized and designed the study, assisted in co-ordination, assisted in appraising quality of papers, made critical revision of the manuscript for important intellectual content, and approved the final manuscript as submitted.

Abstract

Context: Migraine is a common problem in children and adolescents, but few satisfactory prophylactic treatments exist.

Objective: We aimed to investigate the pooled evidence for the effectiveness of using biofeedback to reduce migraine in children and adolescents.

Data sources: A systematic database search was conducted across the databases MEDLINE, EMBASE, CENTRAL, CINAHL, and PsychINFO.

Study selection: Prospective randomized controlled trials of biofeedback for migraine among children and adolescents were located in the search.

Data extraction: Reduction of mean attack frequency and a series of secondary outcomes, including adverse events, were extracted. Risk of bias was also assessed.

Results: Forest plots were created using a fixed-effects model, and mean difference (MD) was reported. Five studies with a total of 137 participants met the inclusion criteria. Biofeedback reduced migraine frequency (MD: -1.97; 95% CI -2.72 to -1.21; $p < 0.00001$), attack duration (MD: -3.94; 95% CI -5.57 to -2.31; $p < 0.00001$), and headache intensity (MD: -1.77; 95% CI -2.42 to -1.11; $p < 0.00001$), compared to a waiting-list control. Biofeedback did not demonstrate an adjuvant effect when combined with other behavioral treatment; neither did it have significant advantages over active treatment. Only 40% of bias judgments were deemed as ‘low’ risk.

Limitations: Methodological issues hampered the meta-analyses. Only a few studies were possible to include, and they suffered from incomplete reporting of data and risk of bias.

Conclusion: Biofeedback seems to be an effective intervention for pediatric migraine, but in light of the limitations, further investigation is needed to increase our confidence in the estimate.

Introduction

Migraine represents a serious problem among children and adolescents. A review of 64 studies estimated the one-year prevalence of childhood migraine to be 9.1%[1]. Still, this probably is an underestimation, due to the common practices of using restrictive screening questions and neglecting probable migraine. A recent study reported a 36% one-year prevalence of all migraine among adolescents[2]. For patients, this means troublesome symptoms and often considerable degrees of disability with time lost from school, friends and other activities[2,3]. From a societal perspective, migraine leads to substantial indirect costs from lost productivity and direct costs for health care[4].

Despite the high prevalence and morbidity, relatively few prophylactic drugs have been proven effective among children and adolescents, and they are all associated with a risk of adverse effects[5]. Non-pharmacologic treatment (e.g., biofeedback), is therefore an attractive alternative. In biofeedback, patients learn to voluntarily modify their bodily reactions through feedback-mediated awareness of physiological parameters[6]. Biofeedback reduces cortical excitability, and affects resonance and oscillations of essential feedback loops in the central nervous system[7,8]. The most frequently used modalities in biofeedback treatment are peripheral skin temperature, blood-volume-pulse, and electromyography.

Many systematic reviews have reported a favorable effect of behavioral treatments for pain conditions[9-14], but they vary greatly in how they have applied meta-analytic methodology. Unfortunately, most of these studies[9,10,12-14] have merged different types of psychological treatment and pain conditions, including tension-type headache and migraine. This does not allow us to say if biofeedback is effective as a migraine

prophylactic. Only Nestoriuc et al.[11] have considered migraine separate from other headache disorders, and biofeedback separate from other psychological treatment. However, their study was restricted to adults.

To fill in this gap of knowledge, we present here the results of a systematic review with a meta-analysis of the effect of biofeedback treatment in pediatric migraine. The objectives were: to assess the efficacy of biofeedback on primarily attack frequency in children and adolescents with migraine; assess the efficacy on secondary endpoints (e.g., attack duration, headache intensity, quality of life, disability, and acute medication use); investigate any potential adverse events associated with the treatment; and conduct a risk of bias assessment of the included studies.

Methods

Criteria for considering studies for this review

Types of studies

Included studies were required to be prospective randomized controlled trials investigating biofeedback as a prophylactic treatment for episodic migraine in children or adolescents. Studies were only included if they were randomized or pseudo-randomized. Due to the low number of studies expected to meet these criteria, no lower limit for number of participants was set.

Types of participants

Participants were children and adolescents up to the age of 18 suffering from episodic migraine. We did not require the use of a specific set of diagnostic criteria (e.g. IHS Classification Committee 1988[15] or ICHD-II 2004[16]), but the diagnosis had to be based on a least some of the distinctive migraine features defined by the International

Headache Society (IHS): Unilateral location, pulsating character, moderate to severe intensity, physical aggravation, accompanying nausea or photo- and phonophobia, and aura[17].

Types of interventions

Studies were eligible if at least one arm represented biofeedback treatment. All modalities of biofeedback were included. Studies were considered eligible when some degree of behavioral treatment was delivered alongside biofeedback during the same session, or when biofeedback was the only difference between the intervention group and comparison group. Eligible comparison groups were active treatment with documented effectiveness; non-pharmacological therapies with documented effectiveness; waiting-list control; or treatment as usual.

Types of outcome measures

Migraine frequency was chosen as the primary outcome of interest[18]. Secondary outcomes pre-specified to be extracted were: Responder rate equal to or greater than 50%, headache intensity, attack duration, disability, quality of life, doses of acute medication, and adverse events[18]. We also aimed to assess effect sizes by sex in the included studies.

Search methods for identification of studies

A medical librarian performed the literature search[19]. The searched databases included MEDLINE, EMBASE, CENTRAL, CINAHL, and PsychINFO. The search was updated on November 23, 2015, and involved a combination of thesaurus and free-text terms optimized to cover randomized control trial studies where patients under the age of 18 had received biofeedback treatment as a prophylaxis for migraine (see Appendix 1 for the complete search strategy for all databases searched). The literature

lists of all reviews encountered on the subject were hand-searched in order to capture potentially relevant studies not detected in the electronic search.

Data collection and analysis

Study selection

Two authors independently screened the results from the literature search to identify eligible studies. In cases where papers could not be excluded based on information in the title and abstract, full texts were obtained and screened. The remaining studies were included in this review. Disagreements were resolved through discussion, and near-eligible studies are referenced in this review with reasons for exclusion.

Data extraction and management

Characteristics of each included study were summarized, including study design and methods; participants demographics and criteria for migraine diagnosis; characteristics of intervention arms; outcomes with method of data collection; and units of measurement. Information on the biofeedback treatment, including type of instrument, modality, setting, and circumstances was extracted. Any additional treatment to biofeedback was reviewed. Raw outcome data was extracted from the studies for meta-analysis. We primarily sought Ns, means, and standard deviations. In such cases where this could not be obtained directly from the paper, the data were calculated in-house from the information provided in the paper. Headache diary outcomes are usually reported over different time periods, and we therefore attempted to standardize the unit of time over which outcomes were measured. Outcome data was assessed at end of treatment and follow-up. End of treatment was considered as the last weeks of treatment when outcomes were assessed, or the first weeks immediately following treatment if outcome assessment was post-treatment. Follow-

up was considered to be 3-12 months after completed treatment, and in cases where more than one follow-up time point was reported, the last one was used. Two authors extracted data and reconciled their findings.

Data synthesis

We used the Cochrane Collaboration software Review Manager (RevMan 5.3) for synthesis of meta-analyses and construction of figures. Raw data from the included studies were entered into the software. In cases where the means and variances of groups were not sufficiently reported, we attempted to calculate the necessary data from the data reported, such as test statistics and error bars in graphs, whenever possible. Scales for outcome assessment were converted to be equivalent. For continuous outcomes we calculated the summary mean difference (MD) with 95% confidence intervals (CIs), using an inverse variance fixed-effects model. For dichotomous outcomes, we calculated the summary odds ratio (OR) with 95% CI with a fixed-effects model. Owing to the low number of participants in each meta-analysis, the Mantel-Haenszel method was used for calculating dichotomous outcomes. We also calculated the number needed to treat to benefit (NNTB) based on an assumed control risk, calculated from the responder rate in the control groups. Statistical heterogeneity was also calculated for each meta-analysis to evaluate the variability of intervention effects across the included studies.

Risk of bias assessment in included studies

Four categories of bias were considered: Selection bias with regard to random sequence generation and allocation concealment; detection bias with regard to blinding of outcome assessors; attrition bias, i.e. selective occurrence and biased handling of protocol deviations and losses to follow-up; and reporting bias determined

by differences between pre-specified measures and reported outcomes. Other potential biases (e.g., biased study design or claim of fraud) were to be reported if encountered. Performance bias was not assessed due to the difficulty of blinding participants and personnel when delivering biofeedback treatment. Each bias was graded as being of 'low', 'high', or 'unclear' risk. The latter was chosen when the information in the paper was insufficient to determine the risk. Two authors performed the assessment independently, and discrepancies were thereafter resolved by discussion and referral with a third author.

Results

Results of search

Study selection

Figure 1 presents a flow diagram of the process for study selection. The electronic search yielded 908 records. After removing duplicates, 639 records remained, and 581 of these were excluded through screening of titles and abstracts. The full-text files of the 58 remaining records were then retrieved and read. Eleven of these studies, and a single study identified through the hand-search[20], i.e., a total of 12 studies, qualified for description in the review. Five of these[21-25] met all the eligibility criteria and are included in data synthesis. The remaining seven studies[20,26-31] are listed with their reason for exclusion in Table 1. Characteristics of included studies in summary are found in table 2. Detailed information may be accessed through appendix 2-6.

Risk of bias

Of the 30 risk of bias items scored for the five studies, 12 (40%) were 'low', 15 (50%) were 'unclear', and three (10%) were 'high'. The three bias items scored 'high' were

limited to two studies[23,24]. Figure 2 gives an overview of the risk of bias assessment. One[24] of the five included studies described an adequate random sequence generation earning a low risk of bias, whereas the other four[21-23,25] lacked description and were assigned unclear risk of bias. For allocation concealment, none of the studies provided sufficient information to ascertain the true risk of bias, and subsequently all were assigned an unclear risk of bias. For the blinding of outcome assessment, Scharff et al.[24] was judged to suffer from a high risk of detection bias because all evaluation, treatment, and follow-up sessions were conducted by a single investigator. The four[21-23,25] remaining studies were assigned an unclear risk of bias status due to insufficient information. Only two of the included studies reported when there were significant differences between completers and non-completers[22,24]. Fentress et al.[25] evaluated 35 patients to obtain a final sample of 18 participants. These 18 were also analyzed, thus giving the study a low risk of bias. Labbé 1984[21] reported dropouts only at follow-up, a time point not included in our analyses, thus giving the study an unclear risk of bias. Labbé 1995[22] recruited 46 participants, but only 30 completed the study. The study reported no significant differences between completers and dropouts, but no information is given on how the dropouts were treated in the analyses, resulting in an unclear risk of bias for the study. In the Sartory et al. study[23], sixteen children could not be contacted at follow-up. Only children with complete data sets are included in the table of means that was used for the meta-analyses, resulting in our analyses being conducted with a substantial departure of participants from the intervention to which they were assigned at randomization. This qualifies for a high risk of bias status. Scharff et al.[24] reported two dropouts after randomization, but before initiation of treatment. No significant differences were found between dropouts and participants with regards to age, psychological measures, or headache characteristics, thus giving

the study a low risk of bias. Four of five[21-23,25] studies reported results of all preplanned outcomes and were assigned a low risk of bias for selective reporting. Scharff et al.[24] was the only study to not report data fully, and was therefore classified as high risk of bias for selective reporting. The study also did not report data sufficient for assessment of depression and anxiety outcomes at post-treatment. No other bias was encountered in the studies.

Data analysis

Four of the five included studies reported outcomes over a one-week time period[21-23,25]. Data from the final study[24] was converted to fit this. Ordinal scales used for outcome assessment were converted to be equal. One study[24] did not report means and measures of spread as numbers. These data were therefore derived by hand from error bars in the graphs. Two studies[21,22] did not report measures of spread, only F-statistics for the ANOVA analyses. To estimate the standard deviation, we calculated the between-group variance of the groups and phases included in the ANOVA analyses, and thereby estimated a within-group variance. One study[25] used non-parametric methods in their analyses. Consequently, no continuous outcomes from this study could be used in the meta-analyses. No investigations of differences in treatment efficacy between girls and boys could be done because none of the included studies reported outcomes by sex.

Results of analyses

In cases where only one study could be entered into a comparison, we chose to present a forest plot for our primary outcome measurement for ease of interpretation.

Biofeedback versus waiting-list control

Four studies, with a total of 84 participants, qualified for comparisons of biofeedback versus waiting-list control[21,22,24,25]. In all four studies, hand-warming biofeedback, with an additional behavioral therapy delivered during the same sessions (appendix 2-6), was compared to a waiting-list control.

Data from three trials[21,22,24] (72 participants) showed that biofeedback significantly ($z=5.10$; $p<0.00001$) reduced the frequency of migraine attacks at the end of treatment compared to waiting-list control (Figure 3). The mean difference between interventions was -1.97 [95% CI (-2.72, -1.21)] attacks per week. Only one study[22] compared biofeedback and waiting-list control at post-treatment follow-up. The study reported significant differences for headache frequency and duration across time for all subjects at 6-month follow-up.

Data from four studies[21,22,24,25] (84 participants) of biofeedback versus waiting-list control were included in an analysis to enumerate the responder rate. The definition of responder rate varied between all of these studies (appendix 2-4, 6). Participants treated with biofeedback showed a significantly higher ($z=4.57$; $p<0.00001$) proportion of responders to treatment at the end of treatment compared with waiting-list control (OR=27.71; 95% CI 6.66 to 115.35) (Figure 4). The number needed to treat to benefit was 2.

Two studies[21,22] (48 participants) were meta-analyzed to assess whether biofeedback reduced the duration of migraine attacks compared to waiting-list control at end of treatment (appendix 3, 4). The analysis showed a mean difference in pain intensity after biofeedback versus waiting-list control of -3.94 [95% CI (-5.57, 2.31)], which was significant ($z=4.75$; $p<0.00001$) (Figure 5). The one study assessing the outcome

at post-treatment follow-up reported maintained improvement for the biofeedback group[22].

Data from two studies[22,24] (52 participants,) were included in a meta-analysis to investigate if biofeedback improved headache intensity compared with waiting-list control (appendix 4, 5). The analysis showed a mean difference in headache duration after biofeedback versus waiting-list control of -1.77 [95% CI (-2.42, -1.11)], which was significant (z=5.30; p<0.00001) (Figure 6). None of the included studies assessed headache intensity at post-treatment follow-up for this comparison.

The secondary outcomes of interest—disability, quality of life, and adverse events—were not assessed by any of the studies comparing biofeedback with a waiting-list control (appendix 2-4, 6). Only one study comparing biofeedback with a waiting-list control assessed the outcome doses of acute medication, and reported a significant reduction over time for medication consumption in both the biofeedback and waiting-list control group. However, no significant difference between the groups at end of treatment and follow-up was reported[21] (appendix 3).

Adjuvant effect of biofeedback

Two of the eligible studies[22,25] had biofeedback as the only difference between two treatment arms, allowing for a meta-analysis of its adjuvant effect. Only one of these[22] (20 participants) reported sufficient data to analyze continuous outcomes. This trial displayed no significant effects, either for migraine frequency (MD=-0.40; 95% CI (-1.64, 0.84); z=0.63; p=0.63; Figure 7) or attack duration (MD=-0.36; 95% CI (-2.80, 2.08); z=0.29; p=0.77; Figure 7), when comparing biofeedback plus autogenic training with autogenic training only. Both studies[22,25] (32 participants) reported the proportion of responders to treatment, and a meta-analysis showed no significant

effect (OR=1.79; 95% CI (0.21, 15.55); z=0.53; p=0.60; Figure 8) for biofeedback as adjuvant treatment in this regard.

Biofeedback versus active treatment

One study[23] compared biofeedback with active control groups. Data were reported for 27 of the original 43 included participants. No significant differences were found in migraine frequency when comparing biofeedback to progressive relaxation, nor when comparing biofeedback to propranolol at the end of treatment or at follow-up (Figure 9). Moreover, the study reported no significant group differences for the outcomes headache intensity, attack duration, and analgesic intake. On the other hand, non-parametric, pre-post within-group analyses demonstrated significant improvement in migraine frequency and intensity for the relaxation group, and significant improvement with regard to migraine frequency, duration, and mood for the biofeedback group. Neither the relaxation group nor the metoprolol group differed significantly from the biofeedback group with regards to responder rate at post-treatment. The study did not assess the outcomes of disability, quality of life, or adverse events.

Biofeedback versus "sham-biofeedback"

One study[24] (23 participants) compared hand-warming biofeedback to hand-cooling biofeedback. No significant between-group benefit was found for migraine frequency at end of treatment or follow-up (Figure 10). However, the proportion of responders to treatment was significantly higher in the hand-warming group (7/13 vs. 1/10; OR 10.50; 95% CI (1.02, 108.58); z=1.97; p=0.049).

Discussion

The present systematic review is the first to attempt to estimate the pooled intervention effect for biofeedback treatment among children and adolescents with migraine. We primarily set out to assess its impact on headache frequency, but also several secondary outcomes defined by IHS[32]. The most robust finding of the review is that biofeedback can reduce the frequency of migraine when compared to a waiting-list control (Figure 3). Biofeedback also seems to reduce attack duration and headache intensity compared to waiting-list controls. However, some pre-specified outcomes were not possible to meta-analyze due to the low number of studies reporting these data.

An adverse event is an outcome that is often neglected; through this review, we had hoped to learn some of its association to biofeedback. The lack of attention to the adverse events outcome became even more apparent upon learning that none of the included studies addressed this outcome.

A ‘low’ risk of bias was found in just 40% of the scores, the remaining being deemed ‘unclear’ or ‘high’. This decreases the confidence in our estimates. There was a substantial lack of description of the randomization process, where four out of five random sequence generation judgments, and all judgments for allocation concealment, were scored ‘unclear’. Considering that blinding is not possible when delivering biofeedback, this risk of bias has not been assessed. Consequently, there is the possibility of a contribution by a placebo effect in the intervention group.

Three of the studies[21,22,24] (appendix 3, 4 and 6) used peripheral skin temperature, one used electromyography[25] (appendix 2), and one[23] (appendix 5) used vasomotor tone

for biofeedback. The two former techniques are based on the fact that increased peripheral skin temperature and decreased muscle tension are associated with a higher parasympathetic tone and a higher degree of relaxation, which in turn is assumed to lead to less migraine. The vasomotor feedback is suggested to have associations with changes in intracranial blood flow similar to those occurring in electromyography or peripheral skin temperature feedback[33], although its physiological basis is not fully understood.

A major limitation of this study is the heterogeneity of the interventions. This raises question of what part of the observed package effect may be attributed to the biofeedback. In the comparison of biofeedback with waiting-list control, we lumped together the somewhat heterogeneous intervention packages (appendix 2-6), assuming that the analyses might give information on the intervention effect of biofeedback among children with migraine. This assumption was further investigated in the analyses of the adjuvant effect of biofeedback.

Biofeedback is regarded as a complete treatment package, not just feedback from a computer[6,34]. Indeed, the characteristics of included studies showed a broad composition of treatment packages (appendix 2-6). Biofeedback as an adjuvant does not seem to increase the effect of other behavioral treatment. Some might use this finding to conclude that biofeedback per se produces no effect, but instead the effect may be attributed to other components of the treatment packages. However, considering the small sample size, the adjuvant analysis is likely to lack sufficient statistical power to exclude the possibility that some differences may exist. The small number of participants eligible to be included warrants further research. In addition, it is possible that biofeedback as a supplement to relaxation therapies would give no

additional effect because the patient has received the maximum effect from the other relaxation strategies.

According to the publications we found, biofeedback has a greater responder rate compared to waiting-list controls, with an NNTB of 2 (Figure 4). However, this information should be treated with caution, given that only one study[25] used the responder rate as defined by IHS[18,32]. Three studies[21,22,24] defined responder rate as a 50% reduction in the average headache intensity, while the final study[23] used a 50% reduction in an index derived by multiplying headache frequency by intensity. Despite these differences, we chose to meta-analyze these outcomes.

Another limitation of this review is the fact that children and adolescents were regarded as one group. Biological and psychological differences between these age groups could hamper interpretation of the results. The included studies only provided age means, and never medians, making it impossible to perform separate subgroup analyses of young children and adolescents as defined by (for example) the Adolescent Health Committee[35]. We may therefore only be certain that the intervention effect is of value for patients under the age of 18.

Our findings are in accordance with the well-established use of behavioral treatment as migraine prophylaxis[36], and with recommendations of biofeedback treatment for migraine in guidelines[37]. Another meta-analysis from 2007 that investigated biofeedback as prophylactic treatment for adults with migraine concluded with a medium effect size[11]. These results, together with our findings, show that biofeedback has a place in the treatment of migraine regardless of age group.

A major strength of the present review is the fact that it analyzed biofeedback separately from other psychological treatments, and migraine separately from other

headache diagnoses. We also present systematic descriptions of all included studies (appendix 2-6) because it serves to enlighten the diversity of treatment compositions and differences in outcome definitions. Further strengthening this review, a comprehensive literature search strategy was used in order to locate all potentially eligible studies. In addition, we were able to estimate continuous data from the sparse data reported in many of the included studies, and then to utilize this information in the analyses. These are data that are not readily available from the papers. We recommend that investigators thoroughly report the number of participants, means, and measures of spread, in order to ease interpretation and comparison, and to allow for future meta-analyses. Based on the positive effectiveness findings and seemingly high tolerability, we recommend biofeedback as prophylactic treatment for childhood migraine.

There was a wide range in the number of treatment sessions, raising questions regarding the importance of treatment dose. Another review of psychological treatment for headaches concluded that higher treatment dose results in better pain scores post-treatment[38]. The studies included in this review delivered biofeedback in a clinic, which is time consuming for the patient and hampers the widespread delivery of treatment, despite its positive results in treating headache. This has led to the emergence of less time consuming approaches, such as prudent limited office treatment (PLOT) and internet-based delivery[14,39,40]. These approaches are obviously promising, and warrant further research.

Another question is whether part of the positive effect of biofeedback treatment packages should be attributed to nonspecific effects, such as effects of attention, suggestion, and expectation. In an attempt to investigate this, one of the included

studies[24] compared hand-warming biofeedback, traditionally assumed to be effective, to 'sham biofeedback' consisting of hand-cooling biofeedback. The study was unable to demonstrate any differences between the groups at end of treatment and follow-up, supporting the idea that nonspecific effects are partially responsible. Again, one should bear in mind the fact that the small number of participants might lack the statistical power to detect a difference.

Biofeedback delivered alongside relaxation therapy or autogenic training seems to be effective in reducing migraine frequency in the pediatric population. Also, the apparent lack of adverse advents should qualify biofeedback as an attractive treatment alternative for pediatric migraine. Despite the positive findings, the number of identified studies and participants were small, and a series of methodological issues hampered proper meta-analyses. Therefore, continued research is warranted.

Acknowledgements

The authors would like to thank Øyvind Salvesen at the Unit for Applied Clinical Research, Norwegian University of Science and Technology, for valuable statistical advice.

References

1. Wober-Bingol C. Epidemiology of migraine and headache in children and adolescents. *Current pain and headache reports.* 2013;17(6):341.
2. Krogh AB, Larsson B, Linde M. Prevalence and disability of headache among Norwegian adolescents: A cross-sectional school-based study. *Cephalalgia.* 2015.
3. Kernick D, Campbell J. Measuring the impact of headache in children: a critical review of the literature. *Cephalalgia.* 2009;29(1):3-16.
4. Mazzotta G, Gallai B, Mattioni A, et al. Cost assessment of headache in childhood and adolescence: preliminary data. *J. Headache Pain.* 2005;6(4):281-283.
5. Termine C, Ozge A, Antonaci F, Natriashvili S, Guidetti V, Wober-Bingol C. Overview of diagnosis and management of paediatric headache. Part II: therapeutic management. *J. Headache Pain.* 2011;12(1):25-34.

