

Playing to (Self-)Rehabilitate: A Month-Long Randomized Control Trial with Brain Lesion Patients and a Tablet Game

Hendrik Knoche, Kasper Hald
Aalborg University
Rendsburggade 14
9000 Aalborg, DK
+45 99402661
{hk|kh}@create.aau.dk

Dorte Richter, Helle Roving Møller Jørgensen
Brønderslev Neurorehabiliteringscenter
Nørregade 77
9700 Brønderslev, DK
+45 976 40152
{dorte.richter|hrmj}@rn.dk

ABSTRACT

We designed and evaluated a whack-a-mole (WAM) style game (see Figure 1) in a clinical randomized controlled trial (RCT) with reminder-assisted but self-initiated use over the period of a month with 43 participants from a post-lesion pool. While game play did not moderate rehabilitative progress indices of standard neuropsychological control tests, it did significantly improve in-game performance when compared to the control group, and its performance indicators were highly accurate in predicting neglect. Patients found playing beneficial to their rehabilitation and attributed gains in the attention training properties of the game. The game showed potential for bedside assessment, insight support, and motivation by providing knowledge about rehabilitative progress.

CCS Concepts

• Human-centered computing→Field studies • Applied computing→Health informatics

Keywords

Self-rehabilitation; game performance; patient insight; stroke; neglect; whack-a-mole; classification

1. INTRODUCTION

Increasing health-care costs and ageing populations will require patients to take more responsibility to improve and maintain their health [24]. Rehabilitation is costly and leaves time for motivated patients to train if they can carry out relevant activities unassisted, find them beneficial, and muster the initiative. Much research focuses on motor recovery, but there is a greater need to address cognitive training [37], and going from proof of concept to providing evidence of effectiveness of interventions, e.g. through randomized controlled trials (RCT). Strokes are the leading cause of severe disability [23], and many of them suffer from neglect with poor insight into their inability to attend to their left visual field. The trend in turning activities beneficial to patients' health into games or gamifying them has seen a large push to tap the intrinsic motivation that can help patients adhere to or increase their the

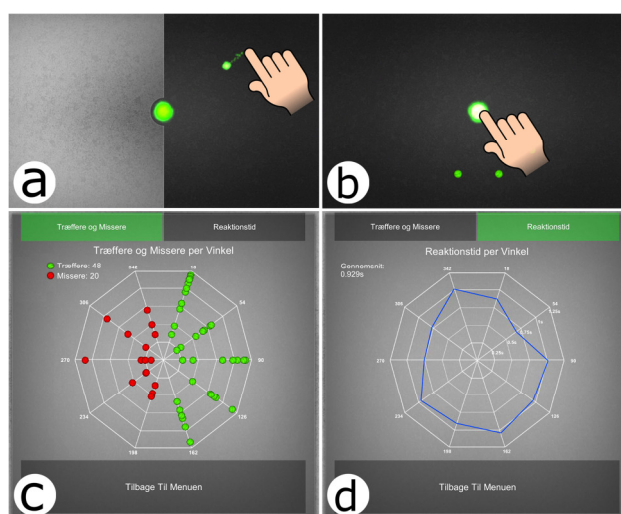


Figure 1: WAM design (left, top) in which a hit target flings back to the center, b) tap on the center button makes multiple targets appear c) number and spatial distribution of hits and misses d) hit delays per axis on which the targets appeared

required an intense and repetitive regimen. Unlike standardized tests games can easily support varying degrees of difficulty as patients might be unwilling to complete tasks that they - due to cognitive, initiative, or motivational deficits - experience or deem too difficult [33]. Research needs to address rehabilitation games, which:

1. allow for long-term, self-initiated, and unsupported play by elastically adapting to a range of impairment degrees,
2. yield performance parameters indicative of patients condition and progress [11],
3. improve patients insight into their own condition [16] and progress, and
4. provide benefits in non-game scenarios.

To this end, we developed a game simple enough even for more severely affected patients but still worthwhile for the majority to play over a four week period in a randomized controlled trial. While playing WAM did not result in measurable gains in neuropsychological control measures its game performance data allowed for a solid classification of neglect patients. Patients used WAM as a way to recognize progress and train their attention. Clinical staff found it a useful a tool for bedside assessment and as

an independent measure that provided grounds for discussions with patients to improve their insight into their condition.