6. Schwartz NM, Schwartz MS. Definitions of Biofeedback and Applied Psychophysiology. In: Schwartz MS, Andrasik F, eds. *Biofeedback*. New York, NY: The Guilford Press; 2003.
7. Siniatchkin M, Hierundar A, Kropp P, Kuhnert R, Gerber WD, Stephani U. Self-regulation of slow cortical potentials in children with migraine: An exploratory study. *Applied Psychophysiology Biofeedback.* 2000;25(1):13-32.
8. Lehrer P, Eddie D. Dynamic processes in regulation and some implications for biofeedback and biobehavioral interventions. *Appl. Psychophysiol. Biofeedback.* 2013;38(2):143-155.
9. Hermann C, Kim M, Blanchard EB. Behavioral and prophylactic pharmacological intervention studies of pediatric migraine: an exploratory meta-analysis. *Pain.* 1995;60(3):239-255.
10. Trautmann E, Lackschewitz H, Kroner-Herwig B. Psychological treatment of recurrent headache in children and adolescents--a meta-analysis. *Cephalalgia.* 2006;26(12):1411-1426.
11. Nestoriuc Y, Martin A. Efficacy of biofeedback for migraine: a meta-analysis. *Pain.* 2007;128(1-2):111-127.
12. Palermo TM, Eccleston C, Lewandowski AS, Williams AC, Morley S. Randomized controlled trials of psychological therapies for management of chronic pain in children and adolescents: an updated meta-analytic review. *Pain.* 2010;148(3):387-397.
13. Eccleston C, Palermo TM, Williams AC, et al. Psychological therapies for the management of chronic and recurrent pain in children and adolescents. *Cochrane Database Syst Rev.* 2014;5:CD003968.
14. Fisher E, Law E, Palermo TM, Eccleston C. Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents. *Cochrane Database Syst Rev.* 2014;2014.
15. Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Headache Classification Committee of the International Headache Society. *Cephalalgia.* 1988;8 Suppl 7:1-96.
16. The International Classification of Headache Disorders: 2nd edition. *Cephalalgia.* 2004;24 Suppl 1:9-160.
17. The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia.* 2013;33(9):629-808.
18. Tfelt-Hansen P, Pascual J, Ramadan N, et al. Guidelines for controlled trials of drugs in migraine: third edition. A guide for investigators. *Cephalalgia.* 2012;32(1):6-38.
19. Rethlefsen ML, Murad MH, Livingston EH. Engaging medical librarians to improve the quality of review articles. *JAMA.* 2014;312(10):999-1000.
20. Osterhaus SO, Lange A, Linssen WH, Passchier J. A behavioral treatment of young migrainous and nonmigrainous headache patients: prediction of treatment success. *Int. J. Behav. Med.* 1997;4(4):378-396.
21. Labbé EL, Williamson DA. Treatment of childhood migraine using autogenic feedback training. *J. Consult. Clin. Psychol.* 1984;52(6):968-976.
22. Labbé EE. Treatment of childhood migraine with autogenic training and skin temperature biofeedback: a component analysis. *Headache.* 1995;35(1):10-13.
23. Sartory G, Müller B, Metsch J, Pothmann R. A comparison of psychological and pharmacological treatment of pediatric migraine. *Behav. Res. Ther.* 1998;36(12):1155-1170.

24. Scharff L, Marcus DA, Masek BJ. A controlled study of minimal-contact thermal biofeedback treatment in children with migraine. *J. Pediatr. Psychol.* 2002;27(2):109-119.
25. Fentress DW, Masek BJ, Mehegan JE, Benson H. Biofeedback and relaxation-response training in the treatment of pediatric migraine. *Dev. Med. Child Neurol.* 1986;28(2):139-146.
26. Allen KD, Shriver MD. Role of parent-mediated pain behavior management strategies in biofeedback treatment of childhood migraines. *Behav. Ther.* 1998;29(3):477-490.
27. Gerber W-D, Petermann F, Gerber-von Muller G, et al. MIPAS-family: Development and evaluation of a behavioural medicine programme for the treatment of chronic paediatric headaches. *Verhaltenstherapie.* 2008;18(4):247-255.
28. Gerber WD, Petermann F, Gerber-von Müller G, et al. MIPAS-Family-evaluation of a new multi-modal behavioral training program for pediatric headaches: clinical effects and the impact on quality of life. *J. Headache Pain.* 2010;11(3):215-225.
29. Guarnieri P, Blanchard EB. Evaluation of home-based thermal biofeedback treatment of pediatric migraine headache. *Biofeedback and self-regulation.* 1990;15(2):179-184.
30. Osterhaus SO, Passchier J, Helm-Hylkema H, et al. Effects of behavioral psychophysiological treatment on schoolchildren with migraine in a nonclinical setting: predictors and process variables. *J. Pediatr. Psychol.* 1993;18(6):697-715.
31. Osterhaus SO, Passchier J, Van der Helm-Hylkema H, de Jong K, et al. The behavioral treatment of juvenile patients with migraine in a nonclinical setting: Effects and observations. *Gedragstherapie.* 1994;27(1):3-18.
32. Silberstein S, Tfelt-Hansen P, Dodick DW, et al. Guidelines for controlled trials of prophylactic treatment of chronic migraine in adults. *Cephalalgia.* 2008;28(5):484-495.
33. McGrady A, Wauquier A, McNeil A, Gerard G. Effect of biofeedback-assisted relaxation on migraine headache and changes in cerebral blood flow velocity in the middle cerebral artery. *Headache.* 1994;34(7):424-428.
34. Schwartz MS, Andrasik F. Headache. In: Schwartz MS, Andrasik F, eds. *Biofeedback A Practicioners Guide*. New York, NY: Guilford Press; 2003.
35. Age limits and adolescents. *Paediatr. Child Health.* 2003;8(9):577.
36. Penzien DB, Irby MB, Smitherman TA, Rains JC, Houle TT. Well-Established and Empirically Supported Behavioral Treatments for Migraine. *Current pain and headache reports.* 2015;19(7):34.
37. Steiner TJ, Paemeleire K, Jensen R, et al. European principles of management of common headache disorders in primary care. *J. Headache Pain.* 2007;8 Suppl 1:S3-47.
38. Fisher E, Heathcote L, Palermo TM, de C, Lau J, Eccleston C. Systematic review and meta-analysis of psychological therapies for children with chronic pain. *J. Pediatr. Psychol.* 2014;39(8):763-782.
39. Andrasik F. Behavioral treatment of headaches: extending the reach. *Neurol. Sci.* 2012;33 Suppl 1:S127-130.
40. Trautmann E, Kroner-Herwig B. A randomized controlled trial of Internet-based self-help training for recurrent headache in childhood and adolescence. *Behav. Res. Ther.* 2010;48(1):28-37.

41. McGrath PJ, Humphreys P, Keene D, et al. The efficacy and efficiency of a self-administered treatment for adolescent migraine. *Pain.* 1992;49(3):321-324.
42. Sargent JD, Green EE, Walters ED. Preliminary report on the use of autogenic feedback training in the treatment of migraine and tension headaches. *Psychosom. Med.* 1973;35(2):129-135.
43. Falkenstein M, Hoormann J, Weitkämper B, Zülch J. Signalverarbeitung bei fotoelektrischer Gefässplethysmographie. *Biomed. Tech. (Berl.).* 1984;29:213-214.

Figure 1: Study flow diagram.

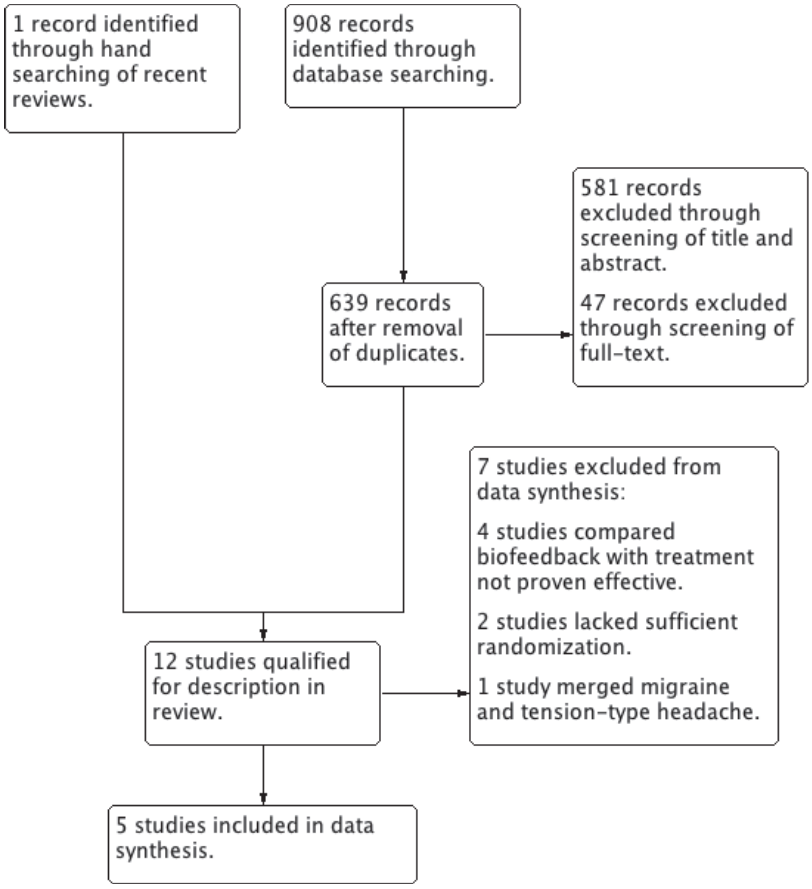


Figure 2: Risk of bias assessment table.

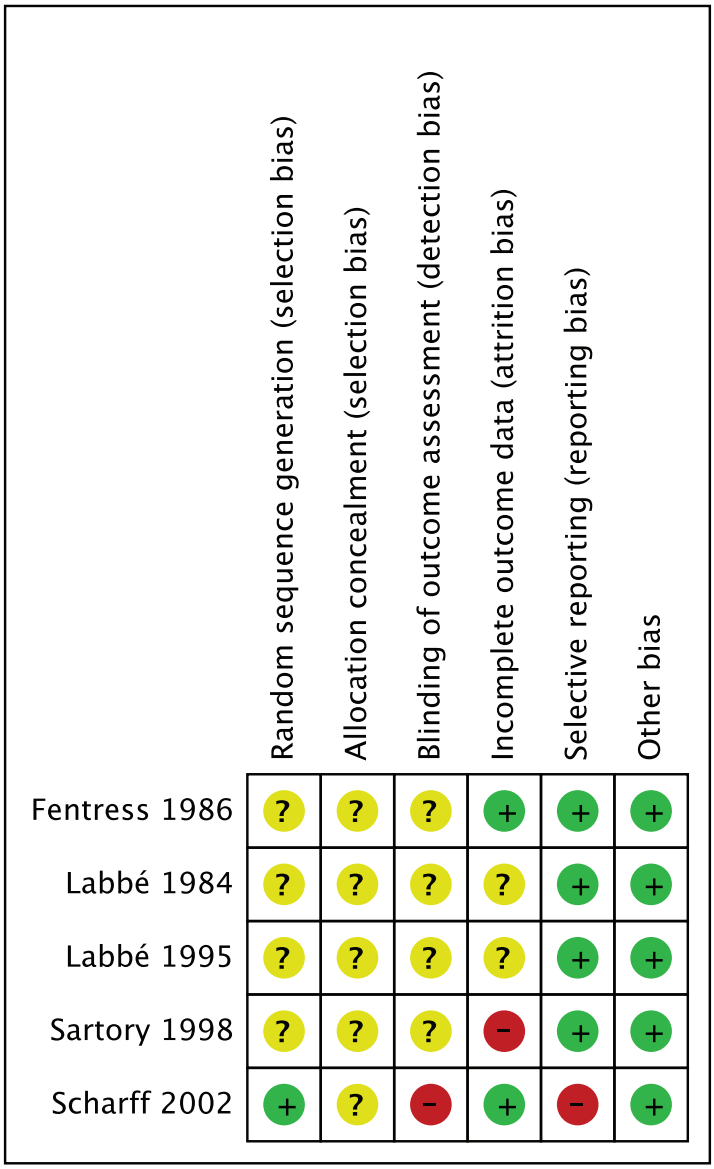
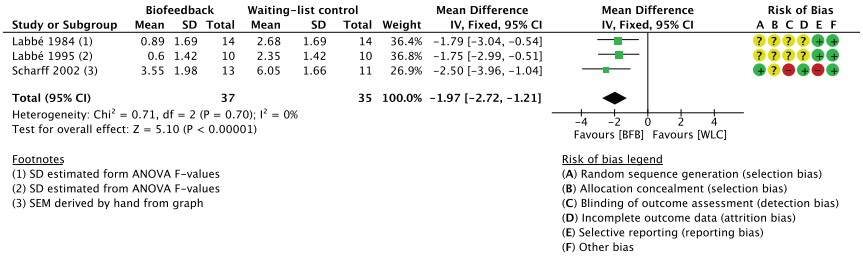


Figure 3: Comparison: Biofeedback (BFB) vs. Waiting-list control (WLC). Outcome: Migraine frequency. 1, 2 = Standard deviations estimated form ANOVA F-values; 3 = Standard error derived by hand from graph.



Footnotes
(1) SD estimated form ANOVA F-values
(2) SD estimated from ANOVA F-values
(3) SEM derived by hand from graph

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of outcome assessment (detection bias)
(**D**) Incomplete outcome data (attrition bias)
(**E**) Selective reporting (reporting bias)
(**F**) Other bias

Figure 4: Comparison: Biofeedback (BFB) vs. Waiting-list control (WLC). Outcome: Responders to treatment.

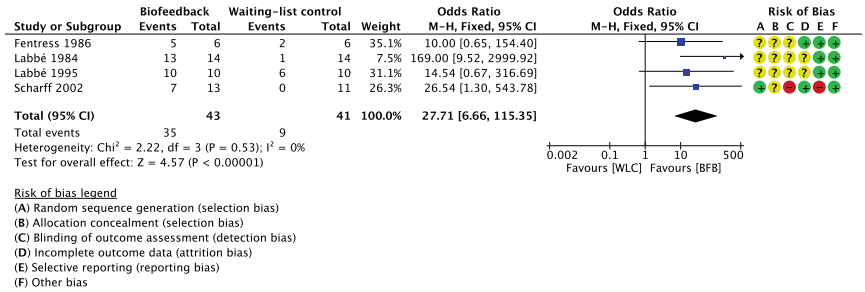
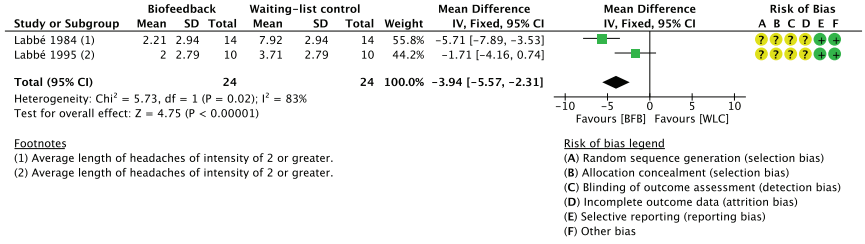


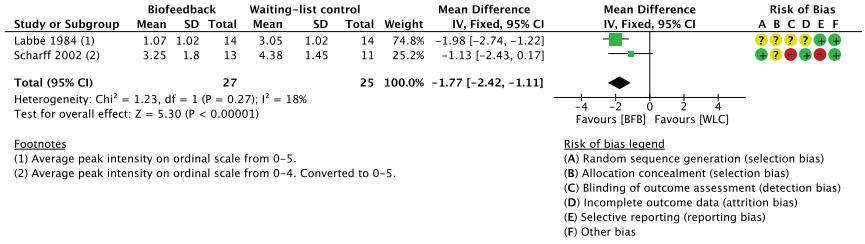
Figure 5: Comparison: Biofeedback (BFB) vs. Waiting-list control (WLC). Outcome: Attack duration. 1, 2 = SD estimated form ANOVA F-values.



Footnotes
(1) Average length of headaches of intensity of 2 or greater.
(2) Average length of headaches of intensity of 2 or greater.

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of outcome assessment (detection bias)
(**D**) Incomplete outcome data (attrition bias)
(**E**) Selective reporting (reporting bias)
(**F**) Other bias

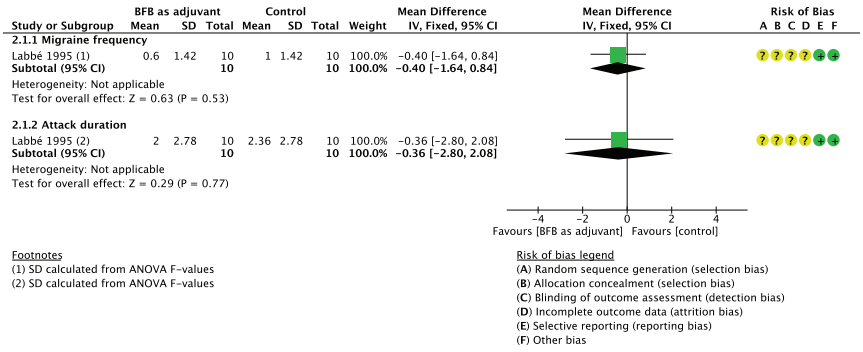
Figure 6: Comparison: Biofeedback (BFB) vs. Waiting-list control (WLC). Outcome: Headache intensity. 1 = Standard deviation estimated form ANOVA F-values; 2 = Standard error derived by hand from graph.



Footnotes
(1) Average peak intensity on ordinal scale from 0-5.
(2) Average peak intensity on ordinal scale from 0-4. Converted to 0-5.

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of outcome assessment (detection bias)
(**D**) Incomplete outcome data (attrition bias)
(**E**) Selective reporting (reporting bias)
(**F**) Other bias

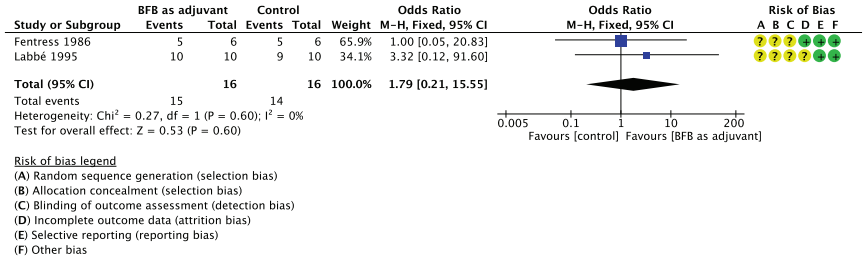
Figure 7: Comparison: Adjuvant effect of biofeedback (BFB). Outcome: Migraine frequency and attack duration. 1, 2 = Standard deviation estimated form ANOVA F-values.



Footnotes
(1) SD calculated from ANOVA F-values
(2) SD calculated from ANOVA F-values

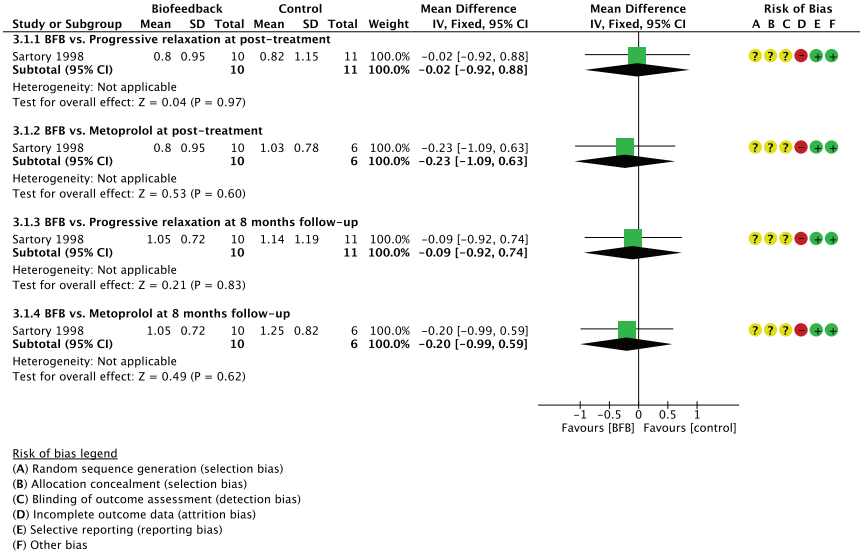
Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of outcome assessment (detection bias)
(**D**) Incomplete outcome data (attrition bias)
(**E**) Selective reporting (reporting bias)
(**F**) Other bias

Figure 8: Comparison: Adjuvant effect of biofeedback (BFB). Outcome: Responder rate.



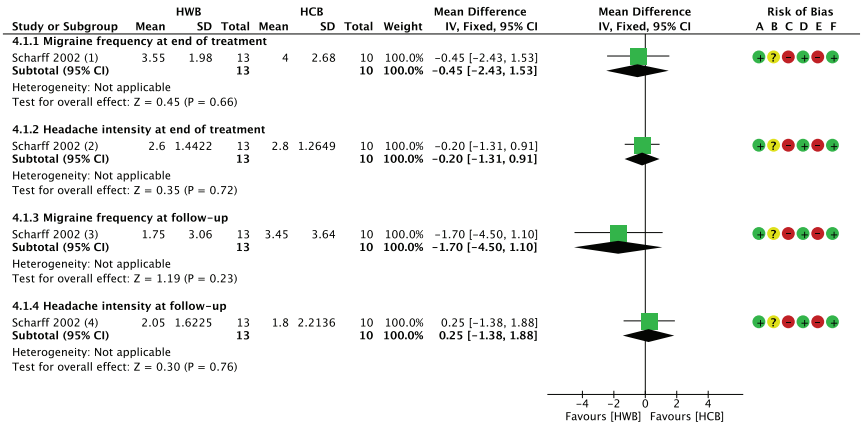
Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of outcome assessment (detection bias)
(**D**) Incomplete outcome data (attrition bias)
(**E**) Selective reporting (reporting bias)
(**F**) Other bias

Figure 9: Comparison: Biofeedback (BFB) vs. Active treatment control. Outcome: Migraine frequency at post-treatment and follow-up.



Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of outcome assessment (detection bias)
(**D**) Incomplete outcome data (attrition bias)
(**E**) Selective reporting (reporting bias)
(**F**) Other bias

Figure 10: Comparison: Biofeedback versus "Sham biofeedback." Outcomes: Migraine frequency and headache intensity. 1,2,3,4 = Standard error derived by hand from graph



Footnotes
(1) Standard error derived by hand from graph
(2) Standard error derived by hand from graph
(3) Standard error derived by hand from graph
(4) Standard error derived by hand from graph

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of outcome assessment (detection bias)
(**D**) Incomplete outcome data (attrition bias)
(**E**) Selective reporting (reporting bias)
(**F**) Other bias

Supplemental information

- Appendix 1. Complete search strategy for all databases searched
- Appendix 2. Characteristics of included study table 1: Fentress 1986
- Appendix 3. Characteristics of included study table 2: Labbé 1984
- Appendix 4. Characteristics of included study table 3: Labbé 1995
- Appendix 5. Characteristics of included study table 4: Sartory 1998
- Appendix 6. Characteristics of included study table 5: Scharff 2002

Paper II

Original Paper

Wireless Surface Electromyography and Skin Temperature Sensors for Biofeedback Treatment of Headache: Validation Study with Stationary Control Equipment

Anker Stubberud[1]; Petter Moe Omland[1,2], PhD; Erling Tronvik[1,3], PhD; Alexander Olsen[4,5], PhD; Trond Sand[1,3], PhD; Mattias Linde[1,3], PhD

[1]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Neurology and Clinical Neurophysiology, St Olavs Hospital, Trondheim, Norway

[3]National Advisory Unit on Headaches, Section of Neurology, St Olavs Hospital, Trondheim, Norway

[4]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

[5]Department of Physical Medicine and Rehabilitation, St Olavs Hospital, Trondheim, Norway

Corresponding Author:
Anker Stubberud
Department of Neuromedicine and Movement Science
Norwegian University of Science and Technology
Nevro Øst, Edvard Griegs Gate 8
Trondheim, 7030
Norway
Phone: 47 73 59 20 20
Email: ankers@stud.ntnu.no

Abstract

Background: The use of wearables and mobile phone apps in medicine is gaining attention. Biofeedback has the potential to exploit the recent advances in mobile health (mHealth) for the treatment of headaches.

Objectives: The aim of this study was to assess the validity of selected wireless wearable health monitoring sensors (WHMS) for measuring surface electromyography (SEMG) and peripheral skin temperature in combination with a mobile phone app. This proof of concept will form the basis for developing innovative mHealth delivery of biofeedback treatment among young persons with primary headache.

Methods: Sensors fulfilling the following predefined criteria were identified: wireless, small size, low weight, low cost, and simple to use. These sensors were connected to an app and used by 20 healthy volunteers. Validity was assessed through the agreement with simultaneous control measurements made with stationary neurophysiological equipment. The main variables were (1) trapezius muscle tension during different degrees of voluntary contraction and (2) voluntary increase in finger temperature. Data were statistically analyzed using Bland-Altman plots, intraclass correlation coefficient (ICC), and concordance correlation coefficient (CCC).

Results: The app was programmed to receive data from the wireless sensors, process them, and feed them back to the user through a simple interface. Excellent agreement was found for the temperature sensor regarding increase in temperature (CCC .90; 95% CI 0.83-0.97). Excellent to fair agreement was found for the SEMG sensor. The ICC for the average of 3 repetitions during 4 different target levels ranged from .58 to .81. The wireless sensor showed consistency in muscle tension change during moderate muscle activity. Electrocardiography artifacts were avoided through right-sided use of the SEMG sensors. Participants evaluated the setup as usable and tolerable.

Conclusions: This study confirmed the validity of wireless WHMS connected to a mobile phone for monitoring neurophysiological parameters of relevance for biofeedback therapy.

(JMIR Biomed Eng 2018;3(1):e1) doi:10.2196/biomedeng.9062

KEYWORDS

biofeedback; mobile phone; app; migraine; pediatric

Introduction

In the emerging era of mobile health (mHealth) and technology, the use of wearable sensors and mobile phone health apps has recently gained attention. This has led to a subcategory of health informatics, labeled mHealth, encompassing the use of mobile phones for medical purposes [1]. In addition to these apps, there is also a wide array of wearable health monitoring sensors (WHMS) [2],which represent a means for patients to access real-time data from a broad range of physiological parameters at home [3-5], thus enabling extensive data acquisition [6]. mHealth is of special interest to the younger generation, which is constantly exposed to and familiarized with such technology. It is also increasing in popularity within the field of headache care and research. In particular, mobile phone–based headache diaries are frequently used [7]. However, there is a potential for extending this mobile technology into the preventive treatment of headache disorders, such as migraine. The bulk of current mHealth research focuses on chronic conditions and delivery of self-educational treatment [8], fitting the description of behavioral headache treatments. Biofeedback, one of the several behavioral headache treatments, is well established and empirically supported [9]. Systematic reviews with meta-analyses demonstrated that biofeedback is effective as a migraine prophylaxis in both the adult and pediatric populations [10,11]. However, the treatment is both time-consuming and costly and therefore not readily available for those in need. Thus, a more optimal approach for behavioral headache treatment has long been sought [12,13]. Biofeedback has the potential to exploit the recent advances in mHealth technology [14,15]. All the while, biofeedback mHealth solutions for other purposes, such as exercise and postcancer swallowing exercises, are being developed [16,17].

Modalities proven effective in biofeedback treatment for headache disorders include surface electromyography (SEMG) and peripheral skin temperature. Both modalities are common in the current development of WHMS [2] and may serve as natural elements in the implementation of biofeedback solutions. Nevertheless, such WHMS sensors have not been validated for use in neurophysiological monitoring for the purpose of biofeedback therapy.

The aim of this study was to assess the validity of WHMS for measuring SEMG and peripheral skin temperature in combination with a mobile phone app. This proof of concept would form the basis for the development of a novel, innovative mHealth system for biofeedback therapy for young persons with primary headache.

Methods

Study Design

In the first phase of the study, we identified suitable WHMS and developed the preliminary software. In the second phase of the study, we recruited healthy volunteers to establish the validity of the chosen WHMS. The study was exploratory in nature, with the main aim to evaluate the validity of the chosen WHMS by assessing the agreement compared with stationary neurophysiological equipment following recommended guidelines for agreement studies [18].

Identification of Sensors

The inclusion criteria and requirements for suitable sensors were (1) wireless setup, (2) small size, (3) low weight, (4) simple to use compared with standard clinical equipment, and (5) low cost.

Software Development

The first version of the app was created as a minimal viable product (MVP). This preliminary version was programmed to serve as the starting point of iterative and incremental rounds of testing [19], allowing subsequent development and fine-tuning of the user interface and software components in an upcoming usability study.

Participants

We considered a sample size of 18 to be sufficient, based on the model for sample size determination in reliability studies presented by Bonett [20] (Multimedia Appendix 1). We set out to recruit 20 healthy volunteers to account for potential dropouts. Participants were recruited as a convenience sample by actively seeking out young individuals from the local research and student community. Exclusion criteria were reduced hearing, vision, or sensibility, and severe neurologic or psychiatric disease.

Equipment

TheNeckSensor (EXPAIN, Oslo, Norway) was selected as the wireless WHMS to measure muscle tension. This is a small, compact bipolar SEMG sensor, with a single SR-R adhesive gel patch containing both electrodes (total patch area, 19.8 cm^2), and no patient ground electrode. For wireless measurement of temperature, we selected the PASPORT Skin/Surface Temperature Probe, PS-2131, combined with PASPORT Temperature sensor, PS-2125, and AirLink, PS-3200 (Pasco, Roseville, CA, USA). Both the sensors transmitted signals via Bluetooth Smart/4.0.

As the stationary equipment, the following AD Instruments (Dunedin, New Zealand) setup was used: (1) SMEG signals recorded with 5-Lead Shielded Lead Wires (MLA2505) and 5-Lead Shielded BioAmp cable (MLA2540) attached to Red Dot 2560 electrodes with a silver/silver-chloride 3.48 cm^2 sensor area (3M Health Care, Germany) fed through a Dual BioAmp, FE135, and PowerLab 8/35; (2) equivalent lead wires, cables, and electrodes for registration of an electrocardiogram (ECG) through a separate Dual BioAmp; and (3) temperature registered through Skin Temperature Pod and Probe, ML309 + MLT422/A fed through PowerLab. The recordings were visualized and analyzed using the LabChart 8 software (AD Instruments, Dunedin New Zealand) installed on a Dell Latitude E4310 laptop.

Experimental Procedure

Participants were seated in a recliner at a 90 degree angle in the neurophysiological laboratory. The 2 electrodes from the NeckSensor were placed over the upper fibers of the right trapezius muscle midway along the line between the spinous

process C7 and the acromion [21,22]. Since simultaneous registrations of SEMG signals from the same location with different sets of surface electrodes are not possible, one set of electrodes from the stationary equipment was placed 2 cm cranially of the NeckSensor, and the other set was placed 2 cm caudally. The interelectrode distance was 4 cm. The "patient ground" electrode for the stationary equipment was placed over the spinous process C7 (Figure 1). The skin beneath the stationary electrodes was washed with alcohol swabs. The 2 skin temperature sensors were attached, without touching each other, to the volar pad of the distal phalange on the second finger with sticky tape, with the stationary sensor placed radially of the 2 sensor electrodes.

Figure 1 shows the scheme of the electrode placements over the upper trapezius fibers. The wireless sensor electrode pair was placed first, midway in the line between the acromion and the spinous process C7. One of the two pairs of stationary sensor electrodes was placed cranially, whereas the other was placed caudally of the wireless sensor electrode pair. The interelectrode distance for each pair was 4 cm.

Initially, each participant was asked to relax for 5 min to allow the skin temperature to increase during relaxation. Relaxation was achieved by asking the participant to do nothing and sit still on the recliner. This served to give a baseline (relaxed) muscle tension measurement. Relaxed trapezius muscle tension (baseline) was recorded in the last 30 s of relaxation. Thereafter, the temperature sensors were detached to allow the measurement of room temperature for the remainder of the procedure. Subsequently, the participant was instructed to complete a series of exercises to activate the upper fibers of the trapezius muscle. Arbitrary angle isometric maximal voluntary contraction (MVC), through shoulder elevation, was completed in 3 repetitions, each lasting for 6 s [22-25]. The SEMG and force were simultaneously registered. The force was recorded by a dynamometer (Manual Muscle Tester, Lafayette Instruments, USA) attached to a fixed sling placed over the acromion. Subsequently, the participant was asked to complete similar sets of contractions at 50% (VC50) and 25% (VC25) of maximal contraction guided by a sound signal from the dynamometer elicited at a corresponding set force. Finally, the participant was asked to complete 4 repetitions of static contractions (15 s each) performed by abducting both shoulders to a 90 degree angle and holding them against gravity [22].

After completing the exercises, the participant was asked to answer a 5-item user evaluation questionnaire. Of these, 3 questions had reply options on a 5-point Likert scale, ranging from "Very dissatisfied" to "Very satisfied," while the remaining 2 questions were open for free comments (Table 1).

Figure 1. Electrode placement.

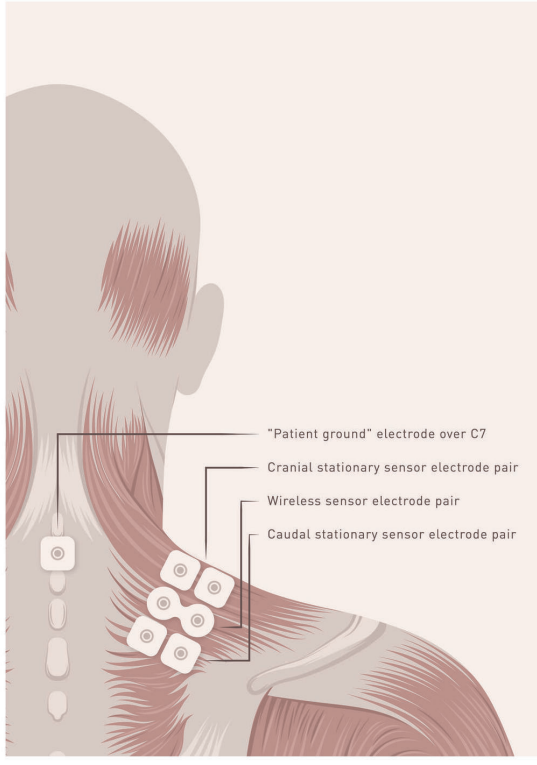


Table 1. Evaluation questionnaire.

Item	Question
1	Did you perceive the wireless sensors as practical to use?
2	To what degree did you feel that the use of shoulder-musculature reflected the feedback in the app?
3	Do you recognize the wireless sensors as safe to use?
4	Did you experience any undesirable harmful effects (if yes, please explain)?
5	Do you have any further comments (if yes, please explain)?

Data Management

The NeckSensor uses a 12-bit ADC resolution sampled at 1024 Hz with a third order 10-480 Hz active bandpass filter. The sensor was programmed to calculate and transmit mean square values internally, with a window width of 40 ms, with no overlap, and a frequency of 25 Hz in order not to overload the Bluetooth capacity. The PowerLab sampled the SEMG signals at 2000 Hz with a fourth order Bessel lowpass filter at 500 Hz and a first order high pass filter at 10 Hz. In addition, a 50 Hz notch analog filter was applied [26]. All stationary recordings were evaluated visually for the presence of ECG artifacts. If found, these were to be corrected by removing the spike-correlated area in the SEMG signal and subsequently replacing the gap with surrounding SEMG activity.

First, the stationary readings were root mean square (RMS) rectified and then averaged over the two sets of electrodes to avoid phase-cancellations. The RMS value was calculated from the mean square values of the wireless sensors. The RMS values for each muscle contraction exercise to be used in the analyses were calculated as the mean of the repetitions for both equipment sets. For the temperature measurements, we calculated the difference in temperature from the start to the end of relaxation and the difference between the temperature at the end of relaxation and room temperature.

Statistics

The means and SD for the RMS values during trapezius muscle exercises and the chosen data temperature points were calculated. Systematic differences between stationary and wireless equipment were assessed with the Wilcoxon signed-rank test.

Mean difference (MD) and limits of agreement (LOA), together with Bland-Altman plots were used as descriptive tools [27].We calculated the intraclass correlation coefficient (ICC) with a two-way, mixed-effects consistency of agreement model. Coefficients for both individual and average agreement were presented. In addition, we calculated the Lin concordance correlation coefficient (CCC) [28-30]. For the ICC and CCC analyses, the data was first transformed to meet assumptions for a two-way analysis of variance model. Then the data was transformed by calculating the natural logarithm after adding 0.1 as a constant to adjust for values being close to zero. The ICC values were interpreted as suggested by Cicchetti et al [31], that is, unacceptable or poor (.00-.40), fair (.41-.60), good (.61-.75), and excellent (.75-1.00). All data were analyzed by using the statistical package Stata version 14 (StataCorp, College Station, TX, USA).

Results

Sensors and Software

The WHMS fulfilling the predefined requirements were identified through pragmatic Internet-searches. The MVP version of the app used in the experimental procedure was programmed to receive data from the wireless sensors and feed raw data back to the user. The raw data were presented as two columns increasing in height with increase in muscle tension and temperature, respectively. The app was programmed to allow connection of any WHMS using Bluetooth.

Participants

A total of 20 healthy participants were recruited and completed the experimental procedure. Of these, 12 were male participants, and their mean age was 24.7 years (SD 2.7, range 18-29 years).

Surface Electromyography Sensor Agreement

We observed no ECG artifacts in the SEMG recordings (Figure 2). Hence, the ECG-related elements were not removed from the SEMG recordings.

Figure 2 shows the raw data of the SEMG activity for the wireless sensor (red), anterior stationary sensor (blue), and posterior stationary sensor (green) from a 24-year-old male participant. The marked areas indicate where the different exercises are performed. The figure exemplifies the absence of ECG artifacts and the similarity of the signals.

Means and standard deviations of the RMS values for the trapezius muscle exercises are presented in Table 2. The wireless sensor showed a lower voltage during trapezius muscle exercises than during all contraction periods and at baseline.

Table 3 summarizes the MD in millivolts (mV) between stationary and wireless equipment with corresponding LOA, for each of the exercises. Compared with the wireless equipment, the stationary equipment indicated a systematically higher voltage during MVC (0.25 mV), VC50 (0.11 mV), VC25 (0.06 mV), static hold (0.07 mV), and baseline (0.04 mV). A Bland-Altman plot, visually presenting the MD and LOA for VC25, is shown in Figure 3. Table 3 also summarizes the ICC and CCC values for the SEMG equipment comparisons.

Figure 2. Raw surface electromyography (SEMG) data. ECG: electrocardiogram; MVC: maximal voluntary contraction; RMS: root mean square; VC50: voluntary contraction at 50% force; VC25: voluntary contraction at 25% force.

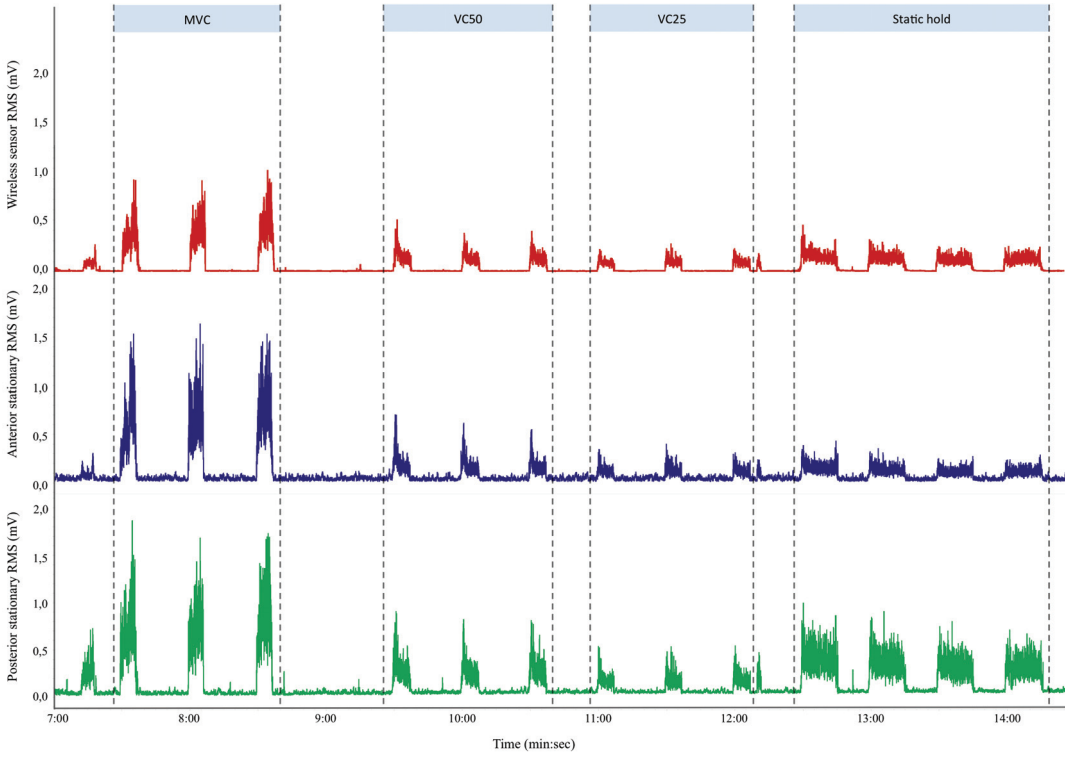


Table 2. Comparison of the means for stationary and wireless equipment.

Exercise	Stationary equipment (SD)	Wireless equipment (SD)	Z-value (*P* value)[a]
MVC[b]	0.62[c] (0.25)	0.37 (0.15)	3.73 (<.001)
VC50[d]	0.26 (0.11)	0.15 (0.06)	3.92 (<.001)
VC25[e]	0.15 (0.05)	0.09 (0.05)	3.73 (<.001)
Static hold	0.16 (0.06)	0.08 (0.03)	3.85 (<.001)
Baseline	0.045 (0.004)	0.01 (0.002)	3.92 (<.001)
Start temperature	28.8[f] (3.4)	28.8 (3.3)	0.75 (=.46)
End temperature	30.7 (3.6)	31.5 (4.0)	3.4 (<.001)
Room temperature	23.0 (0.3)	23.6 (0.4)	3.9 (<.001)

[a]Z-value from Wilcoxon signed-rank test.

[b]MVC: maximal voluntary contraction.

[c]Mean voltage in millivolts RMS.

[d]VC50: voluntary contraction at 50% force.

[e]VC25: voluntary contraction at 25% force.

[f]Mean temperature in degrees Celsius.

Table 3. Indices of agreement between stationary and wireless equipment.

Exercise	Mean difference	Limits of agreement	ICC[a] (95% CI) individual	ICC (95% CI) average	CCC[b] (95% CI)
MVC[c]	0.25[f]	−0.12 to 0.61	.81 (0.57-0.92)	.89 (0.73-0.96)	.52 (0.30-0.73)
VC50[d]	0.11	−0.04 to 0.27	.81 (0.57-0.92)	.89 (0.73-0.96)	.44 (0.23-0.64)
VC25[e]	0.06	−0.03 to 0.15	.66 (0.31-0.85)	.79 (0.47-0.92)	.37 (0.14-0.60)
Static hold	0.07	−0.02 to 0.16	.58 (0.19-0.81)	.73 (0.32-0.89)	.26 (0.06-0.45)
Baseline	0.04	0.03-0.04	.50 (0.09-0.77)	.67 (0.16-0.87)	.01 (0.00-0.01)
Start to end temperature	−0.77[g]	−1.90 to 0.35	.96 (0.91-0.99)	.98 (0.95-0.99)	.90 (0.83-0.97)
End to room temperature	−0.23	−1.74 to 1.28	.98 (0.95-0.99)	.99 (0.97-1.0)	.98 (0.96-1.0)

[a]ICC: intraclass correlation coefficient.

[b]CCC: concordance correlation coefficient.

[c]MVC: maximal voluntary contraction.

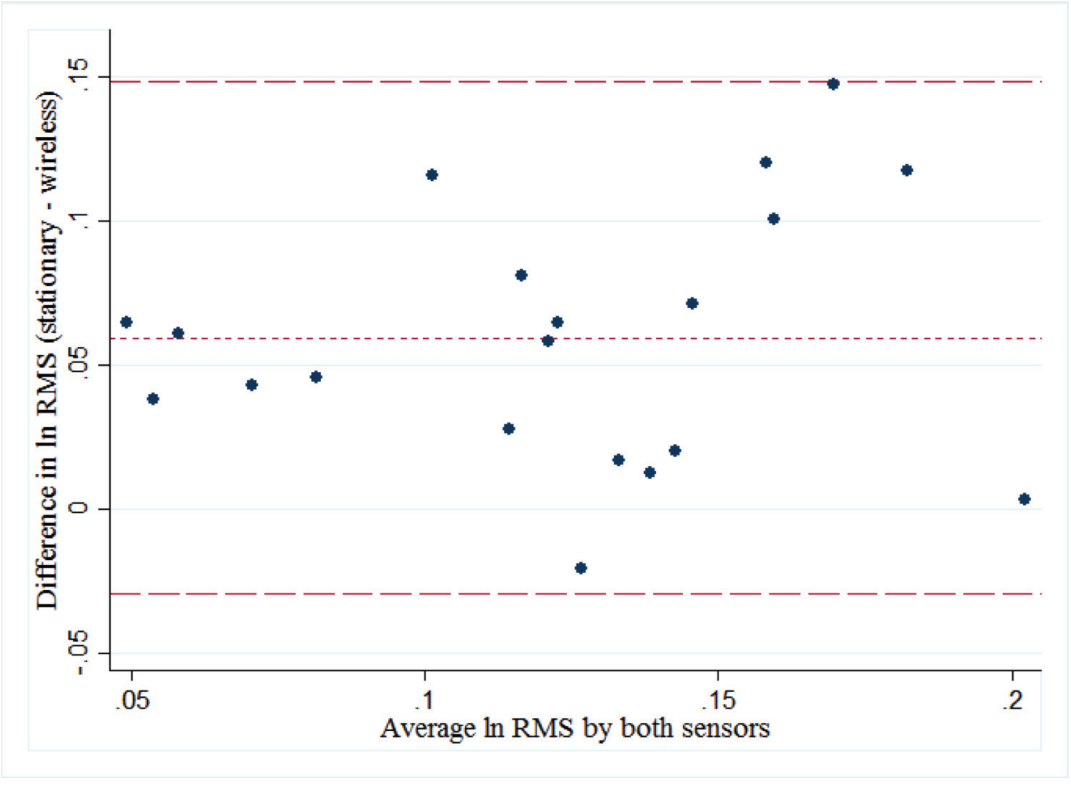
[d]VC50: voluntary contraction at 50% force.