2. BACKGROUND AND RELATED WORK

Unilateral spatial neglect (USN) is a disorder in which patients, despite functioning eyes, have difficulty attending to the left hand side of their visual field. Neglect follows right hemisphere stroke in the acute stage in about 50% of cases [8]. USN patients typically have poor insight into their own condition and exhibit poor coping strategies, e.g., they do not adapt a different head body orientation to counter their impairment *vis-a-vis* their environment as, for example, a patient with hemianopia might. Methods for USN diagnosis include copying pictures with pen and paper, striking off each dot in a dotted letter, judging whether which of two bars (left/right) appears first [28], bisecting lines or cancellation tests in which patients strike out (cancel) target items on a piece of paper with a cluttered array of targets and distractors. USN patients fail to copy the left part of a depicted objects [6], only strike off the right hand part, tend to bisect lines not in the middle but make substantial deviations in the direction of their impaired side, and fail to cancel targets there. Mattingley *et al.* showed that neglect patients could exhibit motor neglect – a difficulty in initiating leftward movements towards targets on the left side of their visual field [22]. Proponents of visual restoration therapy (VRT) posit that by presenting visual stimuli at the border of the field of vision patients can increase their field of vision [29].

Specific instruments to diagnose and quantify neglect include the following. The Line Bisection (LiBi) test requires participants to mark the middle of a series of horizontal lines [14,32]. In the Line cancellation (LiCcl) tests participants should cross out 40 lines that are arranged in seven columns but appear randomly scattered on a sheet of paper [36]. The letter cancellation task (LetCcl) contains five rows of 34 letters of which 40% are targets (E,R) to cancel out without a time limit [14]. Poor performance in these three tests above established cut-off scores on both the left and the right hand side indicate a general attention deficit rather than USN. The Catherine Bergego Scale (CBS) assesses degrees of neglect in ten daily life tasks in the personal, peripersonal, and extrapersonal space from both self- and observer reported (CBS_{obs}) ratings [5]. For example it assesses whether a person exhibits no, mild, average or severe signs of neglect when shaving his face. The difference between ratings the patients’ self-assess and those from clinical staff provide a measure of the patients’ insight deficit into their condition (CBS_{id}). The Symbol Digit Modality Test (SDMT) assesses the scanning and tracking aspect of attention similar to the ones at work in the Letter Cancellation and visual selective attention [34].

The Functional Independence Measure (FIM) battery is a more general and common measure in rehabilitation that assesses the quality of *activities of daily living* on two sub-scales - motor (FIMm) and cognitive (FIMc) [12]. Typical ways of quantifying rehabilitation impact use measures obtained at admission (AD), discharge (DC), and in some cases pre-morbid (PM) or maximum scores attainable (MAX). Rehabilitation impact indices include, absolute and relative gain, effectiveness (REs) and efficiency (REy) – see [9] for details. In this paper we use the revised Montebello Rehabilitation Factor Score (MRFSR) to test gains in CBS, SDMT, and FIM.

$$MRFSR = \frac{(DC - AD)/DC}{(MAX - AD)/MAX} \quad (1)$$

See Koh *et al.*’s overview on factors that moderate rehabilitation indices [19].

2.1 Related work

Applications for (self-)rehabilitation need to be motivating, provide interactivity and progress to be used continuously [35]. Games are sought as a means to tap into the intrinsic motivation they provide to counter the repetitive and uninteresting nature of many rehabilitation activities [21]. HCI research has started to address the design, implementation and evaluation of bespoke games, e.g. for physical therapy [11], rehabilitation [13], enjoyment for children with motor disabilities [15]. Bespoke games for this audience often need to be simple [21], require simplified control schemes or accessibility features [15], and need to adapt to player performance that can vary over the course of a day [3]. Game challenge needs to adapt to player performance or risks becoming boring within the span of one session [1] or too hard and disengage patients as neuropsychological tests do [33]. However, the adaptation of challenge has to happen gently as patients may perceive steep challenge increases as aggressive behavior on the part of the game [7].

Maintaining player motivation for long-term rehabilitation requires the player to either find fun in the games or to recognize a beneficial effect from playing the game. Alankus *et al.* ran a six-week physical rehabilitation program with one participant using a motion-controlled game. The participant reported only playing for fun for the first half of the study, after which she started to recognize increasing performance in everyday life, greatly increasing motivation to continue as well as the desire to set and achieve personal goals [1]. Balaam detailed design tensions between conflicting goals of enjoyable barrier-free game play and rehabilitation needs, i.e., making actions difficult enough to yield both training effects and motivation [3]. According to motivation theory [31] a feeling of growing competence is important for developing and sustaining intrinsic motivation.

Jamieson stressed the importance of communicating the need for assistive technology to people with acquired brain injury to motivate them to engage with it [16]. Helping patients gain insight into their condition is an important concern for neglect patients mirrored by the dedicated measure of the Catherine Bergego Scale.

2.2 Design considerations

The game presented in this paper follows the idea of visual restoration therapy [29] by continually presenting visual stimuli at the border of the field of vision to increase it. The design is based on understood performance measures to quantify performance in cancellation tests (see Dalmaijer *et al.*’s overview [10]). We focus on performance measures that match with the conceptual design of WAM. Rorden *et al.* popularized the *Center of Cancellation* (CoC) - the average of the x,y positions of all cancelled targets from e.g. a line cancellation test [30]. This spatial measure allows for both lateral and near-far (targets towards the bottom are closer than those on the top) neglect discrimination. Rabuffetti found that the temporal measure of *inter-cancellation times* – the time between one and the following cancellation – was higher for patients than for controls [26]. Knoche *et al.* found that empirical parameters from Fitts’ law (*a*, *b*) from modelling rapid touch interactions in games can help predict neglect [18].

3. DESIGN

We implemented a tablet-based game – Whack-a-mole (WAM), in which 6mm small targets (moles) appear and stay for three seconds before disappearing (expiring) and the player hits by tapping on them (see Figure 1b). When targets are present the center button is mostly white (see Figure 1b), otherwise green. To

direct the player’s gaze back to the center, a hit target - in a springing motion - flies back (see Figure 1a) to merge with the center and a center button tap spawns new targets. Initially, targets appear close to the center and successful hits increase the radius in discrete steps of 10% from the current maximum distance hit. Expired targets reduce the radius such that the game adjusts the challenge in each session individually and USN patients play close to the border of their visual field. Targets hit fast (within one second) increase the pitch of the feedback sound with no pitch ceiling for consecutive fast hits. But a single slow hit or expiry brings the pitch straight down to its starting value. After an initial calibration phase with 14 single targets on the right hand side, the game goes through different stages with single, sequential, multiple targets, and multiple targets along with distractors. Advancement through stages depends on game performance and in the case of the stage with distractors on a fixed time schedule. While WAM was mainly developed for USN patients it was built to support competitive play with no ceiling on the number of possible hits.

The game had been iteratively tested and evaluated with patients (both neglect and attention deficit disorder) and staff in individual sessions and a short pilot trial of 10 days [18]. We paid particular attention to both feedback of in-game actions and session performance and game challenge [7]. The tests informed design choices such as target expiry times, their placement, tactual recognition field size, and audio-visual stimuli during gameplay, both for spawning and feedback when hitting them. We removed initial extant information such as in-game point counters so as not to cognitively overload severely impaired players. A resident psychologist at the clinic judged that WAM required and trained sustained attention. The collaborating therapists advised to keep game play and time to review the results to within 10 minutes. After eight minutes, the game ended and depicted the number and spatial positions of both hits and misses (*c.f.* (see Figure 1c), and hit delays along the 10 axes in two summary screens (Figure 1d). However, it did not provide an overview of the progress of the patient over time.

4. STUDY

We designed a field trial using the instruments in Table 1 to see whether self-decided playing of WAM 1) was possible for a range of patient impairments, 2) usage could be predicted from the perceived *fun*, *ease of use* and *benefit* after an initial game play, 3) yielded performance indicators in line with the patients’ recovery, 4) had clear outcome benefits for the participants and 5) whether hard and software choices were compatible with the constraints of a clinical setting. We used a randomized controlled trial to test (6) whether playing WAM had an effect on the rehabilitation indices of measures of:

1. neglect and its associated insight deficit in daily life tasks (CBS) to test whether it reduced impairment or having continued exposure to game results improved patient insight in their impairment,
2. an attentional measure (SDMT) to test whether WAM in-game performance improvements due to training transferred to this related but untrained measure, and
3. more general motor and cognitive skills (FIM).

Patients received a personal tablet (iPad2 or iPadAir in a protective shell) during their trial period and a patient’s table was - if necessary - equipped with some anti-slip rubber sheet for the tablet to rest on while playing. The tablet contained WAM and a gamified Trail Making Test (TMT) [27] – a popular instrument to measure attention and executive functioning - that was stratified

with different difficulty levels. We limit reporting in this paper to the WAM game.

4.1 Data collection

We collected the following control variables whose fulfillment or higher values are associated with decreases in our dependent variables – the patients’ rehabilitation impact indices [19]: age, length of stay in the trial (LoT) and in the clinic (LoS), cognitive impairment (FIMc), time delay from lesion onset to admission into rehabilitation unit (admission delay), gender (female), and USN. Our questionnaires also addressed previous experience with mobile devices coded as no use (N), use (U) and use incl. games (G), WAM’s perceived *benefit*, *fun*, and *ease of use* (experienced difficulty inversed), motivation and family support, lesion details, along with the diagnostic tests for neglect: Line Bisection, Letter Cancellation, and Line Cancellation, and control tests: SDMT, CBS (both observed and self-reported), cognitive (FIM_C) and motor (FIM_M) sub-scores of FIM with their dates (*c.f.* Table 1). We used a three 20cm line staircase version of the LiBi test along with our own version that included five randomly positioned shorter lines from 10cm down to 6mm (in half steps). Along with the questionnaires, an occupational therapist administered the control tests as indicated in Table 1 at entry and exit to the trial and – patient stay duration permitting – during further bi-weekly periodical tests in between.

During the enrollment (entry) patients were introduced to handling the tablet, responding to reminders (alarms scheduled on the tablet), unlocking the tablet screen, starting the game, logging in as themselves by tapping on a button with their name (a second button was labeled *guest*), and an eight minute play-through of each app. They responded to a questionnaire on the perceived benefits, difficulty and fun. On the tablet, the therapist set three alarms compatible with the patients rehab schedule.