[e]VC25: voluntary contraction at 25% force.

[f]Mean voltage in millivolts RMS.

[g]Mean temperature in degrees Celsius.

Figure 3. Surface electromyography (SEMG) sensor agreement. mV: millivolts; RMS: root mean square.



Excellent agreement was found for MVC (ICC .81, 95% CI 0.57-0.92) and VC50 (ICC .81, 95% CI 0.57-0.92). Good agreement was found for VC25 (ICC .66, 95% CI 0.31-0.85). Fair agreement was found for static hold (ICC .58, 95% CI 0.19-0.81) and baseline (ICC .50, 95% CI 0.09-0.77). All participants displayed a decrease in voltage from MVC to VC50, from VC50 to VC25, and from static hold to baseline for both sets of equipment, with the exception of one participant who had a small increase (0.03 mV) in voltage from VC50 to VC25 registered on the stationary equipment (Figure 4).

Figure 3 shows Bland-Altman plot assessing the agreement between stationary and wireless SEMG sensors during voluntary contraction at 25% force. The x-axis represents the average of the two parallel measurements. The y-axis represents the corresponding difference between the 2 measurements. The values are indicated in millivolt RMS.

Figure 4 is a line graph showing the SEMG readings for each participant during MVC, VC50, VC25, static hold, and baseline. The top panel indicates readings with the stationary equipment. The bottom panel indicates readings with the wireless equipment. The values are indicated in millivolt RMS.

Peripheral Skin Temperature Sensor Agreement

Means and standard deviations of the temperature measurements at the 3 selected time points are shown in Table 2. The start temperature between the 2 sets of equipment did not differ significantly (P=.46), but the wireless sensor indicated a higher temperature at the end of relaxation (P<.001) and at room temperature (P<.001; Table 2).

The between-equipment MDs for changes in the temperature are presented in Table 3, along with the LOA and agreement indices. A Bland-Altman plot visually representing the MD and LOA for temperature change during relaxation is depicted in Figure 5. Excellent agreement was found for the change in temperature during relaxation (CCC .90, 95% CI 0.83-0.97) and from end of relaxation to room temperature (CCC .98, 95% CI 0.96-1.0). A rise in temperature was detected among 17 participants on the stationary equipment, and among 18 participants on the wireless equipment. Moreover, a rise in temperature of more than 1°C was detected among 15 participants on both equipment sets (Figure 6).

Figure 5 is a Bland-Altman plot showing the agreement between stationary and wireless equipment for the change in temperature from start to end of relaxation. The x-axis represents the average of the 2 parallel measurements. The y-axis represents the corresponding difference in measurements. The values are in degrees Celsius.

Figure 6 is a line graph showing temperature readings for each participant at the start and end of relaxation and at room temperature. The upper panel represents readings with the stationary equipment. The lower panel represents readings with the wireless equipment. The values are in degrees Celsius.

Evaluation Questionnaire

In total, 19 of the 20 participants perceived the use of wireless sensors as practical (n=14) or very practical (n=5). Likewise, the absolute majority of participants reported that the app feedback reflected the use of shoulder musculature to a large (n=9) or a very large (n=9) degree. All participants regarded the use of wireless sensors as safe (n=2) or very safe (n=18). In contrast, 2 of the 20 participants reported undesirable, harmful effects, with both stating that the removal of the electrodes attached to the stationary equipment was unpleasant.

Figure 4. Surface electromyography (SEMG) sensor line graphs. mV: millivolts; MVC: maximal voluntary contraction; RMS: root mean square; VC50: voluntary contraction at 50% force; VC25: voluntary contraction at 25% force.

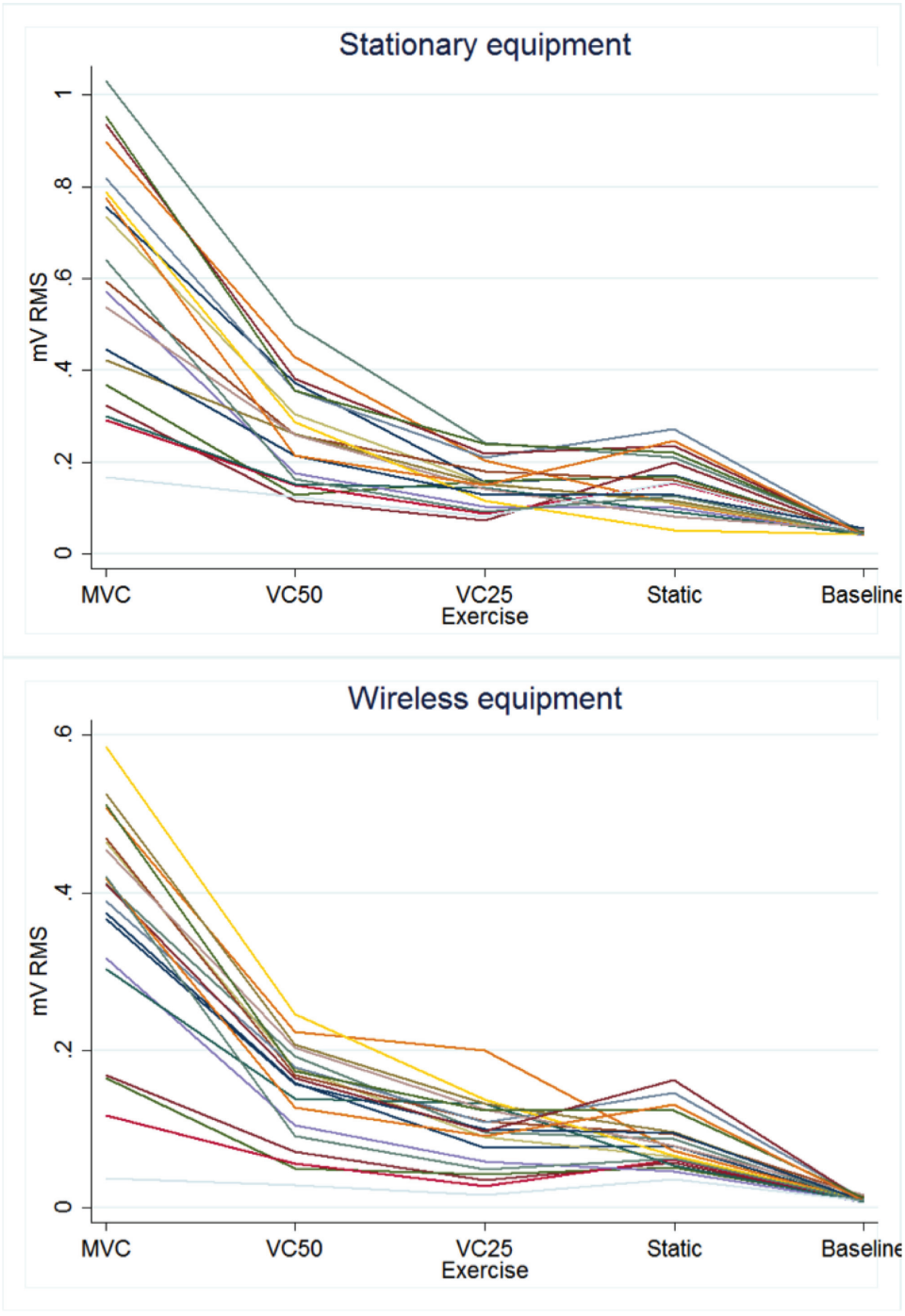


Figure 5. Temperature sensor agreement. mV: millivolts; RMS: root mean square.

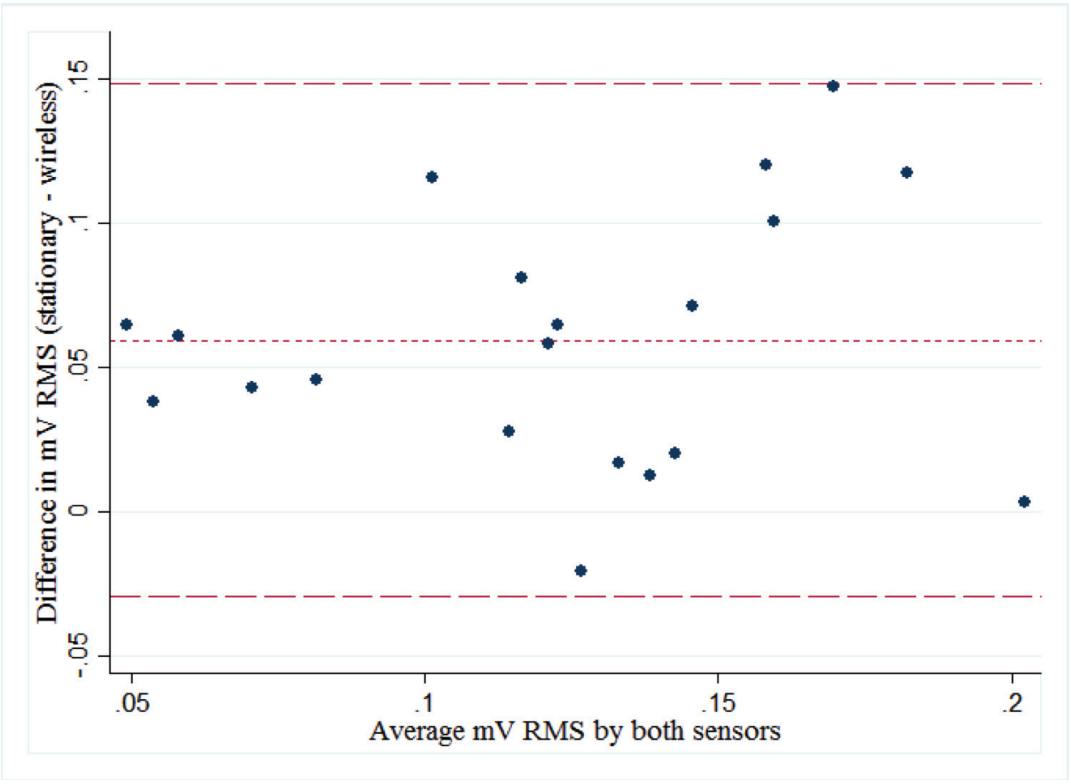
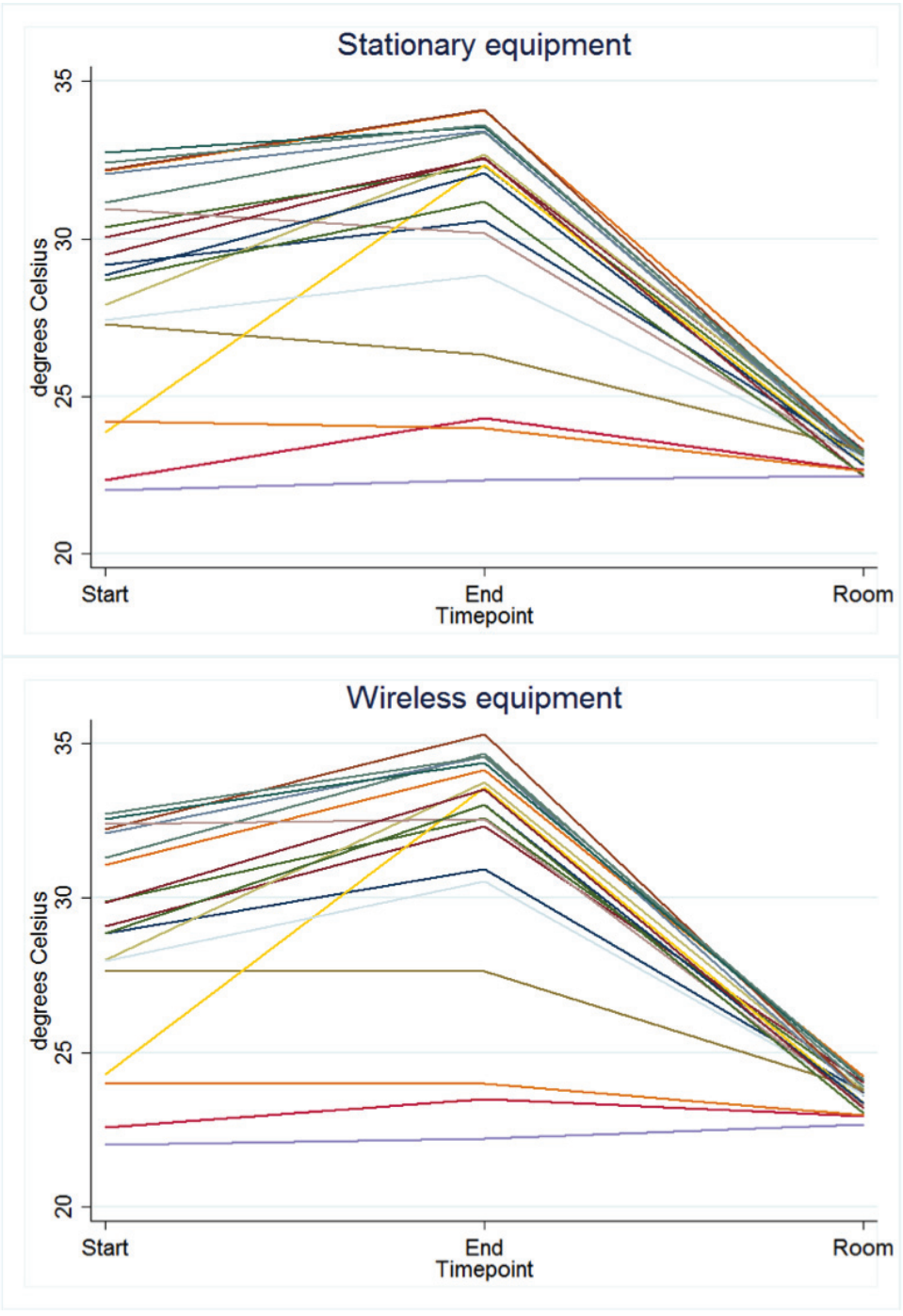


Figure 6. Temperature sensor line graphs.



Discussion

Principal Findings

This study aimed to provide a proof of concept for using a mobile phone and WHMS for biofeedback purposes, in a fashion similar to phase I-II development of new drug treatments [32]. We chose to investigate temperature and SMEG because they are the most commonly used biofeedback modalities [11] and are shown to be especially effective in adolescents [33]. We identified sensors fulfilling a set of predefined criteria that were considered necessary for the sensors to gain acceptance among patients, and thus these sensors were used [34]. The choice of sensors was arbitrary, as long as the predefined criteria were met. Even though the use of other temperature and SEMG sensors would not yield identical results, we argue that our approach has provided a proof of concept.

We found that the use of a wireless temperature sensor had almost perfect agreement regarding the change in finger temperature during relaxation. Furthermore, the use of a wireless SEMG sensor had a fair to excellent agreement for measuring tension in the trapezius muscle. We noted that the wireless SEMG consistently showed a lower voltage than the stationary equipment. The SEMG sensors showed excellent agreement during MVC and VC50, good agreement during VC25, and fair agreement during static hold and baseline. However, under the assumption that the stationary equipment was the most sensitive, it is not surprising that the calculated agreement decreased slightly at lower activity levels since random and equipment-generated noise constituted a larger part of the signal at low EMG-levels. Nonetheless, the wireless SEMG sensor registered consistent changes in muscle tension. We observed no ECG artifacts in the SEMG recordings. Therefore, it can be assumed that the ECG artifacts do not have a relevant influence on the SEMG recorded from closely placed bipolar electrodes on the right shoulder. Moreover, the safety and usability of the setup were highly satisfactory. In conclusion, the wireless sensors are well suited for biofeedback purposes.

Strengths and Limitations

The proper sample size for the study was assessed before recruiting participants (Multimedia Appendix 1). Traditionally, a sample of 15 to 20 participants is deemed sufficient for reliability studies [35]. However, the use of more precise calculations of sample sizes has been previously suggested [36]. Therefore, we used a CI estimation model suggested by Bonett [20] to determine the minimum sample size required. Due to the interindividual variation in our findings, the analyses would possibly have benefited from having a larger sample size because we did not obtain a predefined CI for all analyses.

There is a large degree of variability in individual human anatomical properties that may influence SEMG readings. This includes the thickness of fatty tissues, resting muscle length, velocity of contraction, muscle cross-sectional area, fiber type, posture change, interelectrode distance, skin impedance, age, and sex [22]. We chose to combine the recordings for the 2 pairs of stationary sensor electrodes to approximate the muscle activity of the wireless sensor placed in between. The relative spread of the electrode pairs may have led to EMG crosstalk, and muscle contraction exercises performed by untrained participants may have additionally resulted in movement artifacts, and suboptimal and varying performances [37]. The abovementioned factors may all have limited the precision of our measurements and contributed to a larger degree of interindividual differences, thus lowering individual ICC and CCC values for SEMG agreement. Likewise, the placement of the 2 temperature sensors beside each other on the finger might have led to differences in measurements. Figure 5 shows 1 outlier that displayed a larger increase in temperature by 1°C with the stationary equipment than with the wireless equipment. This differs from the majority that displayed the largest temperature increase with the wireless equipment. Nevertheless, LOA of ±1.5°C is still acceptable [38].

The SEMG signals usually have a frequency distribution with significant energy up to 400 to 500 Hz, requiring a sampling frequency of at least 1000 Hz (preferably 2000 Hz) to meet the Nyquist rate (2 times higher signal frequency) and avoid the so-called aliasing [39]. However, it is known that oversampling above this critical Nyquist rate does not significantly improve the signal quality [40] but will likely lead to higher cost and size of the sensor. The SEMG signals are usually bandpass filtered at 10 to 500 Hz [41], which we consequently chose to do for both setups. Furthermore, we observed that the notch filter, at 50 Hz, for the stationary equipment seemed to be saturated during recordings. After analog filtering, sine waves of 20 ms duration were still present. This may be explained by power-line noise, despite the use of a notch filter [42]. The wireless sensor also applies a notch filter at 50 Hz, which increases the signal-to-noise ratio. In total, we concluded that the wireless SEMG sensor applies appropriate signal processing settings.

We chose different statistical methods for assessing agreement to evaluate different properties of the wireless sensors. The Wilcoxon signed-ranks tests, together with the Bland-Altman plot and LOA, assess the degree of systematic differences and expected variance between measurements. A two-way, mixed-effect ICC model [43] ignores the element of rater variance (raters fixed as the 2 equipment sets), and the estimate can thus serve as an index of consistency [28,30,44,45]. This is useful to assess agreement when having mean differences between 2 measurement methods. We reported both individual and average ICC values, as the average value becomes useful when a large degree of interindividual variance exists or if individual readings are considered unreliable [30]. On the other hand, we also calculated the CCC to evaluate the degree of absolute agreement, that is, the 2 measurement methods showing identical values.

Interpretation

We have compared the WHMS with a gold standard; however, this does not imply that the gold standard is without measurement error. Thus, some lack of agreement is inevitable [46]. As pointed out by Bland and Altman [47], one should keep in mind that correlation coefficients alone do not assess interchangeability of measurement methods. The acceptable level of agreement in order to claim validity is a clinical decision. Considering the intended use of the chosen sensors,

a high degree of absolute agreement is not a necessity, but consistency of agreement is important. We certainly observed that there exists variance in the data, leading to a low degree of absolute agreement. On the other hand, SEMG readings changed similarly and as expected through the experimental procedure for each participant, despite dissimilarities between the 2 equipment sets. This consistency is indeed supported by excellent to fair agreement of ICC values. Furthermore, the wireless SEMG sensor was less reliable at lower voltage, at least in terms of absolute agreement, when compared with our gold standard. A well-designed SEMG setup usually produces a system noise of about 1% of the MVC [48]. Our stationary equipment baseline showed 7% of MVC, which means that there was some inherent noise in the gold standard setup. In contrast, the baseline readings of the wireless sensors amounted to 3% of MVC, which in part may explain the increasing deviation at lower voltages.

Although the SEMG sensor did not demonstrate excellent agreement in all analyses, both SEMG and temperature WHMS appear to be suited for app-based biofeedback. Interestingly, 15 out of 20 participants (75%) managed to raise their temperature by more than 1°C during a single naive session indicating that the setup was simple to master. Moreover, all participants had similar changes in muscle tension through the sets of exercises. However, it is unlikely that the users will be able to decrease their muscle tension throughout the entire duration of a biofeedback session [49]. This means that detecting a change in tension is more important than the absolute values. In line with this, it was recently shown that the feedback itself is more important than lowering muscle tension in the treatment of headache [50]. Taken together, these findings imply that perfect sensor agreement in itself is not a prerequisite for an app-based biofeedback platform. The main focus of app-based biofeedback should be directed at the development of high-quality feedback mechanisms and user interfaces.

Prospects for Future Research

This study confirmed the usability of WHMS in a biofeedback setting and established partial evidence for an upcoming biofeedback app. At any rate, the scientific validation of the sensor is of utmost importance for the value and effectiveness of a future treatment program. The choice to use an MVP app to assess agreement enables iterative and incremental developments. Future research should be carried out to establish further the basis for the use of WHMS for medical purposes in the emerging era of health informatics and mHealth. As an example, similar validation of heart rate variability measurements, which is of interest in biofeedback treatment, has been conducted [51,52,53]. We are currently exploring the user interface and assessing the usability of the app among adolescents with migraine.

Conclusions

This study confirmed the validity of wireless WHMS connected to a mobile phone for monitoring neurophysiological parameters of relevance for biofeedback therapy.

Acknowledgments

The authors wish to thank all the volunteers for participating in the study. The study was funded by strategic seeding grants from the Faculty of Medicine, NTNU Norwegian University of Science and Technology. We would also like to thank Searis AS for the fruitful collaboration and for their help with programming the app, EXPAIN AS for supplying SEMG sensors for use in the study, and the personnel at the Department of Neurophysiology, St Olavs Hospital, for their support with the experimental procedures.

Conflicts of Interest

Anker Stubberud has participated as a nonpaid member of an Expert Panel advising EXPAIN AS during the final phases of product development. Should the research result in a commercially available product, the university and authors may benefit financially from future intellectual property rights.

Multimedia Appendix 1

Sample size determination.

[PDF File (Adobe PDF File), 32KB - biomedeng_v3i1e1_app1.pdf]

References

1. Kay M, Santos J, Takane M. Who.int. 2011. mHealth: new horizons for health through mobile technologies URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2018-02-13] [WebCite Cache ID 6xCFk9ovq]
2. Pantelopoulos A, Bourbakis NG. A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans Syst Man Cybern C Appl Rev 2010;40(1):1-12. [doi: 10.1109/TSMCC.2009.2032660]
3. Dobkin BH, Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. Neurorehabil Neural Repair 2011;25(9):788-798 [FREE Full text] [doi: 10.1177/1545968311425908] [Medline: 21989632]
4. Hanson MA, Powell Jr HC, Barth AT, Ringenberg K, Calhoun BH, Aylor JH, et al. Body area sensor networks: challenges and opportunities. Computer 2009;42(1):58-65.