All in-game interactions were logged with timestamps and spatial coordinates on the devices. This allowed for applying neglect measures such as Center of Cancellation, and performance comparisons (number of hits, misses, and expiries) of left and right hand sides, and temporal modelling of interaction data with Fitts’law [18]. The caretakers filled in paper logs on the patients’ desks when they required help. We obtained feedback both during and after the trial from the caretakers on further observations of and comments from patients and how WAM worked for the clinical staff as an addition to patient activities.

Table 1: Instruments for data collection

Instruments	Entry	Periodical	Exit
FIM Motor	● *		● *
FIM Cognitive	● *		● *
Line Bisection	●	●	●
Letter Cancell.	●		●
Line Cancell.	●		●
SDMT	●	●	●
CBS _{obs,id}	●		●
supervised app use	●		●
app questionn.	●	●	●
game interaction logs	during WAM use		
problem logs	incident based		
clinic. staff interviews	throughout trial		

* obtained at submission and discharge from clinic

4.2 Data preparation

To test whether playing WAM had an effect on the patients' rehabilitation impact indices we computed the rehabilitation index (MRF_{SR}) for our control measures: SDMT, CBS_{obs}, CBS_{id}, FIM_C, and FIM_M. As can be seen in Table 1, we obtained the base scores at entry and at exit from the trial. However, the FIM scores were only available at admission and at discharge from the clinic. We relied on modeled FIM scores from a linear regression to account for gains during non-trial times. In the absence of pre-morbid test scores we used the maximum scores possible for FIM_C (35), FIM_M (91), and CBS (30, inverted). For the maximum score for SDMT we relied on the age specific normative test scores plus three times the normative standard deviation (both from [20]) for each participant. In absence of prior knowledge we assumed linear recovery regarding the patients' attention ability as measured through SDMT tests. To reduce noise in the SDMT scores (test-retest reliability is 0.8 in healthy adults) we used linear regression modeled scores from the obtained SDMT scores (entry, periodical(s), exit) wherever periodical test results were available.

In Table 2, line bisection scores are the average absolute deviation from the middle of the lines: zero percent being the middle and 100% being either end of the line. The letter and line cancellation scores are computed as the imbalance of omissions on either half of the page; the percentage points of omissions, relative to the entire page, on one half of the page subtracted from the percentage points of the other half. For example, if on the left hand side 4 out of 10 and on the right 6 out of 10 targets were omitted then the scores would be 6/20-4/20=10%.

From the logged data we extracted the following parameters on a per game session basis both for targets on the left and right hand side of the screen: Averages of hit delays, number of hits and expiry counts, x- and y target touch offsets between a target hit coordinates and the center of the target., and Fitts law model variables (*a* and *b*) of all moles distances and their corresponding hit delays. To obtain a *center of hit* measure (CoH) we computed the average x and y position in millimeters from the center (left negative, right positive) and summed them up.

4.3 Participants

Patients at the clinic who volunteered to participate were excluded from the study if they could not:

- give informed consent,
- complete a game session with a therapist's assistance due to poor eyesight or hearing, lack of arm-hand mobility, or cognitive capabilities, or
- respond to the alarms set on the tablet.

Originally, 52 patients of a rehabilitation clinic volunteered to participate and were randomly assigned to either treatment or control group. Out of these 42 yielded complete data sets; a move to a different facility was the most common for drop-out. Thirty-one men and eleven women (63 years old on average, SD: 14.8) completed the trial. Table 2 summarizes the participant profiles according to control variables and neuropsychological test measures. We relied on Jehkonen *et al*'s test suite (line bisection, line cancellation and letter cancellation [17]) and their cut-offs from the literature for a robust neglect classification. A positive outcome in one of the three test classified participants as having neglect. Four participants suffered from neglect (three in the treatment group). Seven participants (3 in the treatment group) had above cut-off scores on both the left and right hand side of the paper tests and the health care professional in charge of conducting all tests (and co-author of this paper) classified them

Table 2: Participant demographic data of Treatment and Control group, and the 14 USN patients (6 in test, 8 in the control group) and their neuropsychological test scores at entry