5. Schüll ND. Data for life: wearable technology and the design of self-care. BioSocieties 2016;11(3):317-333 [FREE Full text]
6. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. Crit Public Health 2013;23(4):393-403. [doi: 10.1080/09581596.2013.794931]
7. Hundert AS, Huguet A, McGrath PJ, Stinson JN, Wheaton M. Commercially available mobile phone headache diary apps: a systematic review. JMIR Mhealth Uhealth 2014;2(3):e36 [FREE Full text] [doi: 10.2196/mhealth.3452] [Medline: 25138438]
8. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
9. Penzien DB, Irby MB, Smitherman TA, Rains JC, Houle TT. Well-established and empirically supported behavioral treatments for migraine. Curr Pain Headache Rep 2015 Jul;19(7):34. [doi: 10.1007/s11916-015-0500-5] [Medline: 26065542]
10. Stubberud A, Varkey E, McCrory DC, Pedersen SA, Linde M. Biofeedback as prophylaxis for pediatric migraine: a meta-analysis. Pediatrics 2016 Aug;138(2) [FREE Full text] [doi: 10.1542/peds.2016-0675] [Medline: 27462067]
11. Nestoriuc Y, Martin A. Efficacy of biofeedback for migraine: a meta-analysis. Pain 2007 Mar;128(1-2):111-127. [doi: 10.1016/j.pain.2006.09.007] [Medline: 17084028]
12. Andrasik F. Behavioral treatment of headaches: extending the reach. Neurol Sci 2012 May;33 Suppl 1:S127-S130. [doi: 10.1007/s10072-012-1073-2] [Medline: 22644187]
13. Schwartz MS, Andrasik F, editors. Biofeedback: A Practitioner's Guide. New York City: Guilford Press; 2017.
14. Minen MT, Torous J, Raynowska J, Piazza A, Grudzen C, Powers S, et al. Electronic behavioral interventions for headache: a systematic review. J Headache Pain 2016;17:51 [FREE Full text] [doi: 10.1186/s10194-016-0608-y] [Medline: 27160107]
15. Luxton DD, McCann RA, Mishkind MC, Reger GM, Bush NE. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pr 2011;42(6):505-512 [FREE Full text] [doi: 10.1037/a0024485]
16. Constantinescu G, Loewen I, King B, Hodgetts W, Rieger J, Brodt C. Designing a mobile health app for patients with dysphagia following head and neck cancer: a qualitative study. JMIR Rehabil Assist Technol 2017;4(1):e3. [Medline: 28582245]
17. O'Reilly M, Duffin J, Ward T, Caulfield B. Mobile app to streamline the development of wearable sensor-based exercise biofeedback systems: system development and evaluation. JMIR Rehabil Assist Technol 2017;4(2):e9. [doi: 10.2196/rehab.7259]
18. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. Int J Nurs Stud 2011 Jun;48(6):661-671. [doi: 10.1016/j.ijnurstu.2011.01.016] [Medline: 21514934]
19. Larman C, Basili VR. Iterative and incremental developments: a brief history. Comput 2003;36(6):47-56. [doi: 10.1109/MC.2003.1204375]
20. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med 2002 May 15;21(9):1331-1335. [doi: 10.1002/sim.1108] [Medline: 12111881]
21. Hermens HJ, Freriks B, Disselhorst-Klug C, Rau G. Development of recommendations for SEMG sensors and sensor placement procedures. J Electromyogr Kinesiol 2000 Oct;10(5):361-374. [Medline: 11018445]
22. Criswell E. Cram's Introduction to Surface Electromyography. 2nd edition. Sudbury, MA: Jones & Bartlett Publishers; 2010.
23. Mathiassen SE, Winkel J, Hägg GM. Normalization of surface EMG amplitude from the upper trapezius muscle in ergonomic studies - A review. J Electromyogr Kinesiol 1995 Dec;5(4):197-226. [Medline: 20719652]
24. Burden A. How should we normalize electromyograms obtained from healthy participants? What we have learned from over 25 years of research. J Electromyogr Kinesiol 2010 Dec;20(6):1023-1035. [doi: 10.1016/j.jelekin.2010.07.004] [Medline: 20702112]
25. Knutson LM, Soderberg GL, Ballantyne BT, Clarke WR. A study of various normalization procedures for within day electromyographic data. J Electromyogr Kinesiol 1994;4(1):47-59. [doi: 10.1016/1050-6411(94)90026-4] [Medline: 20870546]
26. Chowdhury RH, Reaz MB, Ali MA, Bakar AA, Chellappan K, Chang TG. Surface electromyography signal processing and classification techniques. Sensors (Basel) 2013 Sep 17;13(9):12431-12466 [FREE Full text] [doi: 10.3390/s130912431] [Medline: 24048337]
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
28. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
29. Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. Theriogenology 2010 Jun;73(9):1167-1179 [FREE Full text] [doi: 10.1016/j.theriogenology.2010.01.003] [Medline: 20138353]
30. Barnhart HX, Haber MJ, Lin LI. An overview on assessing agreement with continuous measurements. J Biopharm Stat 2007;17(4):529-569. [doi: 10.1080/10543400701376480] [Medline: 17613641]

31. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
32. Schmidt B. Proof of Principle studies. Epilepsy Res 2006 Jan;68(1):48-52. [doi: 10.1016/j.eplepsyres.2005.09.019] [Medline: 16377153]
33. Sarafino EP, Goehring P. Age comparisons in acquiring biofeedback control and success in reducing headache pain. Ann Behav Med 2000;22(1):10-16. [Medline: 10892524]
34. Bergmann JH, McGregor AH. Body-worn sensor design: what do patients and clinicians want? Ann Biomed Eng 2011 Sep;39(9):2299-2312. [doi: 10.1007/s10439-011-0339-9] [Medline: 21674260]
35. Fleiss JL. Design and Analysis of Clinical Experiments. Hoboken, NJ: John Wiley & Sons; 2011.
36. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998 Jan 15;17(1):101-110. [Medline: 9463853]
37. Merletti R, di Torino P. Standards for reporting EMG data. J. Electromyogr Kinesiol 1999;9(1):3-4 [FREE Full text]
38. Kelechi TJ, Michel Y, Wiseman J. Are infrared and thermistor thermometers interchangeable for measuring localized skin temperature? J Nurs Meas 2006;14(1):19-30. [Medline: 16764175]
39. Ives JC, Wigglesworth JK. Sampling rate effects on surface EMG timing and amplitude measures. Clin Biomech (Bristol, Avon) 2003 Jul;18(6):543-552. [Medline: 12828904]
40. Durkin JL, Callaghan JP. Effects of minimum sampling rate and signal reconstruction on surface electromyographic signals. J Electromyogr Kinesiol 2005 Oct;15(5):474-481. [doi: 10.1016/j.jelekin.2005.02.003] [Medline: 15935959]
41. van Boxtel A. Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. Psychophysiol 2001 Jan;38(1):22-34. [Medline: 11321618]
42. De Luca CJ, Gilmore LD, Kuznetsov M, Roy SH. Filtering the surface EMG signal: movement artifact and baseline noise contamination. J Biomech 2010 May 28;43(8):1573-1579. [doi: 10.1016/j.jbiomech.2010.01.027] [Medline: 20206934]
43. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. Stat Med 1994;13(23-24):2465-2476. [Medline: 7701147]
44. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1(1):30-46. [doi: 10.1037/1082-989X.1.1.30]
45. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1966 Aug;19(1):3-11. [doi: 10.2466/pr0.1966.19.1.3] [Medline: 5942109]
46. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun;8(2):135-160. [Medline: 10501650]
47. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med 1990;20(5):337-340. [Medline: 2257734]
48. Clancy EA, Farry KA. Adaptive whitening of the electromyogram to improve amplitude estimation. IEEE Trans Biomed Eng 2000 Jun;47(6):709-719. [doi: 10.1109/10.844217] [Medline: 10833845]
49. Rausa M, Palomba D, Cevoli S, Lazzerini L, Sancisi E, Cortelli P, et al. Biofeedback in the prophylactic treatment of medication overuse headache: a pilot randomized controlled trial. J Headache Pain 2016 Dec;17(1):87 [FREE Full text] [doi: 10.1186/s10194-016-0679-9] [Medline: 27655371]
50. Rains JC. Change mechanisms in EMG biofeedback training: cognitive changes underlying improvements in tension headache. Headache 2008 May;48(5):735-6; discussion 736. [doi: 10.1111/j.1526-4610.2008.01119_1.x] [Medline: 18471128]
51. Munster-Segev M, Fuerst O, Kaplan SA, Chan A. Incorporation of a stress reducing mobile app in the care of patients with type 2 diabetes: a prospective study. JMIR mHealth and uHealth 2017;5(5):e75. [Medline: 28554881]
52. Uddin AA, Morita PP, Tallevi K, Armour K, Li J, Nolan RP, et al. Development of a wearable cardiac monitoring system for behavioral neurocardiac training: a usability study. JMIR Mhealth Uhealth 2016;4(2):e45 [FREE Full text] [doi: 10.2196/mhealth.5288] [Medline: 27106171]
53. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR mHealth and uHealth 2017;5(3):e34. [Medline: 28302596]

Abbreviations

CCC: concordance correlation coefficient
ECG: electrocardiogram
ICC: intraclass correlation coefficient
LOA: limits of agreement
MD: mean difference
mHealth: mobile health
MVC: maximal voluntary contraction
MVP: minimal viable product

RMS: root mean square
SEMG: surface electromyography
VC50: voluntary contraction at 50% force
VC25: voluntary contraction at 25% force
WHMS: wearable health monitoring sensors

Edited by G Eysenbach; submitted 28.09.17; peer-reviewed by M Minen, YCP Arai; comments to author 09.12.17; revised version received 21.01.18; accepted 23.01.18; published 23.02.18

Please cite as:
Stubberud A, Omland PM, Tronvik E, Olsen A, Sand T, Linde M
Wireless Surface Electromyography and Skin Temperature Sensors for Biofeedback Treatment of Headache: Validation Study with Stationary Control Equipment
JMIR Biomed Eng 2018;3(1):e1
URL: http://biomedeng.jmir.org/2018/1/e1/
doi:10.2196/biomedeng.9062
PMID:

Paper III

Headache
© 2020 The Authors. Headache: The Journal of Head and Face Pain
published by Wiley Periodicals, Inc. on behalf of American Headache Society.
ISSN 0017-8748
doi: 10.1111/head.13772
Published by Wiley Periodicals, Inc.

Research Submission

Biofeedback Treatment App for Pediatric Migraine: Development and Usability Study

Anker Stubberud, MD ; Erling Tronvik, PhD ; Alexander Olsen, PhD ; Gøril Gravdahl, MSc; Mattias Linde, PhD

Objective.—The objective of this study was to develop and investigate the usability of a biofeedback treatment smartphone app for adolescent migraine sufferers.

Background.—Biofeedback is effective in treating pediatric migraine. However, biofeedback is not widely used due to the necessity of a trained therapist and specialized equipment. Emerging digital technology, including smartphones and wearables, enables new ways of administering biofeedback.

Methods.—In a prospective open-label development and usability study, 10 adolescent migraine sufferers used a newly developed biofeedback app with wearable sensors that measured their muscle tension, finger temperature, and heart rate. Three iterative rounds of usability testing, including a 2-week home testing period, were completed. A biofeedback algorithm, combining and optimizing the 3 physiological modalities, and several algorithms for sham-treatment were created. Usability was evaluated statistically and summarized thematically.

Results.—Five of ten participants completed all 3 rounds of usability testing. A total of 72 biofeedback sessions were completed. Usability scoring was consistently high, with median scores ranging from 3.5 to 4.5 on a 5-point scale. The biofeedback optimization algorithm correlated excellently to the raw physiological measurements ($r = 0.85$, $P < .001$). The intervention was safe and tolerable.

Conclusion.—We developed an app for young migraine sufferers to receive therapist-independent biofeedback. The app underwent a rigorous development process as well as usability and feasibility testing. It is now ready for clinical trials.

Key words: mHealth, smartphone, wearables, headache, adolescent

Abbreviation: mHealth mobile health

(*Headache* 2020;60:889-901)

From the Department of Neuromedicine and Movement Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway (A. Stubberud, E. Tronvik, G. Gravdahl, and M. Linde); National Advisory Unit on Headaches, Department of Neurology and Clinical Neurophysiology, St. Olavs Hospital, Trondheim, Norway (E. Tronvik, G. Gravdahl, and M. Linde); Department of Psychology, NTNU Norwegian University of Science and Technology, Trondheim, Norway (A. Olsen); Department of Physical Medicine and Rehabilitation, St. Olavs Hospital, Trondheim, Norway (A. Olsen).

Address all correspondence to A. Stubberud, Department of Neuromedicine and Movement Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway, email: anker.stubberud@ntnu.no

Accepted for publication January 24, 2020.

INTRODUCTION

Pediatric migraine is highly prevalent and associated with the substantial deterioration of social functioning and mental health.[1,2] There are few viable options for prophylactic medications, with many options having limited efficacy or adverse effects.[3,4] However, behavioral prophylaxis appears to be a valid treatment option for pediatric pain and headache.[5,6] Specifically, biofeedback is one of the most prominent behavioral approaches, and meta-analytical evidence suggests that it is effective in treating pediatric migraine.[7]

Despite being effective, biofeedback has limited population coverage. This is possibly because it is time-consuming and costly with its provision traditionally through specialist clinics. Typically, to be effective, biofeedback treatment requires a trained therapist, as well as specialized equipment measuring surface electromyography, peripheral skin temperature, or heart rate.[8] However, new digital technologies, including wearable sensors and the use of smartphones for medical purposes mobile health (mHealth), provide new possibilities.[9] Recent research suggests that behavioral mHealth interventions for headache are feasible, but development processes and usability testing remain insufficient.[10] Additionally, efficacy measures are uncertain.[11] Currently, there are no biofeedback smartphone applications available specifically targeted at pediatric migraine.[12] To address this, we have recently performed a study showing that wearable sensors are suitable for biofeedback,[13] similar to studies that have validated the use of wearables for other medical purposes.[14,15] Nonetheless, mHealth treatment is entirely dependent on robust development and usability testing to ensure adherence and efficacy.[10,16]

We present a development and usability study aimed at (1) developing a new biofeedback app for adolescents with migraine and evaluating and improving its feasibility and usability; (2) developing and optimizing an algorithm for the multimodal combination of data from selected physiological measures to provide personalized and therapist-independent biofeedback; and (3) developing a sham biofeedback paradigm to be used as a control in efficacy trials.

METHOD

Study Design and Participants.—The study was designed as a prospective open-label iterative and incremental development and usability study at St. Olavs University Hospital in Trondheim, Norway, from September 2017 to June 2018. Ten adolescent migraine sufferers (aged 13-17 years) were recruited from the municipality using social media and the hospital intranet. No statistical power calculation was conducted prior to the study, and the sample size was based on recommendations for usability studies. All diagnoses were confirmed by a consultant neurologist with headache expertise. The participants completed 3 cycles of usability testing with a smartphone biofeedback app. The first two were conducted in a makeshift usability lab, while the final cycle was performed over 14 days

Conflict of Interest: NTNU and St. Olavs Hospital, Trondheim University Hospital may benefit financially from the commercialization of the proposed treatment through future possible intellectual properties. This may include financial benefits to the authors of this article. Dr. Stubberud is a co-founder of the Nordic Brain Tech AS, a spin-off company that was established based on the proposed treatment in this and previous studies at the NTNU Norwegian University of Science and Technology. Dr. Stubberud is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology. Dr. Tronvik is a co-founder of the Nordic Brain Tech AS, a spin-off company that was established based on the proposed treatment in this and previous studies at the NTNU Norwegian University of Science and Technology. Dr. Tronvik is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology. Dr. Olsen is a co-founder of the Nordic Brain Tech AS, a spin-off company that was established based on the proposed treatment in this and previous studies at the NTNU Norwegian University of Science and Technology. Dr. Olsen is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology. Mrs. Gravdahl declares no potential conflicts of interest concerning the research, authorship, or publication of this article. Dr. Linde is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology.

Funding: The study received funding for cooperative projects between the Department of Neuromedicine and Movement Science and Department of Psychology NTNU Norwegian University of Science and Technology.

Registration: The study was approved by the regional committee for medical and health research ethics (2017/582-3) and registered at the Norwegian Centre for Research Data (project number: 54571) prior to enrollment of participants.

at home. After the usability testing, the data collected were used to develop an algorithm to process and combine multimodal physiological data for biofeedback, develop an algorithm for sham-treatment, and finalize the app design to be used and further tested in clinical trials. The study was approved by the regional committee for medical and health research ethics (2017/582-3) and the Norwegian Centre for Research Data (project number: 54571).

Inclusion criteria were age between 12 and 18 years; migraine with or without aura (MWA or MWoA) diagnosed according to the International Classification of Headache Disorders 3;[17] 2 to 6 attacks per month; not using prophylactic migraine medication; experience with using an iPhone® (Apple Inc.); and informed signed consent provided by their guardian. Exclusion criteria were lack of proficiency in the Norwegian language; reduced vision, hearing, or sensibility to a degree that hampered study participation; or if they had any serious neurological or psychiatric disorders.

Biofeedback Setup.—The biofeedback setup consisted of 3 sensors measuring muscle tension, finger temperature, and heart rate, all transmitting signals via Bluetooth® Smart/4.0 to an iPhone® 6 or newer. A small compact bipolar surface electromyography sensor (NeckSensor™; EXPAIN AS, Oslo, Norway) was used for measuring muscle tension from the upper trapezius muscle fibers. A PASPORT Skin/Surface Temperature Thermistor Probe, PS-2131 (Pasco, Roseville, CA, USA) was held between the index finger and thumb of the right hand to measure finger temperature. Finally, a MIO Fuse™ (Mio Global, Physical Enterprises) heart rate wristband was used to measure heart rate over the dorsal aspect of the left wrist.

Usability Evaluation and App Development.—Usability evaluation and biofeedback app development consisted of 3 iterative cycles. Each cycle included the following steps: (1) app programing and design; (2) intervention review by a neurologist, neuropsychologist, computer engineer, and medical student; and (3) usability testing by adolescent migraine sufferers. The initial version of the app-user interface was based on a literature review and evaluation from a previous study that validated wearable sensors as suitable for biofeedback.[13]

The first two usability-testing cycles were completed as one-hour sessions in a consultation room at the hospital. During the first cycle, the participants were initially given an introduction, a description of the rationale of the treatment, and instructions on how to use the app. Subsequently, they were asked to set up the equipment, start the app, and complete a ten-minute biofeedback session. Participants were not trained or instructed in relaxation or stress management techniques. For the second cycle, the participants completed 3 sessions of 5 minutes, with 20 minutes rest time between each session. The final cycle was conducted at home for 2 weeks. The participants were provided with sensors to be used with their personal iPhone®. They downloaded the app from a webpage and were asked to complete a daily biofeedback session of 10 minutes. Following this, they completed a headache diary in the app. After each usability cycle, the participants were asked to complete a comprehensive, structured age-appropriate user evaluation (Supporting Information 1). The user evaluation form was based on commonly structured surveys such as the *Post Study System Usability Questionnaire*, the *System Usability Survey*, and a recently developed mobile app rating scale.[18] The 5 main domains included in the evaluation were (1) engagement; (2) functionality; (3) design; (4) information; and (5) understanding of the biofeedback. The user evaluation also included questions regarding any discomfort they experienced while using the app or sensors and an open-ended adverse events assessment. During the 2 first sessions, 1 of the investigators was present to assist participants with completing the evaluation. Experiences and findings from the intervention review and usability testing from each cycle were used to implement changes to the app for the next iteration of testing. Descriptive analyses of changes to the app interface and development were summarized by a simple thematic analysis categorized under the same 5 domains as the questionnaire.[18]

Biofeedback Algorithm Development.—The biofeedback algorithm was designed to give a compound feedback signal based on all 3 input parameters, that is, muscle tension, finger temperature, and heart rate. To optimize feedback, 2 settings of the algorithm were individually adjusted to each user. First, the default upper and lower measurement limits for the 3 physiologic

parameters were defined based on normalizing graphs of participant data. A factor was then defined as to how the upper and lower limits would be adjusted between each session. Based on the upper and lower individual physiological limits, a 0-100 score for each parameter was created. Second, we defined an internal weighting factor for combining the 3 parameter scores. This was to ensure that a lack of improvement in 1 parameter for a session and absence of a decreasing score would still result in a moderate positive combined score. These variable factors were decided based on the usability evaluation and confirmed as suitable using a regression analysis after the final iteration.

We also developed a set of sham-algorithms by manipulating the raw data. The sham algorithms were visually and statistically analyzed to evaluate if they produced sufficient disruption between the physiological data and feedback, while, importantly, still retaining masking and motivation for the user.

Data Management and Statistics.—The average number of hours of daily smartphone use, general experience with apps, and experience with wearable sensors were averaged over the 3 cycles for each participant. Usability evaluations were scored on a 5-point Likert interval scale, ranging from 1-"completely disagree" to 5-"completely agree." These scores were averaged over each domain for all participants. We used the principle of last observed value carried forward (LOCF) for missing data from dropouts in the usability analyses. We also made an analysis of complete data to serve as a comparison to the imputed data. Baseline feedback score and change in feedback score (ie, the change from the start to the end of a session) for surface electromyographic voltage, skin temperature, and heart rate were registered for all completed sessions. Combined unweighted "raw" scores were created using an equal 33.3% weighting for each of the 3 physiologic parameter scores, while biofeedback algorithm weighted change values were calculated using the above-described biofeedback algorithm. We used only complete data for analyses of physiological measurements without imputing data.

Data were reported as means, standard deviations (SD), medians, and interquartile ranges (IQR). Usability scores were compared between cycles with a two-tailed Wilcoxon signed-rank test and summarized with medians and IQR. We calculated the Pearson correlation coefficient to assess the association between the combined unweighted scores and biofeedback algorithm scores and described the association using a two-tailed linear regression analysis. The regression analysis was applied to evaluate if the biofeedback algorithm would provide a non-random and systematic improvement in feedback scores. All normality assumptions were checked by visual inspection of histograms. P values $<.05$ were considered statistically significant.

This is the primary analysis of data collected in this study. A priori we planned for analyses to compare scores across usability cycles and analyze for correlation between the raw feedback scores and the algorithm scores. Analyses of correlation between familiarity with apps and wearables and usability scores were also planned a priori but were omitted as data were underpowered and not suited for regression analyses.

All statistical analyses were performed and figures were made using Stata v14 (Stata Corp, College Station, TX, USA) and Python v3.6 (Python Software Foundation) with the pandas v0.20.3, NumPy v1.17.2, matplotlib v3.1.1, and scikit-learn v0.21.3 libraries.

RESULTS

Participants and Demographics.—Ten participants with a mean age of 15 ± 1.6 years (range, 13-17 years) were included in the study. Seven were male. One participant did not attend the first cycle. In the second cycle, 2 dropped out, and 1 did not attend. In the final cycle, 2 additional participants dropped out, and 1 had problems with making the setup work properly. Five participants completed all usability cycles and 5 of 10 participants dropped out (50% attrition rate). The average daily previous smartphone usage was 3.7 ± 1.6 hours. The median value familiarity with previous smartphone apps was 4 (good familiarity), with a mean of 4.0 ± 0.8, while the median value familiarity with wearable sensors was 1 (very little familiarity), with a mean of 1.5 ± 1.0. A total of 72 biofeedback sessions were completed throughout the study, with an average per participant of 8.4 in the 2 weeks of the third cycle.