	Treatm. (T) Avg and (SD)	Control (C) Avg and (SD)	Neglect Avg and (SD)
Age	63.6 (17.0)	60.7 (12.7)	73.0 (7.9)
Gender	13M / 7F	14M / 4F	4M
handed *	7 l / 12 r / 1 a	8 l / 10 r	4 r
LoS†	73.0 (30.1)	72.9 (34.7)	64.8 (26.4)
LoT‡	31.1 (8.0)	32.4 (8.7)	30.5 2.4
Lesion side†	9 l / 9 r / 2 u	9 l / 5 r / 1 b / 3 u	4 r
Adm. Delay	46.5 (71.6)	26.5 (22.1)	37.3 (38.3)
Mobile use**	8N / 11U / 1G	8N / 9U / 1G	2N / 2U
FIM motor	52.4 (26.5)	48.0 (34.1)	35.9 (24.3)
FIM cognitive	63.3 (16.4)	49.4 (17.9)	50.0 (13.6)
Line Bisection	5.8% (4.4)	6.0% (6.2)	46.3% (33.1)
Letter Cancell.	3.4% (3.4)	4.7% (5.8)	23.8% (12.0)
Line Cancell.	0.1% (0.6)	0.6% (2.0)	21.5% (21.6)
Norm. SDMT	0.16 (0.07)	0.18 (0.10)	0.05 (0.17)
CBS observed	8.7 (6.67)	7.6 (6.4)	19.6 (7.4)
CBS insight def	5.1 (5.26)	3.9 (4.0)	12.8 (9.8)

* left (l), right (r), ambidextrous (A); † LoS – Length of stay at clinic; ‡ LoT – length of trial; † left(l), right (r), both (b), unknown (u), ** no use (N), use (U), use incl. gaming (G)

as attention deficit disorder cases. We refer to the patients by initials for neglect (N), attention deficit (A) and other participants (P) along with their participation number (PID).

We found no significant differences between the control and treatment groups for the control variables in Table 2.

5. RESULTS

After having experienced the apps for the first time, the participants in the treatment group found WAM *easy to play* (mean five-point Likert score 4.0, SD=1.3), *fun* (3.6, SD=1.3) and *beneficial* for their rehabilitation (3.6, SD=1.3). These opinions were similar for patients with and without neglect. At exit participants in the treatment group found WAM *easier to play* (4.3), similar in *fun* (3.5) but less *beneficial* (2.8, SD=1.4) than at entry. In comparison to the treatment group, the control group found WAM at entry *easier to play* (4.2), similarly *fun*, but less *beneficial* (3.3). The control group's opinions of WAM remained similar at exit in terms of *fun* and *easy to play*, and with a smaller reduction in perceived benefit (3.0).

The most common problems for which participants required assistance during the trial related not to playing the game but charging the tablet, waking it from sleep and disabling the set alarms, which annoyed roommates. Some patients felt stressed by the alarms, since they were not able to act on the alarm or could sometimes not remember its purpose - to remind them to initiate their self-training. A few participants needed support up to the first five times to start up the tablet, open and start the game. "I never thought I could figure it out. But after having been shown it 3-4 times I could even start the game up and play when I had time - and I actually ended up finding it pretty fun - I asked my children for an iPad for Christmas." (P37)

All but one patient (P30, f, 71yrs) managed to play WAM by themselves. Neither P30 nor P21 could perform the SDMT test at their entry test but P21 had no problems playing WAM. Usage

occurred generally between 8:00 and 21:00, mostly during the morning (9-11), early afternoon (15-16) and early evening (18-19) throughout the week (7.3min per day) with a reduction over the weekends (4.2min). Some of them went home over the weekend and could take along the tablet but we did not track who did.

On average, the treatment participants played WAM for 5.6 minutes per day (SD 1.8). Usage varied hugely between participants (see Figure 2 for an overview). We used Jenks natural breaks optimization [25] to classify their per day use in minutes into four (no, low, medium, high) levels of use:

1. up to 0.6 - one participant (down to PID30 in Figure 2)
2. up to 3.7 - seven participants (down to PID 25)
3. up to 10.7 - 12 participants (down to PID 18)
4. up to 15.1 - three participants (down to PID32)

In other words the medium use group played a little more than one game and the high use group two games per day. We tested for potential novelty effects and their wear offs. Due to varying participation duration in the trial we compared the first nine days after enrolment with the subsequent nine days in the trial to investigate possible novelty effects. We found no reductions in average daily usage after the beginning of the trial but a few participants who had been playing clearly stopped after a while (PID 20, 25, 12, 35).

Usage of WAM was positively correlated to a moderate degree with numerical values of the perceived *fun* ($r=.43$, $N=23$) and *benefit* ($r=.48$) reported at entry to the trial. The more fun the participants reported and the more they deemed to benefit the more they would later play. The perceived ease of use did not correlate with play time. Treatment group participants without neglect whose perceived *benefit* from playing WAM dropped significantly during the course of the trial did not play less than those whose perceived *benefit* remained the same or improved. One participant (P37) reported having played only out of obligation to the trial and not due to any perceived or assumed health benefits.

Participants derived fun from quick successions of hits that were possible in multi-target stages and competing with themselves and another. *It is fun when you go faster and faster (P2). There is some competition to the game. It is fun (P4). Monotonous, but good that you have to be quick (N1).* Becoming faster in the game became a goal in itself for some. *It is good to practice getting quicker (P24).* While for some participants the difficulty level was just right *“I have played this the most, have an easier time remembering how to play it.” (AD20)* the more able participants found that the degree of challenge could have been higher. The participants associated difficulty with the number of targets and distractors on the screen and not the speed at which they could hit these as long as it was within the expiration time. *Was fun after a while ... the difficulty ramps up too slowly (P43), There could be more dots on-screen, it could be more difficult (P51). The game is easy to get started with - though ... I might have wanted more challenges – it got harder, but also got a little boring over time. (PID32).*

We found a significant difference in the change of hitting performance over time between the treatment and control group – the treatment group participants improved their performance by 90 additional targets on average by the end of the trial - more than the control group (33 additional targets) $t(30.9)=-2.7$, $p<.05$ - indicating a training effect. Some patients attributed their performance gains in WAM to improved concentration from playing WAM. *It is concentration that is trained (P9). I think it*

helps me a lot. It helps you think faster (P18). One participant specifically valued the game not requiring reading. *Being able to practice concentration, without being able to read and write, which can often be an obstacle (P43).*

For rehabilitation impact we used linear regressions including *age, gender, length of stay, admission delay, having neglect, and cognitive impairment (FIMc)* as predictors of the rehabilitation impact index (MRFSR) of FIM_C , FIM_M , SDMT, CBS_{id} , and CBS_{obs} . We found no significant effects of WAM play time on any of these. The rehabilitation impact index scores of FIM_M were moderated by *age, having neglect, and length of stay*, FIM_C and CBS_{obs} by *length of stay*. Higher values of the predictors were associated with reductions of their rehabilitation indices as known from the literature.

The result screen was instrumental in judging performance and progress. Many patients found it sufficient to compare the outcome of a game session with their current remembered best score to gauge their improvements. *I like to compete with myself ... I can try to reach more hits next time (P2).* One participant discovered differences depending on the time of day by comparing his attained scores to remembered ones. *The results screen is nice. You can see whether you are quicker at certain times of day (P52).* A couple of participants desired more support to compare game performance with historical scores as a manifestation of progress and as motivation. *I lack being able to see my progress from game to game. It would be a carrot for me (P52).* Another one inquired about normative data on performance

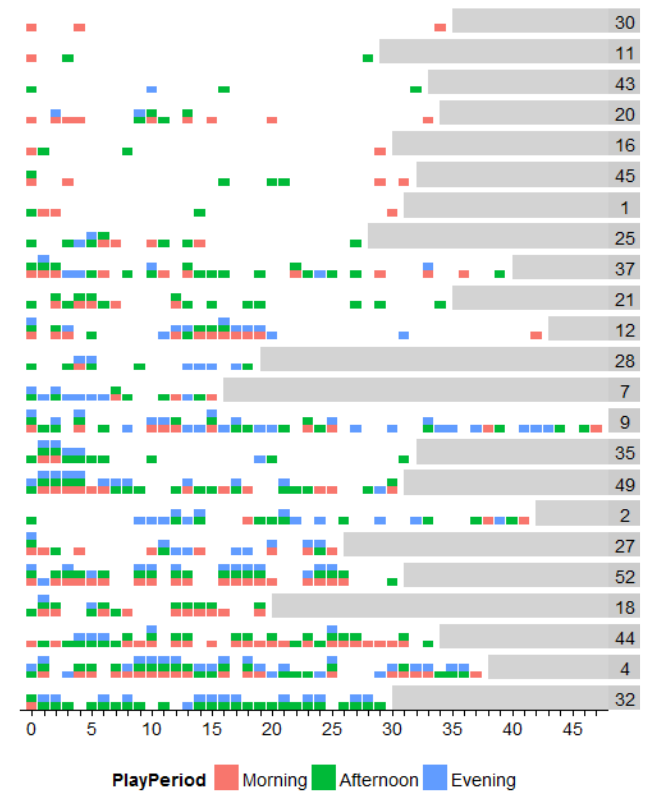


Figure 2: WAM usage by participant (rows) over time in the trial ordered by usage in minutes per day – each box indicates at least one game started during the morning (before noon), afternoon (before 6pm), or evening (after 6pm). Apparent order violations, e.g. PID20, are due to incomplete sessions.

and recovery. *The result screen is good, but you need an explanation of what you are working towards and how you have been doing so far* (P27).

The clinical staff liked that the patients were able to engage in meaningful activities that did not require their supervision. The therapist found WAM results a useful point of reference in discussions with the patients. Specifically, patients could not blame poor performance on other exogenous factors, e.g. in a mundane situation where they might explain away their inability to attend to important events by pointing to distracting factors. The game had been designed to not provide additional stimuli. Therefore, the game made for and was understood by the patients as a reliable and objective measure. After the end of the trial, the therapist in charge continued using WAM in her day-to-day work with patients and three participants asked for the game to continue playing after discharge.