Usability Metrics and App Development.—Figure 1 shows the median and IQR usability scoring for the 5 primary domains of usability assessment

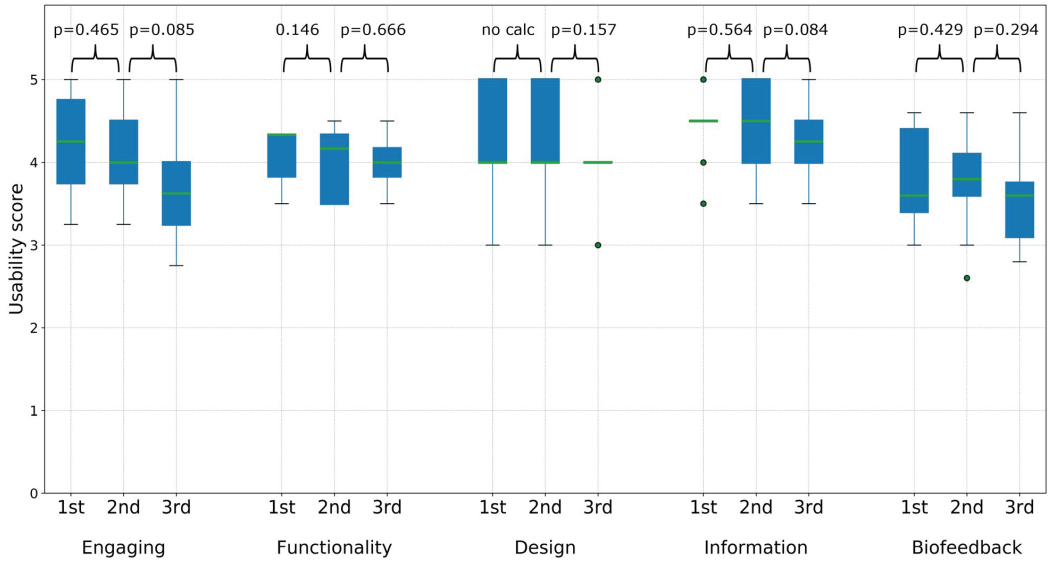


Fig. 1.—Boxplot showing the usability scoring for the 5 main domains through the 3 usability cycles. Green horizontal lines represent medians, blue boxes represent inter quartile ranges (IQR), whiskers represent IQR × 1.5, and green dots represent outliers. Usability scores were compared between cycles with a two-tailed Wilcoxon signed-rank test. No statistically significant changes were found between iterations for each of the 5 domains. Test statistics and *P* value for the domain of design from first to second cycle was not calculable because all ranks were tied. [Color figure can be viewed at wileyonlinelibrary.com]

Table 1.—Thematic Summary of the Most Important App Development and Interface Changes Implemented After Each Cycle. Themes are Classified According to the 5 Main Usability Domains[18] – Biofeedback, Design, Engaging, Functionality, and Information

Cycle 1	Biofeedback – Feedback as separate visualizations for each parameter instead of a combined circle implemented
	Design – Bright light, and especially bright blue colors avoided
	Functionality – Enabled easier navigation and flow between app screens
Cycle 2	Biofeedback – Feedback sometimes perceived as too sensitive was thereby smoothed over a short window
	Engaging – Included reminder function at set timepoint daily
	Information – Provided better information on how to use and connect sensors in the app
Cycle 3	Biofeedback – Algorithm for weighted and individualized feedback optimized
	Design – Finished design with a desirable color palette of dark and green
	Functionality – Included interactive diary that easily allows for viewing previous sessions and headache entries

for each of the 3 iterations. No statistically significant difference was found from the first to the second cycle and from the second to the third cycle for any of the domains (Fig. 1). The complete data analysis without LOCF imputation resulted in lower estimates in the third cycle domains of engaging and biofeedback, with medians (IQR) at 3.3 (2.9-3.6) and 3.2 (3.0-3.5), respectively. Thematic descriptions of the changes in the app interface and app development implemented after each cycle are provided in Table 1. Figures 2 and 3 depict the app interface before the first iteration and after the final iteration.

Biofeedback Algorithm and Sham Development.—The mean value ±2 standard deviations was used to establish the default upper and lower measurement limits for all physiologic parameters. Out of the 251,874 data points for muscle tension measurements, 95% of the values fell within a range of 0.01-0.16 mV.

Fig. 2.—Sample of screenshots from the first version of the app. Screenshot 1 shows the 3 physiological parameters combined as 1 feedback visualization. Screenshot 2 shows the first edition of the headache diary with a subscreen or roll-down pane for each question. [Color figure can be viewed at wileyonlinelibrary.com]

Fig. 3.—Sample of screenshots from the final version of the app. Screenshot 1 shows the easily navigable headache diary. Screenshot 2 shows the feedback as 3 separate feedback-indicators, 1 for each physiologic parameter. Screenshot 3 shows one of the pictures for instructions on connecting sensors. Screenshot 4 shows the headache diary overview. [Color figure can be viewed at wileyonlinelibrary.com]

Similarly, from the 18,572 data points for heart rate measurements, 95% of the values fell within a range of 46 to 90 beats per minute. Finally, out of the 20,734 data points for the finger temperature measurements, 95% of the values fell within a range of 25.3-39.0°C. This supernormal upper-temperature limit is caused by uncertainty in the absolute measurements of the temperature sensor. Therefore, the default upper limit was set to 37.0°C and was allowed to vary around this limit. The default lower temperature limit was set according to the 2.5‰.

Data from 42 completed biofeedback sessions in the third cycle were used to calculate "raw" unweighted scores and biofeedback algorithm scores on a 0-100 scale. The unweighted baseline score was 64.1 ± 10.6 and the end-session unweighted scores were 72.0 ± 10.3. Applying the biofeedback algorithm to the same dataset yielded a baseline session score of 64.3 ± 10.6 and an end-session score of 78.5 ± 10.7. A Pearson's product-moment correlation analysis established a strong positive correlation between the change in unweighted and biofeedback algorithm scores, $r(40) = 0.85$, $P < .001$. The corresponding linear regression established that the unweighted scores accounted for 72% of the variation in the crude biofeedback algorithm scores with the following regression equation: biofeedback algorithm scores = 7.41 + 0.85 × (unweighted score), $F(1, 40)$, $P < .001$. Figure 4 is a scatter plot showing the regression line of fit to visualize the linear correlation and illustrate how the biofeedback algorithm results in an improved feedback score, whereas a sham-algorithm leads to random feedback scores.

Four principal approaches were attempted to develop sham biofeedback. These are described in detail and evaluated in Table 2 and Figures 4 and 5. Sham biofeedback, where the feedback is distorted by a sine-wave fluctuation, was considered the most suitable. This sham was judged to give incorrect feedback, but not to the degree that would promote unmasking.

Safety and Tolerability.—From the evaluation questionnaires, 12 out of the 20 ratings relating to intervention discomfort were "very little discomfort," while the remaining 8 were "little discomfort." Out of the 20 ratings relating to sensor discomfort, 14 were "very little discomfort," 5 were rated as "little discomfort," and 1 was rated as "very great discomfort." No serious adverse events were reported.

DISCUSSION

Principal Findings.—We developed a new mHealth biofeedback intervention for young migraine sufferers that is suitable for self-administration. The intervention includes an algorithm that gives optimized and personalized compound feedback based on 3 physiological parameters proven to be effective in migraine prophylaxis. The intervention was perceived as safe and received consistently high usability scores throughout the 3 cycles of usability testing.

Interpretation.—We developed an app with an algorithm that combines 3 physiological parameters, as opposed to traditional biofeedback where 1 parameter is used.[19] This optimization algorithm was implemented to overcome the challenge that not all biofeedback users experience an influence over the physiological parameter measured,[20] and that different parameters may be useful for different users. For instance, if a user excels at raising their finger temperature, but has trouble lowering their heart rate, the algorithm will fade out the latter throughout the session and thereby chose a more appropriate and "personalized" parameter for the individual. Comparably, the parameter that is most efficient for each user will be given the heaviest weighting in the combined feedback score. This feature was implemented believing that it is likely to result in relevant and useful feedback for a larger group of potential users. Moreover, the intervention did not include commonly used adjuvant therapies such as relaxation training and stress management techniques. This was a deliberate decision made to investigate both if a therapist may be completely excluded from the usual biofeedback treatment "package," and to see if the app itself may to a certain degree replace the therapist. The algorithm was also deemed as suitable after a regression analysis, where the algorithm yielded systematically improved scores, with a significant proportion still attributable to the raw data. This confirmed the desired effect of the biofeedback algorithm to give moderate positive combined feedback despite lack of continuous "improvement" in a physiological parameter. Additionally, the app enables personalized scoring of physiological parameters in an age group that is

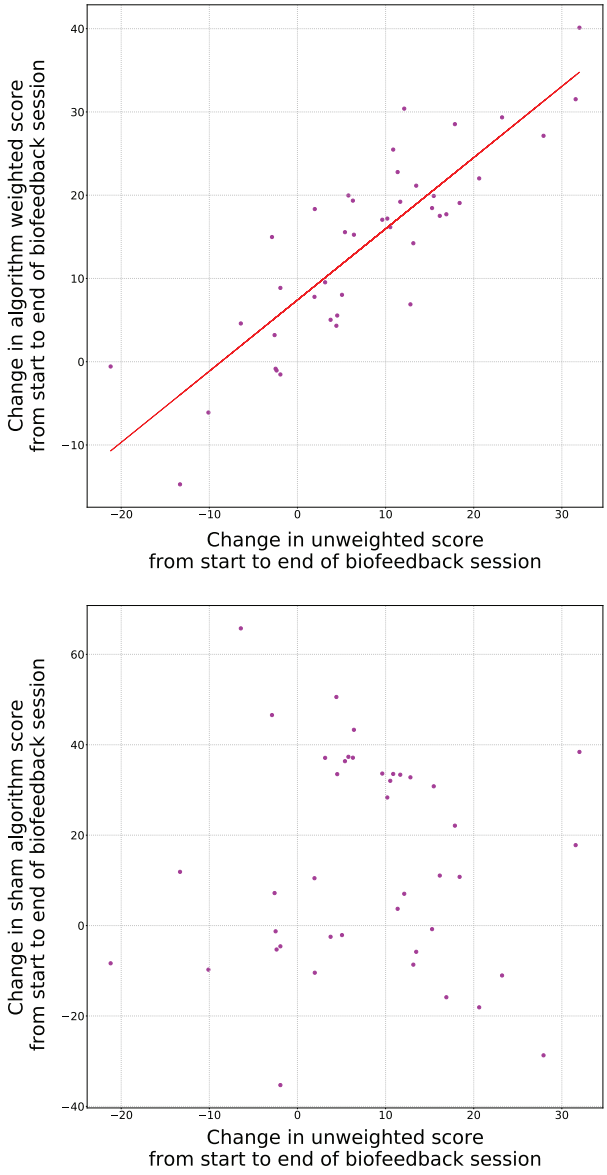


Fig. 4.—Scatterplots comparing "raw" unweighted feedback scores to the biofeedback algorithm scores (upper plot) and sham scores (lower plot) based on 42 biofeedback sessions completed by the 5 participants in the third usability cycle. The values on the axes represent change in feedback score from the start to the end of a biofeedback session (ie, a user's performance during that session). The upper scatterplot shows biofeedback algorithm change values plotted against unweighted change values. The biofeedback algorithm change values are generally higher than the unweighted change values indicating that the biofeedback algorithm improves feedback scores. The predicted improvement in scores is given by the regression equation (red line) for the linear regression model: Biofeedback algorithm score = 7.41 + 0.85 × (unweighted score), $F(1, 40)$, $P < .001$. Together, this illustrates that the biofeedback algorithm scores are improved while still preserving a relationship to the actual "raw" unweighted data. The lower scatterplot shows the inverted sham algorithm change values plotted against unweighted change values. Contrary to the biofeedback algorithm, there is no clear relationship between the sham change values and the unweighted change values. This sham gives a very random feedback and was thus deemed as unsuited because it would likely promote unmasking. [Color figure can be viewed at wileyonlinelibrary.com]

Table 2.—Sham Biofeedback Alternatives

Sham Name	Description	Evaluation
Inverted weighting	Inverting each parameter score and weighting of the 3 physiologic parameters	Applying the inverted weighting algorithm to the raw data yielded a baseline session score of 36.4 ± 12.4, and an end-session score of 50.3 ± 19.4. Moreover, it produced a nearly random feedback score with no clear relationship to the unweighted scores (Fig. 4) and was thus deemed unsuited
Sine-wave fluctuations	Applying a sine-wave fluctuation multiplier of amplitude a to the raw combined data: Sham = $\sin(r \times w \times a)$	The sine-wave fluctuation was evaluated at wavelength (w) 0.1, 0.05, and 0.01π; and at amplitudes (a) 0.05, 0.10, 0.15, and 0.20. The sine-wave fluctuations produced a sham signal deemed to be sufficiently disrupted form the raw data, but still not giving obvious signal deviations in cases such as voluntary contractions and loss of sensor contact. The most suited sine-wave sham version is visualized in Figure 5
Random fluctuations	Applying a pseudo-random fluctuation multiplier of amplitude a to the raw combined data: Sham = $r \times$ random $\{\text{random} \in R \mid 1-a \leq \text{random} \leq 1+a\}$	The pseudo-random fluctuation multiplier was evaluated at frequencies 1, 0.5, 0.33, 0.25, and 0.1 Hz; and at amplitudes (a) 0.10, 0.15, and 0.25. The random fluctuations were evaluated as producing sufficient disruption of the signal, but to a degree that might promote unmasking, and thus deemed unsuitable. The most suited random fluctuation sham is visualized in Figure 5
Full disruption	Providing a feedback signal completely separated from the actual physiological measurements	A full disconnection between the input physiologic data and the feedback visualization, for example, by presenting a completely random feedback, was evaluated as unsuited because it would easily lead to both unmasking and demotivation with the user

known to display great variance in their physiological properties.[21,22] Together, this provides robust therapist-independent treatment.

In addition to the development of the biofeedback algorithm, we rigorously tested and evaluated several sham-treatments. An empirical evaluation of the sham algorithms would have been beneficial to accurately ascertain what type of sham would best perform in a controlled trial. However, this was not prioritized in this current study as our main aim focused on usability. Nevertheless, this paper presents several potential shams of which both random fluctuation and sine-wave fluctuation shams were considered suitable. The inverted weighting and full disruption shams should be avoided because they may promote unmasking of the sham control.

The intervention was considered tolerable and safe. One participant reported "very great discomfort" after using the intervention. This is most likely due to these questions having inverted scoring as compared to the majority of questions in the evaluation. We have previously captured experiences of unpleasantness when removing the electromyography electrodes,[13] but this was not the case in this study. Finally, serious adverse events were not expected and have not been reported in the literature.[7,19] No serious adverse events occurred during our study.

Throughout this study, we aimed to assess and improve the feasibility and usability of the app, which is essential to obtaining satisfactory adherence and an effective treatment.[23,24] Such a rigorous usability approach yields important results that are highly informative for further development, and critical for planning clinical trials. It may be considered as similar to the phase I-II development of new drug treatments.[25] Similar studies carried out within other medical fields have also detected and addressed several issues regarding the feasibility and usability of mHealth interventions. Among these, several[26,27] also used an iterative approach, which is an established usability strategy.[28] Altogether, this highlights the necessity of development and usability studies when creating mHealth interventions. In our study, the usability scorings were consistently high. We evaluated the effect of changes

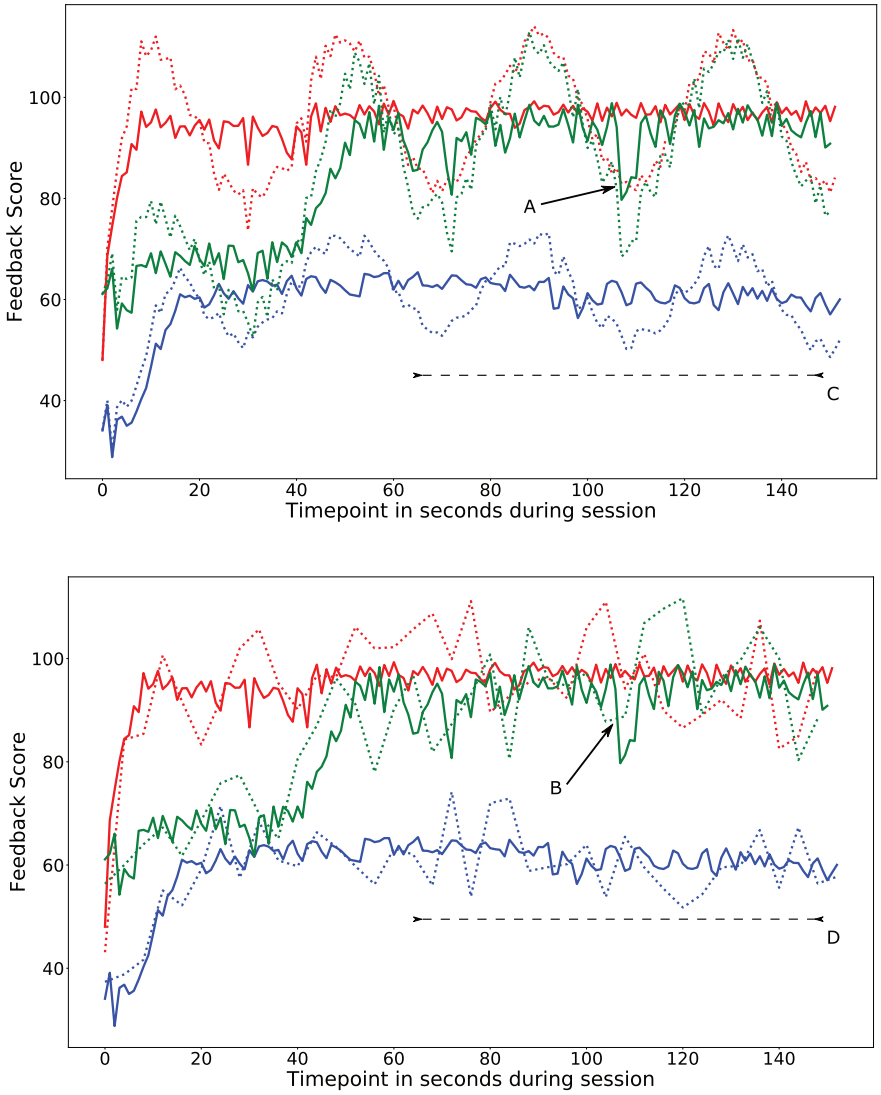


Fig. 5.—Lineplots with the "raw" unweighted feedback scores (solid lines) and sham scores (dotted lines) for 3 representative participants by color. The upper plot shows the sine-wave fluctuation sham scores (dotted lines) with a wavelength of 0.05π and a multiplication amplitude of 0.15. The lower left plot shows the random fluctuation sham scores (dotted lines) with a frequency of 4 Hz and a multiplication amplitude of 0.15. The figures are intended to illustrate how the sham feedbacks are experienced by the user, as compared to the "raw" unweighted feedback scores. Arrow A and B points to a timepoint in the green participant's biofeedback session with a sudden drop in "raw" feedback scores as may occur upon shrugging the shoulders (sudden increase in electromyographic voltage) or losing contact with the finger heart rate sensor (sudden fall in heart rate). In the upper plot, arrow A points to a corresponding decrease in the sine-wave fluctuation sham score. In the lower plot, arrow B points to a moment where the random sham results in a sudden increase in feedback scores, despite an obvious drop in the "raw" feedback score. Such randomness might promote unmasking, thus making a random fluctuation sham less suited. Moreover, the dotted horizontal lines C and D represent a time period where the "raw" feedback score for the red and blue participant is relatively stable. During this time period, the sine-wave fluctuations gives incorrect feedback which is slow and smooth, whereas the random fluctuation gives sharp and sudden changes in feedback score further promoting unmasking. [Color figure can be viewed at wileyonlinelibrary.com]

in each iteration on the usability scores, but no systematic statistically significant differences between cycles were found. This may be explained by both the original high scores and the sample size not providing sufficient power to detect a difference. On the contrary, the high attrition rate should also be considered as a measurement of usability, and dropping out of the study may simply be the result of a participant not enjoying the app. Likewise, missing data as a result of attrition certainly impacts interpretation of usability scoring. If all dropouts were in fact not liking the app, the overall usability scoring would have been poorer. In addition to the quantification of the usability, the participants were asked several open-ended questions and interviewed during the evaluation. Their comments provided valuable qualitative input for a thematic analysis on how the app could be improved. The app-user interface and usability were qualitatively improved with each iteration even though this was not evident in the usability scoring. We believe this resulted in a final solution that is more likely to meet the desired needs of a larger group of users as compared to an undocumented product directly being implemented in clinical efficacy trials.

Limitations and Strengths.—Several factors limit this study and make us reluctant to draw firm conclusions. First, the questionnaire was not validated for our specific study but rather based on common usability surveys and validated mHealth questionnaires.[29] Such questionnaires can be susceptible to response bias,[30] including acquiescence bias, in which participants automatically endorse statements to please the interviewer.[31] This may explain the high usability scorings in the initial cycle and the lower scores in home testing. Second, the first two usability cycles were conducted in a controlled environment, not fully representative of the intended use. In addition, the home testing session was conducted over a shorter than recommended period.[32,33] Together, this adds some uncertainty concerning the adherence to the intervention. Third, the study had a moderate sample size and suffered from attrition. This may represent poor usability and decrease the confidence in our findings. Nonetheless, the sample size was chosen according to recommendations for usability studies. Some researchers even argue that a sample size of approximately 5 participants is sufficient to uncover the majority of usability problems,[34,35] while others argue that such a small sample size is insufficient and that sample sizes should be customized to individual studies.[36] We ultimately chose a sample size of 10 people stratified across the adolescent age range to ensure essential usability problems were uncovered, while also receiving an evaluation from the whole heterogeneous age spectrum.

This is the first study of adolescent migraine that uses mHealth to deliver migraine therapy and enables biofeedback treatment to be provided to a broader population. The optimizing algorithm included in the intervention makes it superior to traditional monitoring that requires a trained therapist for interpretation. This will, in turn, lower costs and increase availability. Moreover, the intervention was developed by a multidisciplinary team, including neurologists with headache expertise, a neurophysiologist, and software engineers based on the guidelines for developing mHealth apps[37] and guidelines for behavioral treatment trials.[33] It used sensors that have previously been validated as appropriate for biofeedback and we involved the target group throughout the whole development process. These factors all helped to improve the final product.[23,38] By using the same set of participants for all cycles of usability testing, we also overcame the challenge that biofeedback as a psychophysiological training method requires several rounds of exposure to master.[8] This also allowed us to complete a large number of biofeedback sessions and repeated usability tests on the same individuals. Altogether, we believe that this new intervention has the potential to be effective and reach a broader population in need.

CONCLUSION

In this study, we developed a new biofeedback treatment app targeted at young migraine sufferers. The treatment includes wearable sensors, validated as appropriate for biofeedback, and a feasible and usable app developed specifically for the target population. Some study findings were limited by the low sample size, attrition, and response bias. Future studies should determine whether the migraine intervention developed in this study has a clinical effect on the migraine burden in adolescents.

Acknowledgments: The authors thank all participants for taking part in the study.