We ran a cross-validation on the per session data of WAM to see how well game play performance can predict whether players suffered from neglect. From all WAM sessions in which participants had hit at least ten targets after the calibration phase we selected randomly half as the training data set for a stepwise binary logistic regression. Three parameters were significant predictors of the binary outcome variable: CoHx in mm from the centre of the screen, $p < .01$, Fitts' a left (a_L), $p < .001$, and Fitts' b left (b_L), $p < .01$. On the test set these predictors (see Eq. 2) yielded a classification accuracy of 98.3% and 98.2% on the training set.

$$z = -10.42 + 0.21CoH_x + 5.2a_L + 9.54b_L \quad (1)$$

Figure 4 depicts the corresponding ROC curve yielding a 0.939 area under the curve (auc). Misclassified sessions are depicted in red for each parameter in Figure 3. Misclassifications were mostly (11/14) false negatives of sessions from USN patients. For example, all three of N16 and 3/5 of N1's sessions were misclassified. N16 had not tested positive for neglect in either LiBi nor LiCcl and in LetCcl had an above cut-off score only on the right side. N1 recovered during the trial, which was mirrored by improvements in CoH_x. Two of N21's four misclassified sessions appeared to be outliers in terms of performance in CoHx and hits and the therapist had seen his wife helping on occasion, who might have tried the game under his login.

Another way of spotting input anomalies was through lateral touch bias. While right-handed input had a consistent rightward bias left-handed entry showed a leftward bias. Most patients with

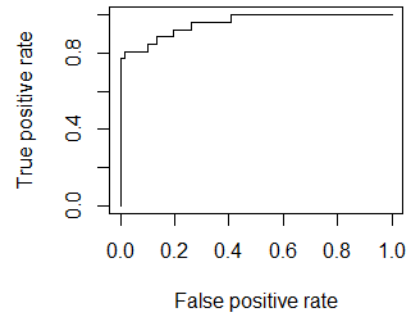


Figure 4: ROC curve of the neglect classifier

left-side lesions had to resort to using their left hand for input. With a simple classification based on positive or negative touch bias aggregate from a WAM session mapped to right and left-handed input we were able to correctly classify 95.1% of the 707 sessions' input hands. We excluded data from an ambidextrous participant P44 from this set.

During our initial testing we had observed a large number of 'unintentional' touch events, which were too far away from targets and typically clustered on the ipsi-lateral side of the input hand. We attributed these entries to the patients' hand making contact with the touch screen. We included a sharp sounding notification as feedback for players when touch input did not result in a hit. One patient pointed out that he found it difficult to keep his hand hovering above the tablet and that "Using a pen helps as to avoid unintentionally touching the screen" (P43).

6. DISCUSSION

We found no evidence that playing WAM in the self-administered amounts observed in the study had measurable effects on FIM_M, FIM_C, CBS_{obs}, CBS_{id} or SDMT. However, at 5.6 minutes on average per day - a fraction of the patients' supervised rehabilitative efforts (around 4 hours per day) - we should not expect to find measurable effects. Especially when the length of stay at the clinic, which correlated directly with supervised rehabilitation, had no significant effect on SDMT and CBS_{id} either. Nevertheless, 5.6 minutes training per day yielded statistically significant performance gains for the treatment compared to the control group.

We found that the WAM performance indicators *Center of Hit* and Fitts' law's b components had high predictive accuracy for

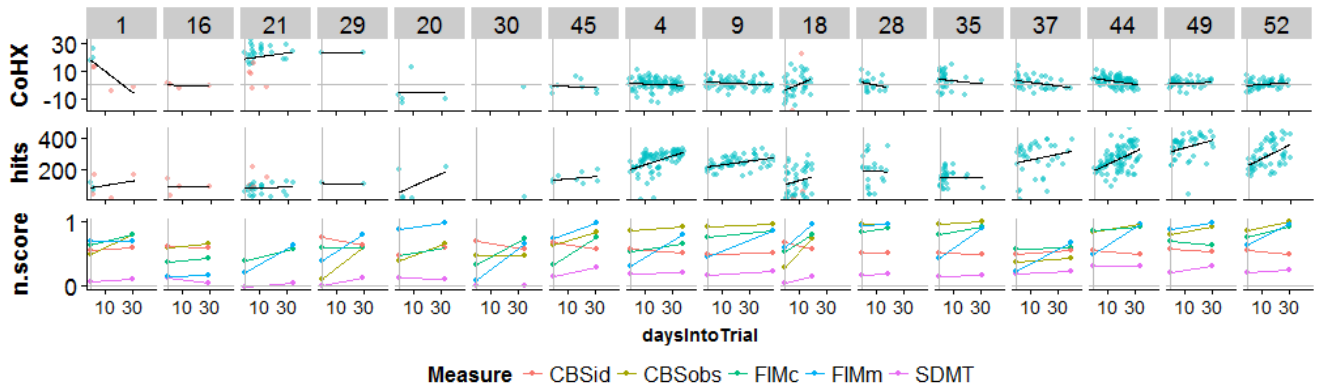


Figure 3: Per session performance aggregates and their linear trend lines of all USN patients (1, 16, 21, 29), all attention deficit patients (20, 30, 45) in the treatment group and a few examples (4-52) of other participants over time. Top: Center of Hit along the x-axis (CoHx), Middle: number of hits; Bottom: normalized scores of control tests