STATEMENT OF AUTHORSHIP

Category 1

(a) Conception and Design
Anker Stubberud, Erling Tronvik, Alexander Olsen, Mattias Linde

(b) Acquisition of Data
Anker Stubberud, Erling Tronvik, Gøril Gravdahl, Mattias Linde

(c) Analysis and Interpretation of Data
Anker Stubberud, Erling Tronvik, Alexander Olsen, Mattias Linde

Category 2

(a) Drafting the Manuscript
Anker Stubberud

(b) Revising It for Intellectual Content
Anker Stubberud, Erling Tronvik, Alexander Olsen, Gøril Gravdahl, Mattias Linde

Category 3

(a) Final Approval of the Completed Manuscript
Anker Stubberud, Erling Tronvik, Alexander Olsen, Gøril Gravdahl, Mattias Linde

REFERENCES

1. Wober-Bingol C. Epidemiology of migraine and headache in children and adolescents. *Curr Pain Headache Rep*. 2013;17:341.
2. Krogh AB, Larsson B, Linde M. Prevalence and disability of headache among Norwegian adolescents: A cross-sectional school-based study. *Cephalalgia*. 2015;35:1181-1191.
3. El-Chammas K, Keyes J, Thompson N, Vijayakumar J, Becher D, Jackson JL. Pharmacologic treatment of pediatric headaches: A meta-analysis. *JAMA Pediatr*. 2013;167:250-258.
4. Powers SW, Coffey CS, Chamberlin LA, et al. Trial of amitriptyline, topiramate, and placebo for pediatric migraine. *N Engl J Med*. 2017;376:115-124.
5. Fisher E, Law E, Dudeney J, et al. Psychological therapies for the management of chronic and recurrent pain in children and adolescents. *Cochrane Database Syst Rev*. 2018;9:CD003968.
6. Trautmann E, Lackschewitz H, Kroner-Herwig B. Psychological treatment of recurrent headache in children and adolescents – A meta-analysis. *Cephalalgia*. 2006;26:1411-1426.
7. Stubberud A, Varkey E, McCrory DC, Pedersen SA, Linde M. Biofeedback as prophylaxis for pediatric migraine: A meta-analysis. *Pediatrics*. 2016;138:e20160675.
8. Schwartz MS, Andrasik F. *Biofeedback: A Practitioner's Guide*. New York: Guilford Publications; 2017.
9. Stubberud A, Linde M. Digital technology and mobile health in behavioral migraine therapy: A narrative review. *Curr Pain Headache Rep*. 2018;22:66.
10. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There'sa pain app for that": Review of patient-targeted smartphone applications for pain management. *Clin J Pain*. 2015;31:557-563.
11. Minen MT, Torous J, Raynowska J, et al. Electronic behavioral interventions for headache: A systematic review. *J Headache Pain*. 2016;17:51.
12. Mosadeghi-Nik M, Askari MS, Fatehi F. Mobile health (mHealth) for headache disorders: A review of the evidence base. *J Telemed Telecare*. 2016;22: 472-477.
13. Stubberud A, Omland PM, Tronvik E, Olsen A, Sand T, Linde M. Wireless surface electromyography and skin temperature sensors for biofeedback treatment of headache: Validation study with stationary control equipment. *JMIR Biomed Eng*. 2018;3:e1.
14. Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med Sci Sports Exerc*. 2017;49:1697-1703.
15. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol*. 2017;2:104-106.
16. de la Vega R, Miró J. mHealth: A strategic field without a solid scientific soul. A systematic review of pain-related apps. *PLoS ONE*. 2014;9:e101312.
17. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition. *Cephalalgia*. 2018;38:1-211.
18. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR mHealth uHealth*. 2015;3:e27.
19. Nestoriuc Y, Martin A. Efficacy of biofeedback for migraine: A meta-analysis. *Pain*. 2007;128:111-127.
20. Rains JC. Change mechanisms in EMG biofeedback training: Cognitive changes underlying

improvements in tension headache. *Headache*. 2008; 48:735-736.
21. Fleming S, Thompson M, Stevens R, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. *Lancet*. 2011;377:1011-1018.
22. Semizel E, Öztürk B, Bostan OM, Cil E, Ediz B. The effect of age and gender on the electrocardiogram in children. *Cardiol Young*. 2008;18:26-40.
23. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: A systematic literature review. *J Med Syst*. 2015;39:1.
24. Zapata BC, Fernández-Alemán JL, Toval A, Idri A. Reusable software usability specifications for mHealth applications. *J Med Syst*. 2018;42:45.
25. Schmidt B. Proof of principle studies. *Epilepsy Res*. 2006;68:48-52.
26. Huguet A, McGrath PJ, Wheaton M, et al. Testing the feasibility and psychometric properties of a mobile diary (myWHI) in adolescents and young adults with headaches. *JMIR mHealth uHealth*. 2015;3:e39.
27. Hughes CM, Hintze A, Padilla A, et al. Development of a mHealth system for post-stroke upper limb rehabilitation in medically underserved populations: An iterative usability study. In: *2018 IEEE Global Humanitarian Technology Conference (GHTC)*: IEEE; 2018:1-8.
28. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: Mixed-methods approach. *JMIR mHealth uHealth*. 2019;7:e11656.
29. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): Development and validation study. *JMIR mHealth uHealth*. 2019;7:e11500.
30. Furnham A. Response bias, social desirability and dissimulation. *Pers Individ Dif*. 1986;7:385-400.
31. Knowles ES, Nathan KT. Acquiescent responding in self-reports: Cognitive style or social concern? *J Res Pers*. 1997;31:293-301.
32. Tfelt-Hansen P, Pascual J, Ramadan N, et al. Guidelines for controlled trials of drugs in migraine: third edition. A guide for investigators. *Cephalalgia*. 2012;32:6-38.
33. Penzien DB, Andrasik F, Freidenberg BM, et al. Guidelines for trials of behavioral treatments for recurrent headache, first edition: American Headache Society Behavioral Clinical Trials Workgroup. *Headache*. 2005;45(Suppl. 2):S110-S132.
34. Schmettow M. Sample size in usability studies. *Commun ACM*. 2012;55:64-70.
35. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. In: *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*: ACM; 1993:206-213
36. Hwang W, Salvendy G. Number of people required for usability evaluation: The 10±2 rule. *Commun ACM*. 2010;53:130-133.
37. Chatzipavlou IA, Christoforidou SA, Vlachopoulou M. A recommended guideline for the development of mHealth apps. *mHealth*. 2016;2:21.
38. Shneiderman B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. London: Pearson Education India; 2010.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

Manuscript IV

Self-Administered Biofeedback Treatment App for Pediatric Migraine: A Randomized Pilot Study

Stubberud A, MD[1]; Linde M, MD, PhD[1,2]; Brenner E, MD, PhD[1,2]; Heier M, MD, PhD[3]; Olsen A, PhD[4,5]; Anne Hege Aamodt, MD, PhD[6]; Gravdahl GB, MSc[1,2]; Tronvik E, MD, PhD[1,2]

[1] Department of Neuromedicine and Movement Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway

[2] Norwegian Advisory Unit on Headaches, Department of Neurology, St. Olavs Hospital, Trondheim, Norway

[3] Department of Clinical Neuroscience for Children, Oslo University Hospital, Oslo, Norway

[4] Department of Psychology, NTNU Norwegian University of Science and Technology, Trondheim, Norway

[5] Department of Physical Medicine and Rehabilitation, St. Olavs Hospital, Trondheim, Norway

[6] Department of Neurology, Oslo University Hospital, Oslo, Norway

Prepared for: *Headache*

Abstract word count: 305

Word count: 3748

Figures: 2

Tables: 3

References: 35

Supplementary material: None

Corresponding author:

Anker Stubberud

Department of Neuromedicine and Movement Science, NTNU Norwegian University of Science and Technology, Norway

Tel.: +47 45229174

Email: anker.stubberud@ntnu.no

Conflicts of interest

NTNU Norwegian University of Science and Technology and St. Olavs Hospital, Trondheim University Hospital may benefit financially from the commercialization of the proposed treatment through future possible intellectual properties. This may include financial benefits to the authors of this article.

Dr. Stubberud is a co-founder of the Nordic Brain Tech AS, a spin-off company that was established based on the proposed treatment in this and previous studies at the NTNU Norwegian University of Science and Technology. Dr. Stubberud is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology.

Dr. Linde is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology.

Dr. Brenner declares no potential conflicts of interest concerning the research, authorship, or publication of this article.

Dr. Heier declares no potential conflicts of interest concerning the research, authorship, or publication of this article.

Dr. Olsen is a co-founder of the Nordic Brain Tech AS, a spin-off company that was established based on the proposed treatment in this and previous studies at the NTNU Norwegian University of Science and Technology. Dr. Olsen is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology.

Dr. Aamodt declares no potential conflicts of interest concerning the research, authorship, or publication of this article.

Mrs. Gravdahl declares no potential conflicts of interest concerning the research, authorship, or publication of this article.

Dr. Tronvik is a co-founder of the Nordic Brain Tech AS, a spin-off company that was established based on the proposed treatment in this and previous studies at the NTNU Norwegian University of Science and Technology. Dr. Tronvik is a co-inventor of the proposed treatment in this study and may benefit financially from a license agreement between Nordic Brain Tech AS and NTNU Norwegian University of Science and Technology.

Keywords

mHealth, smartphone, wearables, headache, adolescent

Funding

The study received funding for cooperative projects between the Department of Neuromedicine and Movement Science and Department of Psychology NTNU Norwegian University of Science and Technology.

Abbreviations

AE – adverse events

CI – confidence interval

IQR – interquartile range

ITT – intention to treat

MD – mean difference

mHealth – mobile health

mITT – modified intention to treat

SEMG – surface electromyography

SD - standard deviation

Acknowledgements

The authors thank all study participants.

Abstract

Objective. To investigate the effect size, safety, and tolerability of a biofeedback treatment app among adolescent migraine sufferers.

Background. Biofeedback is an effective treatment for pediatric migraine but is under-utilized because it is typically time- and cost-consuming. The emergence of smartphone apps and wearables has previously led to the development of a therapist-independent biofeedback app.

Methods. This was a prospective, 3:1 ratio randomized, sham-controlled, double-blind, pilot study with 16 adolescent migraine sufferers randomized to eight weeks of biofeedback treatment (n=12) or sham-biofeedback (n=4). The pre-specified and primary objective of the study was to observe changes in outcomes within the active treatment group. The sham control group was included in a minor ratio primarily to evaluate its feasibility. The primary outcome was change in headache frequency from baseline to end of treatment (weeks 5-8). Secondary outcomes included response rate, headache intensity, daily functioning, abortive drug consumption, and adverse events. A modified intention to treat analysis was performed, including participants completing at least seven biofeedback sessions in weeks 1-4 and weeks 5-8.

Results. Adherence was poor with 40% (±36%) of planned biofeedback sessions completed during weeks 5-8. Within the biofeedback group, a not statistically significant reduction in headache frequency was observed at weeks 1-4 (2.92 days/month, 95% CI -1.00 to 6.84, p=0.145) and weeks 5-8 (1.85 days/month, 95% CI -2.01 to 5.72, p=0.395). The biofeedback group experienced a median of one fewer headache days/month vs. sham that did not reach significance (95% CI -4.0 to 9.0, p=0.83). The only adverse event observed was a case of mild skin rash.

Conclusions. We observed a small reduction in headache frequency in the active treatment group that was not statistically significant. Findings were likely undermined by low adherence and underpowered analyses but indicate that a therapist-independent biofeedback treatment app has potential to be an effective, tolerable and inexpensive treatment option.

Registration:

Clinicaltrials.gov identifier: NCT04106505

Ethics approval: Regional ethics committee REK Midt 2018/35

Introduction

Pediatric migraine is highly prevalent and associated with substantial deterioration of social functioning and mental health.[1, 2] Those in need of prophylactic treatment are faced with few viable options as most pharmacological prophylaxes have limited efficacy or unacceptable adverse effects.[3-6] However, behavioral therapies, and especially biofeedback, appears to be a suitable treatment option for children and adolescents with headache.[7-9]

During biofeedback, individuals learn to voluntarily modify their bodily reactions through feedback from their own physiological processes. Commonly used physiological parameters are peripheral skin temperature, frontal or trapezius muscle surface electromyographic voltage (SEMG) and blood-volume-pulse.[10] Traditionally, biofeedback is delivered in a clinic with suited measurement devices and a trained therapist. The therapist assists with the technical use of the measurement devices and provides the user with insights on how to interpret and modify the physiological parameters. Regular biofeedback training reduces central nervous system arousal, render individuals more resilient to environmental stressors, and ultimately lower migraine burden.[11, 12] Unfortunately, the time-consuming and cumbersome nature of the treatment has resulted in limited population coverage.[13]

The rapidly growing use of wearables and smartphone mobile applications (apps) for medical purposes (mHealth) allows for simpler ways of administering biofeedback.[14] mHealth poses many potential areas of application in headache medicine, but most of these remains to be explored.[15] Specifically, no app-based biofeedback as prophylaxis for migraine in children and adolescent exists.[16, 17] To start filling this gap of knowledge we have validated the use of wearables suited for biofeedback and

developed a self-administered therapist-independent biofeedback treatment app for pediatric migraine sufferers.[18, 19]

Here we aimed to conduct a pilot study with a primary objective to investigate the effect size, safety, and tolerability of a biofeedback treatment app among adolescent migraine sufferers. Secondly, we aimed to evaluate the feasibility of a sham-biofeedback app and compare it to the active treatment. The study was intended to guide study design, choice of control group, and sample size calculation for future clinical trials.

Methods

Study design and participants

The study was designed as a prospective, 3:1 ratio randomized, sham-controlled, double-blind, pilot study conducted at St. Olavs Hospital, Trondheim, Norway; and Oslo University Hospital, Oslo, Norway, with planned enrollment from January 2019 to June 2020. The study comprised a four-week baseline period, followed by an eight-week intervention period with either a biofeedback treatment app or a sham biofeedback app. We planned on recruiting 40 participants—to ensure at least 25 in the main intervention group—as this represents a number where further increase in precision with increased sample size is minimal.[20] However, recruitment proceeded unexpectedly slow and was terminated prematurely in March 2020 due to the SARS-CoV2 pandemic. Thus, 23 adolescent migraine sufferers were recruited through repeated advertisements at pediatric clinics in the municipality, local mainstream media, social media patient groups, and the intranet at the university hospital in Trondheim. The study was approved by the regional ethics committee (Identifier:

2018/35) and the Norwegian Medicines Agency (Identifier: 18/12060-9). The study was registered at clinicaltrials.gov (Identifier: NCT04106505). Written informed consent was obtained from all patients and their guardians.

Inclusion criteria were (A) age between 12 and 18 years; (B) diagnosis of migraine with or without aura according to the international classification of headache disorders (ICHD-3);[21] and (C) two to eight migraine attacks per month. Exclusion criteria were (A) participant not speaking Norwegian; (B) reduced sensibility, hearing or vision to a degree that impairs proper use of the app; (C) severe psychiatric or neurologic disease and; (D) participant currently using migraine prophylaxis.

Eligible participants met with a consultant neurologist or pediatrician with headache expertise to confirm the migraine diagnosis. During baseline, participants were instructed to daily register maximal headache intensity, average headache intensity, functioning in daily activities, and abortive drug consumption in a paper headache diary. After a minimum 28-day baseline period participants were randomly assigned to one of the two intervention groups by a computer-generated block-randomization list. In each block of four, participants had a 75% chance of being allocated to the biofeedback group and a 25% chance of being allocated to the sham group. Participants were asked to download the app and enter a 5-digit number to unlock the app. The 5-digit number was drawn by the enrolling physician sequentially from a list of 40 numbers. One random in every four numbers resulted in downloading a sham-version of the app while the other three numbers resulted in downloading the proper biofeedback app. Both versions of the app looked alike and no pattern in the 5-digit number or the randomization list could reveal which version of the app was given. This ensured blinding of participants, healthcare providers and investigators. Blinding of outcome assessors was not possible due to the 3:1 randomization ratio. Breaking of

the randomization was performed only after follow-up of the last participant, when the software developers revealed if the 5-digit number corresponded to the biofeedback or sham version of the app.

During treatment, participants were asked to complete daily headache diary entries (the same questions as in the paper diary) and biofeedback sessions within the app. Participants were also encouraged to contact investigators with inquiries on how to use the equipment, report errors or shortcomings regarding both hardware and software, and take notes of any adverse events (AE) and report these to the researchers. Finally, participants met with one of the researchers at the end of the two-month intervention period for evaluation, adverse event questioning, and to return the equipment.

Interventions

The active treatment arm comprised a self-administered treatment app, including biofeedback training, instructions for self-delivery, and a headache diary. The app gave a push-reminder to complete a headache diary entry and a biofeedback session of 10 minutes duration daily. The headache diary entry had to be completed to start a biofeedback session. Prior to commencing treatment, participants were given basic information on the rationale behind biofeedback treatment. They were also given instructions on how to use the equipment and software, and how to complete a biofeedback session. Sham biofeedback was achieved by adding sine-curve fluctuations to the correct feedback signal and thereby partly disrupting the true connection between the input of physiological parameters and the feedback. The looks and contents of the normal app and the sham app were completely similar. The only difference was the internal software algorithm, which was inaccessible to the

user and investigators. All participants in both groups were given the same information and instructions. Participants were not instructed in relaxation techniques or stress management techniques. The intervention and sham are described in detail elsewhere.[19]

The biofeedback source signal was produced by wireless wearable sensors measuring muscle tension, finger temperature, and heart rate. The bipolar surface electromyography sensor (NeckSensor™; EXPAIN AS, Oslo, Norway) was used for measuring SEMG muscle tension from the upper trapezius muscle fibers. The PASPORT Skin/Surface Temperature Thermistor Probe, PS-2131 (Pasco, Roseville, CA, USA) was held between the index finger and thumb of the right hand to measure finger temperature. The MIO Fuse™ (Mio Global, Physical Enterprises) photoplethysmography heart rate wristband was used to measure heart rate over the dorsal aspect of the left wrist. All sensors transmitted signals via Bluetooth® Smart/4.0 to an iPhone® 6 or newer.

Outcomes

The primary outcome was change in the frequency of headache days from baseline to end of treatment. Secondary outcomes were responder rate (more than 50% reduction in headache frequency); change in maximal and average pain intensity recorded on a ordinal 4-point scale (0=no headache, 3=severe headache); change in functioning in daily activities recorded on a ordinal 4-point scale (0=no problems with daily activities, 3=severe problems with daily activities); change in number of days with abortive drug consumption; and AEs. Participants were asked specifically to report any skin reactions, nausea and dizziness, and any additional AEs were recorded.

Headache-related functioning in daily activities and average pain intensity was not pre-specified in the protocol and was included in the headache diary prior to enrollment as per trial guideline recommendations.[22] While the pre-specified and primary objective of this pilot study was to observe the outcomes *within* the biofeedback group only, we also conducted post-hoc comparative analyses of outcomes *between* the two groups. We also conducted a second post-hoc response rate analysis, changing the response threshold to 30% or greater reduction in headache frequency. Finally, we included a post-hoc analysis of mean change in biofeedback physiological parameters from the start to the end of sessions.

Data management and statistical analyses

This is the first analysis of data collected in this study. The analysis was conducted after all patients completed the final visit or terminated participation. At all visits, data was collected and recorded on a paper clinical report form. Paper headache diaries were collected at the end of the baseline period. Baseline headache data was calculated from the last 28 days of the baseline period. The SEMG, temperature, and heart rate measurements for each biofeedback session, along with headache diary data was transferred daily to a secure database. A priori we planned to conduct an intention to treat (ITT) analysis of all randomized patients comparing baseline data to the last 28 days (weeks 5-8) of treatment. However, because several participants did not complete any biofeedback sessions during weeks 5-8 (and thus did not receive treatment and had no headache diary entries) and to avoid imputing data, we conducted a modified ITT (mITT) analysis. To be included in the mITT analysis participants were required to have completed at least 7 of the planned 28 headache diary entries in weeks 5-8. Because all participants completed at least seven biofeedback sessions and headache diary entries during weeks 1-4, we also included

an analysis comparing baseline to weeks 1-4. We used only available data in the analyses with no imputation of data.

Adherence was evaluated as the proportion of completed treatment sessions and headache diary entries (out of 56 planned sessions in the eight weeks following treatment start). The mean SEMG, temperature, and heart rate measurements from the first and last minute of sessions lasting more than five minutes were summarized. We also calculated the median of the ten largest values, the median of the ten smallest values, and the overall mean for the SEMG, temperature, and heart rate recordings from each biofeedback session. The latter data was visualized by plotting the average value across all individuals for each completed session with a moving average smoothing function with a window width of three sessions.

Data was reported as means, standard deviations (SD), medians, and interquartile ranges (IQR). Within-group changes were analyzed with a two-tailed Wilcoxon signed-rank test and summarized with mean differences (MD) with 95% confidence intervals (CI). A two-tailed Mann-Whitney U test was used to compare changes in outcomes between the two groups and median effect estimates with 95% CI were produced with the Hodges-Lehman estimator. Finally, to analyze for changes in the physiological measurements between the start and end of biofeedback sessions we performed a two-tailed paired t-test and summarized the findings using MDs with 95% CI. Normality assumptions were based on visual inspection of histograms. P-values were evaluated at the 0.05 significance level.

All statistical analyses and figures were made with Python (v.3.7.7, Python Software Foundation) with the following open-source packages: matplotlib v.3.2.1, NumPy v.1.18.2, pandas v.0.20.3, PyNonpar v.0.2.0, scipy v.1.4.1, and seaborn v.0.10.0.

Results

Twenty-three participants were recruited, 18 from St. Olavs University Hospital and five from Oslo University Hospital. Seven participants were excluded or dropped out during the baseline period, and 16 patients were randomized (Figure 1). Twelve participants were randomized to the biofeedback group and four were randomized to the sham group. All randomized participants were analyzed at weeks 1-4. Seven participants in the biofeedback group and two in the sham group were analyzed at weeks 5-8. Participant demographics are provided in Table 1. The proportions of completed biofeedback sessions in the biofeedback group were 0.58±0.29 during weeks 1-4 and 0.40±0.36 during weeks 5-8. The proportions of completed biofeedback sessions in the sham group were 0.65±0.32 during weeks 1-4 and 0.30±0.33 during weeks 5-8. Three out of four participants allocated to the sham group believed they received sham treatment, whereas one of the participants in the biofeedback group believed they received sham treatment.

Outcomes in the biofeedback group

A not statistically significant mean reduction in headache frequency of 2.92 days/month (95% CI -1.00 to 6.84, p=0.145) was reported during weeks 1-4. A not statistically significant mean reduction in headache frequency of 1.85 days/month (95% CI -2.01 to 5.72, p=0.395) was reported during weeks 5-8. No statistically significant changes in maximal headache intensity, average headache intensity, headache related daily functioning, or abortive drug consumption were observed within the biofeedback group (Table 2). In the biofeedback group, 4 out of 12 (33.3%) participants were considered responders at weeks 1-4, and 2 out of 7 (28,6%) participants were considered responders at weeks 5-8. Moreover, 9 out of 12 (75%)

participants experienced ≥30% reduction in headache frequency during weeks 1-4 and 2/7 (28,6%) experienced a ≥30% reduction in headache frequency during weeks 5-8.

Between group comparisons

No statistically significant difference in change in headache frequency between the two groups was reported during weeks 1-4 (0.5 headache days/month, 95% CI -9.0 to 16.0, *P*=1.0), and weeks 5-8 (-1.0 headache days/month, 95% CI -9.0 to 4.0, *P*=0.760). There was no statistically significant difference between the two groups in any of the secondary outcomes (Table 3).

Physiological measurements

Table 4 summarizes the physiological measurements at the biofeedback session start and session end in the biofeedback group. Within sessions, participants achieved a statistically significant increase in finger temperature (4.43° Celsius; 95% CI 4.02 to 4.84; p<0.001), increase in heart rate (5.63 beats per minute; 95% CI 3.26 to 8.01; p<0.001), and reduction in SEMG voltage (15.11 millivolts; 95% CI 6.56 to 23.68; p=0.0006). Across all sessions, we observed a slightly increasing trend in maximum finger temperature, and a slightly decreasing trend in minimum heart rate and maximum muscle tension. Figure 2 visualizes the SEMG, temperature and heart rate measurements across all sessions in the biofeedback group.