Data points in red denote session data that was misclassified (false positive or false negatives) by our regression model.

neglect classification. But the model was not sensitive enough to detect the mild neglect case of N16, who did not test positively on LiBi, LiCcl, and LetCcl either. While earlier research found that interaction with this type of rapid touch interaction games yielded Fitts law data that allowed comparisons with healthy people we found that in many cases sessions yielded negative Fitts *b* values, meaning that hitting targets further away took (after an initial reaction time) less time than targets closer by indicating a longer initial response time. Still, this model component was a significant predictor of neglect as it summarizes the temporal performance of hitting targets along with Fitts' *a* component.

Regardless of having neglect or not the patients found playing WAM initially *beneficial, easy to use and fun*. Whether the drop in perceived *benefit* from entry to exit in the treatment group was due to the game becoming *easier to use* and less challenging (*c.f.* [4]), disappointment in experienced vs. expected training gains, or less benefit from the game at a later stage in rehabilitation remained unclear. Participants found WAM helpful for their ability to concentrate. We had not explicitly designed the game for this but their beliefs matched the resident psychologist's assessment of WAM – requiring sustained attention - and the statistical analysis of game performance that compared the treatment to the control group. That we did not find significant effects of playing WAM on SDMT could be due to the fact that the scanning and tracking attention measured by the SDMT test was too dissimilar to the attention required for better performance in and trained by WAM.

The participants competed with themselves and some of them with one another. To this end and for better motivation and documentation of progress our participants sought access to their performance history, which matches previous findings in rehab applications [2,7,35]. In a shorter ten day pilot trial [18] with four participants (two neglect and two attention deficit patients) this concern had not emerged.

The clinical staff welcomed this form of self-initiated and administered rehabilitation. But responding to the alarms appropriately and to use them to start self-training had to be learned and in some cases assisted several times both verbally and in more difficult cases physically. Future studies and interventions need to budget for these activities. The therapist found WAM useful for bedside assessment and as a neutral reference point providing tool to illustrate the patient their weaknesses to improve their insight. The ability to infer input hand from WAM data was of particular interest to the therapist who is often confronted with or worried about non-compliance in exercises targeting the patient's weaker side. Overall, however, clinical staff had hoped the patients would make more use of the app and keep to the suggested three sessions per day for an overall involvement of 30 minutes per day (including retrieving the tablet and reviewing the results). But given differences in interest and what people find motivating we should not assume that WAM will be a good fit for all people to self-rehabilitate [4]. In our case WAM usage was not related to gender, age, mobile use, perceived benefit, ease of use, or fun.

Rather than having a fixed time limit, the game should adapt the length based on the user's ability to concentrate. Given the large fluctuations in game performance future work needs to address how to measure concentration and the effort the patients put into a game, build multi-dimensional outlier detection, and better tune the challenge in an elastic way for each participant depending on, e.g. time of day (*c.f.* [4]) and specifically for severely impaired and high performers. This was especially important for sessions in

which participants hit only very few targets, which increased the likelihood for a misclassification.

Unlike typical randomized controlled trials our study could not control the exposure to the game posing a threat to internal validity due to self-selection bias. Controlling exposure, however, would go against the study's aim of investigating rehabilitative gains in self-initiated self-rehabilitation and the current setup allowed for disambiguating WAM training gains from rehabilitative gains.

Controlling for activities the control group engaged in would have been helpful but was beyond the budget. Running the study on the clinical side already required roughly two months more of clinical staff time than what had been budgeted for. This was due to more time required for signing up participants (e.g. relatives and or patients repeatedly wanted to know more information), retrieving the patients for enrolment, periodical, and exit tests, and helping with initial problems or changing schedules.

7. CONCLUSION

We turned an understood neuropsychological measuring concept (Center of Hit) in neglect quantification into a game, which was simple enough to be played by all but one participant, showed a higher sensitivity range than, e.g. the SDMT test, and allowed patients to realize and become aware of performance gains. Playing the game did not result in measurable gains in SDMT, CBS or FIM measures but the patients ascribed in game performance improvements to concentration training gains from the game. Playing the game did not harm the participants; nor did the physical setup conflict with the clinical routine life apart from the auditory alarms. We found potential for WAM and similar solutions for bedside assessment, insight support, and providing knowledge and documentation of performance and its progress as motivation for rehabilitation activities.

8. ACKNOWLEDGMENTS

We thank the patients and staff at Brønderslev Neurorehabiliteringscenter for participating, Uffe Seilman for assessing the attention training potentials of WAM, Torben Christensen at Ideklinikken and patient