Safety and tolerability

One single AE was reported by a participant experiencing a mild skin rash related to the SEMG electrode patch. The rash lasted for a week without treatment. None of the other pre-specified AEs were reported.

Discussion

Principal findings

To the best of our knowledge, this is the first trial investigating the use of a mHealth biofeedback intervention designed specifically for migraine in adolescents. Overall, the study suffered from attrition, difficulties in the recruitment process and prematurely terminated data collection due to the SARS-CoV-2 pandemic. No statistically significant reduction in headache frequency in the active treatment group or superiority over sham was observed. Still, several patients experienced a meaningful reduction in headache frequency, and the intervention was nearly free of AEs. The findings should be used as guidance in planning and designing future studies of therapist-independent app-based biofeedback treatment.

Interpretation

Meta-analyses have found that biofeedback is effective in treating pediatric migraine, at least when compared to a waiting list control.[8, 9] Treatment effect is typically in the range of 35-50% reduction in headache frequency.[23] In this study, we observed an approximate 20% reduction in headache frequency, which is lower than the typical treatment effect. Several factors may contribute to understanding why we observed a limited treatment effect that was not statistically significant.

Firstly, the nature of the biofeedback intervention used in the present study was quite different from traditional biofeedback. Usually, the treatment is administered as a "treatment package" with regular therapist-contact sessions and combined with adjunctive behavioral therapies such as relaxation and stress management. The therapist aids the user in achieving the "correct" self-control, and the treatment package promotes several of the non-specific effects seen with biofeedback, such as

expectancy, conditioning, and regular contact and procedural repetitions.[24] In the present study, participants were given a very minimalistic intervention, only consisting of a brief introduction to the concept of biofeedback and brief instructions on how to use the equipment and perform a session. Thereafter, learning self-control was entirely based on operant conditioning from the feedback instruments. Participants appeared to quickly learn to increase temperature and lower muscle tension within biofeedback sessions. However, there was no clearly evident improvement across sessions, and we also observed a paradoxical increase in heart rate within sessions. A real-world therapist could potentially have helped to modulate the self-control towards the assumed "correct" state, which is hypothesized to predict positive outcomes.[25] Moreover, the absence of therapist contact and adjunctive therapies may have led a reduction in the non-specific effects, further explaining the limited treatment effect.[24] Even though previous studies have found that limited-contact biofeedback may be as efficacious as traditional biofeedback,[26-28] these still employed much more comprehensive treatment packages than was used in the present study. On the other hand, a more similar study, investigating the effect of one single biofeedback training session, followed by self-directed practice sessions observed a reduction in headache frequency from 12.9 to 9.7 days/months, which is more in line with our findings.[29]

Secondly, the adherence rate to biofeedback treatment in the present study was low, potentially resulting in reduced treatment effects. A systematic review found that the adherence to behavioral interventions among children varied between 52% to 86%.[30] This is superior to what we observed, especially in weeks 5-8. There are no clear estimates of how much adherence influences treatment outcome, but lower adherence is believed to undermine the efficacy of behavioral interventions.[31] A study of app-

based progressive muscle relaxation as a prophylactic treatment for migraine in adults found that highly adherent users (defined as two or more session per week) had a significantly greater reduction in headache frequency than users with low adherence.[32] This supports our findings, where the reduction in headache frequency in the biofeedback group was greatest in weeks 1-4, the period where adherence was the highest.

Thirdly, the limited data in the study likely means that there was insufficient power to detect a statistically significant change in headache frequency. A priori we planned to recruit 40 participants, to ensure at least 25 in the biofeedback group. This is twice the number that was allocated to biofeedback treatment, and a larger sample size may indeed have revealed a statistically significant reduction in headache frequency. Still, it is unlikely that the pre-specified sample size would have had the power to detect a difference between the active treatment and sham.

Finally, issues with the use of sham-control and identification of therapeutic gains in studies of biofeedback are important to discuss. Studies have found that the biofeedback per se does not necessarily influence treatment effect,[33] in line with the notion that headache improvement by biofeedback is mainly driven by non-specific effects.[24] It has even been shown that instrumental conditioning in the opposite direction than what is hypothesized to lead to headache improvement—i.e. hand-cooling rather than hand-warming—produces similar treatment effects.[28] The sham group in our study experienced a reduction in headache frequency, suggesting that the improvement in all clinical outcomes is caused by placebo and regression to the mean, and supporting the notion that there is no significant therapeutic gain.[34] Still, the choice to conduct the study as a randomized sham-controlled trial was mainly to evaluate the suitability and feasibility of such a sham. The fact that the sham was only

a partial disruption of the biofeedback signal and that the adherence to sham in the first four weeks of treatment was high suggests that the sham signal may be "too similar" to true biofeedback, thus producing a treatment effect. In addition, participants in the biofeedback group had an idea that they might be receiving sham, which together with the "similar" sham may explain the small difference in treatment effects observed between the two groups.

Even though this study failed to demonstrate a convincing treatment effect of app-based biofeedback treatment we believe there is a rationale for continued research. Firstly, the mobile setup and self-administration allow for widespread biofeedback use. This may help overcome the limited use because of its time- and resource-demanding nature. Secondly, the treatment has a significant cost benefit over traditional biofeedback. The total consumer price will likely be constituted of only a one-time purchase of sensors, and no regular consultation costs. Finally, the treatment has a highly beneficial AE profile. Only one case of AEs was observed, and previous studies using the same setup observed similar AE profiles.[18, 19] This is superior to the most commonly used prophylactic drugs, which all have several AEs in the pediatric population.

There are several measures that should be considered for future iterations and studies of the similar app-based biofeedback treatments. The intervention should include more comprehensive instructions, guidance during biofeedback sessions, and even adjunctive therapies such as relaxation. Such features should be intelligently implemented into the app to ensure therapist-independence and may facilitate the effect of the treatment packages observed in traditional biofeedback. In addition, measures should be taken to keep adherence high through means such as regular reminders, motivation, and gamification.[35] Next, the use of a sham control group

should be carefully considered. As we experienced in this study, it is difficult to create a biofeedback sham that accurately mimics the effects of a proper placebo. A more fruitful approach might be to show non-inferiority compared to the most commonly used prophylactic medications, and the study should be powered to detect small treatment effects.

Study limitations

The main limitation of this study is the small sample size. This has clearly reduced the precision of our estimates and limited interpretability of clinical outcomes both in the biofeedback and sham groups. In addition to the small sample size, the study suffered from attrition and missing data. Several participants were excluded or declined to participate, and the overall adherence was low resulting in missing data, which further decrease confidence in our estimates.

Conclusion

In this study we observed a small reduction in headache frequency in the active treatment group that was not statistically significant nor superior over sham. The limited treatment effect may in part be explained by the minimalistic nature of the intervention, low adherence rates, attrition, and underpowered analyses. Still, the observed reduction in headache frequency suggests that an almost completely therapist independent biofeedback app may be an effective, highly tolerable and cheap treatment option, provided significant alterations to the treatment setup and study design are made. Future iterations of the intervention should include a more comprehensive intervention and ensure increased adherence through means such as gamification. Future studies of the intervention should strongly consider using an

active comparison group and be powered to detect small, but clinically relevant, treatment effects.

References

1. Wober-Bingol C. Epidemiology of migraine and headache in children and adolescents. *Current pain and headache reports* 2013; 17: 341. 2013/05/24. DOI: 10.1007/s11916-013-0341-z.
2. Krogh AB, Larsson B and Linde M. Prevalence and disability of headache among Norwegian adolescents: A cross-sectional school-based study. *Cephalalgia* 2015. DOI: 10.1177/0333102415573512.
3. El-Chammas K, Keyes J, Thompson N, et al. Pharmacologic treatment of pediatric headaches: a meta-analysis. *JAMA Pediatr* 2013; 167: 250-258. 2013/01/30. DOI: 10.1001/jamapediatrics.2013.508.
4. Powers SW, Coffey CS, Chamberlin LA, et al. Trial of Amitriptyline, Topiramate, and Placebo for Pediatric Migraine. *N Engl J Med* 2017; 376: 115-124. 2016/10/28. DOI: 10.1056/NEJMoa1610384.
5. Termine C, Ozge A, Antonaci F, et al. Overview of diagnosis and management of paediatric headache. Part II: therapeutic management. *J Headache Pain* 2011; 12: 25-34. 2010/12/21. DOI: 10.1007/s10194-010-0256-6.
6. Oskoui M, Pringsheim T, Billinghurst L, et al. Practice guideline update summary: Pharmacologic treatment for pediatric migraine prevention: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the American Headache Society. *Neurology* 2019; 93: 500-509. 2019/08/16. DOI: 10.1212/wnl.0000000000008105.
7. Trautmann E, Lackschewitz H and Kroner-Herwig B. Psychological treatment of recurrent headache in children and adolescents--a meta-analysis. *Cephalalgia* 2006; 26: 1411-1426. DOI: 10.1111/j.1468-2982.2006.01226.x.
8. Stubberud A, Varkey E, McCrory DC, et al. Biofeedback as prophylaxis for pediatric migraine: a meta-analysis. *Pediatrics* 2016; 138 2016/07/28. DOI: 10.1542/peds.2016-0675.
9. Fisher E, Law E, Dudeney J, et al. Psychological therapies for the management of chronic and recurrent pain in children and adolescents. *Cochrane Database of Systematic Reviews* 2018.
10. Schwartz MS and Andrasik F. *Biofeedback: A practitioner's guide*. Guilford Publications, 2017.
11. Lehrer P and Eddie D. Dynamic processes in regulation and some implications for biofeedback and biobehavioral interventions. *Appl Psychophysiol Biofeedback* 2013; 38: 143-155. DOI: 10.1007/s10484-013-9217-6.
12. Siniatchkin M, Hierundar A, Kropp P, et al. Self-regulation of slow cortical potentials in children with migraine: An exploratory study. *Applied Psychophysiology Biofeedback* 2000; 25: 13-32. DOI: 10.1023/A:1009581321624.
13. Penzien DB, Irby MB, Smitherman TA, et al. Well-Established and Empirically Supported Behavioral Treatments for Migraine. *Current pain and headache reports* 2015; 19: 34. 2015/06/13. DOI: 10.1007/s11916-015-0500-5.

14. Stubberud A and Linde M. Digital technology and mobile health in behavioral migraine therapy: a narrative review. *Current pain and headache reports* 2018; 22: 66.
15. Lalloo C, Jibb LA, Rivera J, et al. "There'sa pain app for that": Review of patient-targeted smartphone applications for pain management. *The Clinical journal of pain* 2015; 31: 557-563.
16. Mosadeghi-Nik M, Askari MS and Fatehi F. Mobile health (mHealth) for headache disorders: A review of the evidence base. *J Telemed Telecare* 2016; 22: 472-477.
17. Minen MT, Torous J, Raynowska J, et al. Electronic behavioral interventions for headache: a systematic review. *J Headache Pain* 2016; 17: 51. 2016/05/11. DOI: 10.1186/s10194-016-0608-y.
18. Stubberud A, Omland PM, Tronvik E, et al. Wireless surface electromyography and skin temperature sensors for biofeedback treatment of headache: Validation study with stationary control equipment. *JMIR Biomedical Engineering* 2018; 3: e1.
19. Stubberud A, Tronvik E, Olsen A, et al. Biofeedback Treatment App for Pediatric Migraine: Development and Usability Study. *Headache: The Journal of Head and Face Pain* 2020; 60: 889-901.
20. Johanson GA and Brooks GP. Initial scale development: sample size for pilot studies. *Educ Psychol Meas* 2010; 70: 394-400.
21. Headache Classification Committee of the International Headache Society (IHS) The International Classification of Headache Disorders, 3rd edition. *Cephalalgia* 2018; 38: 1-211. DOI: 10.1177/0333102417738202.
22. Tfelt-Hansen P, Pascual J, Ramadan N, et al. Guidelines for controlled trials of drugs in migraine: third edition. A guide for investigators. *Cephalalgia* 2012; 32: 6-38.
23. Penzien DB, Rains JC and Andrasik F. Behavioral management of recurrent headache: three decades of experience and empiricism. *Applied psychophysiology and biofeedback* 2002; 27: 163-181.
24. Autret A, Valade D and Debiais S. Placebo and other psychological interactions in headache treatment. *The journal of headache and pain* 2012; 13: 191-198. 2012/03/01. DOI: 10.1007/s10194-012-0422-0.
25. Lisspers J, Öst L-G and Skagerberg B. Clinical effects of biofeedback treatment in migraine: the relation to achieved self-control and pretreatment predictors. *Cognitive Behaviour Therapy* 1992; 21: 171-190.
26. Guarnieri P and Blanchard EB. Evaluation of home-based thermal biofeedback treatment of pediatric migraine headache. *Biofeedback and self-regulation* 1990; 15: 179-184.
27. Burke EJ and Andrasik F. Home- vs. clinic-based biofeedback treatment for pediatric migraine: Results of treatment through one-year follow-up. *Headache: The Journal of Head and Face Pain* 1989; 29: 434-440.
28. Scharff L, Marcus DA and Masek BJ. A controlled study of minimal-contact thermal biofeedback treatment in children with migraine. *Journal of pediatric psychology* 2002; 27: 109-119. Clinical Trial; Comparative Study; Randomized Controlled Trial; Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, P.H.S.
29. Powers SW, Mitchell MJ, Byars KC, et al. A pilot study of one-session biofeedback training in pediatric headache. *Neurology* 2001; 56: 133.

30. Ramsey RR, Ryan JL, Hershey AD, et al. Treatment adherence in patients with headache: a systematic review. *Headache: The Journal of Head and Face Pain* 2014; 54: 795-816.
31. Gewirtz A and Minen M. Adherence to behavioral therapy for migraine: knowledge to date, mechanisms for assessing adherence, and methods for improving adherence. *Current pain and headache reports* 2019; 23: 3.
32. Minen MT, Adhikari S, Seng EK, et al. Smartphone-based migraine behavioral therapy: a single-arm study with assessment of mental health predictors. *NPJ digital medicine* 2019; 2: 1-9.
33. Mullinix JM, Norton BJ, Hack S, et al. Skin temperature biofeedback and migraine. *Headache* 1978; 17: 242-244. 1978/01/01. DOI: 10.1111/j.1526-4610.1978.hed1706242.x.
34. Nestoriuc Y and Martin A. Efficacy of biofeedback for migraine: a meta-analysis. *Pain* 2007; 128: 111-127. DOI: 10.1016/j.pain.2006.09.007.
35. Pramana G, Parmanto B, Lomas J, et al. Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. *JMIR serious games* 2018; 6: e9.

Table 1. Participant demographics.

	Biofeedback group (n=12)	**Sham group (n=4)**
Age, mean±SD (range)	15±1.6 (13-18)	14±2.3 (12-16)
Female, n (%)	10 (83%)	1 (25%)
Migraine aura, n (%)	9 (75%)	2 (50%)
Other headache disorders:		
TTH, n (%)	8 (67%)	3 (75%)
MOH, n (%)	1 (8%)	1 (25%)
Tried triptans, n (%)	9 (75%)	3 (75%)
Tried migraine pharmacoprophylaxis, n (%)	3 (25%)	1 (25%)

Table 2. Median and mean estimates of headache outcomes at baseline, weeks 1-4 and weeks 5-8 *within* the biofeedback group. The two rightmost columns show the mean difference with 95% confidence intervals and p-values of a two-tailed Wilcoxon signed rank test comparing outcomes at weeks 1-4 and weeks 5-8 versus baseline. IQR=interquartile range; SD=standard deviation; MD=mean difference; CI=confidence interval

		Baseline	**Week 1-4**	**Week 5-8**	**Baseline vs weeks 1-4, MD (95% CI); p-value**	**Baseline vs weeks 5-8, MD (95% CI); p-value**
Headache frequency	Median (IQR)	10.0 (7.0-14.0)	8.0 (4.5-14.0)	8.0 (5.5-14.0)	-2.917 (95% CI -6.838 to 1.004); *P*=0.145	-1.857 (95% CI -5.723 to 2.009); *P*=0.395
	Mean (SD)	11.2 (5.5)	8.2 (6.0)	9.6 (7.3)		
Maximum intensity	Median (IQR)	1.7 (1.7-1.9)	1.8 (1.4-1.9)	2.0 (1.7-1.9)	-0.311 (95% CI -0.973 to 0.352); *P*=0.754	-0.079 (95% CI -0.806 to 0.649); *P*=0.735
	Mean (SD)	1.9 (0.5)	1.5 (0.8)	1.7 (0.8)		
Average intensity	Median (IQR)	1.5 (1.4-1.7)	1.4 (1.2-1.7)	1.6 (1.2-1.7)	-0.209 (95% CI -0.721 to 0.303); *P*=0.347	-0.14 (95% CI -0.889 to 0.608); *P*=0.735
	Mean (SD)	1.5 (0.3)	1.2 (0.7)	1.4 (0.7)		
Daily functioning	Median (IQR)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	-0.167 (95% CI -0.414 to 0.081); *P*=0.157	-0.143 (95% CI -0.492 to 0.207); *P*=0.317
	Mean (SD)	1.0 (0.0)	0.8 (0.4)	0.9 (0.4)		
Abortive drug consumption	Median (IQR)	6.5 (2.8-10.0)	3.0 (1.0-10.0)	7.0 (4.0-10.0)	-3.083 (95% CI -6.89 to 0.724); *P*=0.092	-1.857 (95% CI -6.792 to 3.078); *P*=0.446
	Mean (SD)	7.2 (6.0)	4.2 (4.2)	6.7 (4.6)		

Table 3. Changes in headache outcomes in the biofeedback group vs. sham group. Note that negative values in the two rightmost columns indicates a favor towards the biofeedback group. BFB=biofeedback; IQR=interquartile range; CI=confidence interval

	Group	Baseline, Median (IQR)	Weeks 1-4, Median (IQR)	Weeks 5-8, Median (IQR)	Difference in change score from baseline to weeks 1-4, Median (95% CI); p-value	Difference in change score from baseline to weeks 5-8, Median (95% CI); p-value
Headache frequency	BFB	10.0 (7.0-14.0)	8.0 (4.5-11.5)	8.0 (5.5-14.0)	0.5 (95% CI -9.0 to 16.0); *P*=1.0	-1.0 (95% CI -9.0 to 4.0); *P*=0.760
	Sham	12.5 (8.2-18.2)	8.0 (5.5-11.0)	7.5 (6.2-8.8)		
Maximum intensity	BFB	1.7 (1.7-1.9)	1.8 (1.4-2.0)	2.0 (1.7-2.2)	0.1 (95% CI -0.6 to 0.6); *P*=0.585	0.2 (95% CI -1.6 to 0.8); *P*=0.883
	Sham	2.2 (2.0-2.3)	2.0 (1.9-2.1)	2.2 (2.2-2.2)		
Average intensity	BFB	1.5 (1.4-1.7)	1.4 (1.2-1.6)	1.6 (1.2-1.7)	-0.2 (95% CI -0.8 to 0.3); *P*=0.303	-0.4 (95% CI -2.0 to 0.6); *P*=0.464
	Sham	1.5 (1.3-1.8)	1.8 (1.6-1.8)	2.1 (2.0-2.2)		
Daily functioning	BFB	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	0.0 (95% CI 0.0 to 0.0); *P*=0.460	0.0 (95% CI -1.0 to 0.0); *P*=0.789
	Sham	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)		
Abortive drug consumption	BFB	6.5 (2.8-10.0)	3.0 (1.0-5.8)	7.0 (4.0-9.0)	-5.0 (95% CI -10.0 to 3.0); *P*=0.301	-2.0 (95% CI -12.0 to 4.0); *P*=0.769
	Sham	3.0 (1.0-5.0)	2.5 (0.0-6.2)	6.0 (5.5-6.5)		

Table 4. Physiological measurements in the biofeedback sessions. The table shows the mean physiological measurements of the first and last minute of sessions with a duration of at least five minutes. Because participants completed different number of sessions, we compared the average of the two first sessions with the average of the two middle sessions and the two last sessions. Note that while there is a slight increase in end session temperature from the two first sessions to the two last sessions, the amplitude of within-session change is diminished throughout sessions. SD=standard deviation; bpm=beats per minute; SEMG=surface electromyography; mV=millivolts

		Two first sessions	**Two middle sessions**	**Two last sessions**
Peripheral skin temperature, °Celsius (SD)	Session start	30.7 (4.0)	31.6 (3.9)	32.8 (3.4)
	Session end	36.3 (4.1)	36.1 (3.9)	37.6 (1.6)
Heart rate, bpm (SD)	Session start	71.0 (24.7)	77.5 (17.5)	74.0 (20.7)
	Session end	81.7 (10.5)	80.9 (8.6)	79.0 (6.0)
Trapezius SEMG voltage, mV (SD)	Session start	15.1 (18.2)	19.5 (32.5)	20.5 (32.6)
	Session end	8.8 (2.2)	8.5 (1.6)	17.0 (37.1)

Figure 1. CONSORT flow diagram.

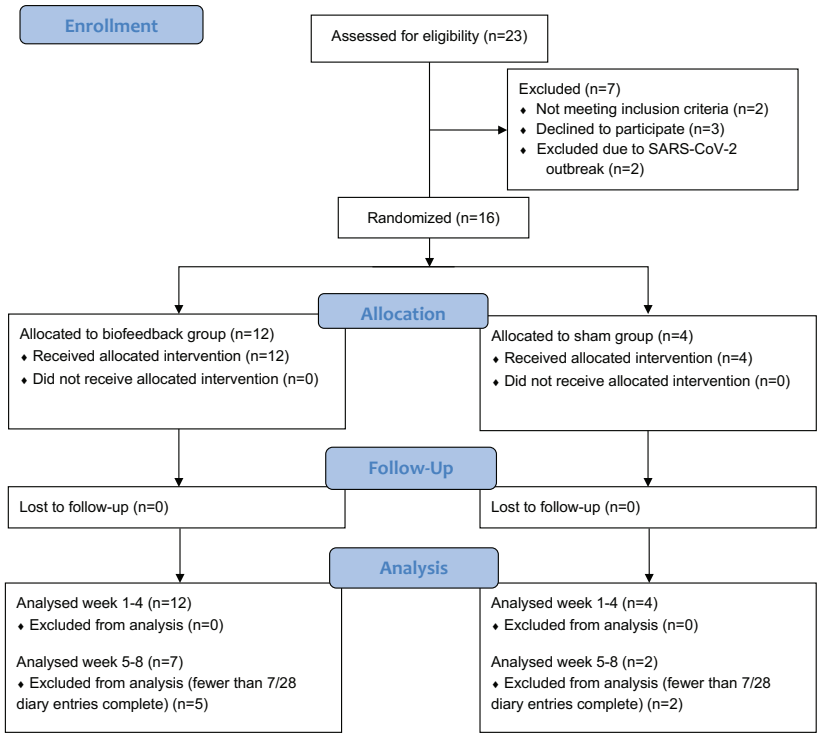


Figure 2. Visualization of raw physiological data over the course of all biofeedback sessions. The thick lines represent mean raw physiological value averaged over a moving window of three sessions. The shaded gray area represents the corresponding moving average of one standard deviation from the mean. Note the slight increasing trend in finger temperature, and slight decreasing trend in minimum heart rate and maximum muscle tension up to session 40. The number of participants completing more than 40 sessions was low, thus yielding the highly variable trends in session 40-60.

ISBN 978-82-326-5158-0 (printed ver.)
ISBN 978-82-326-5159-7 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology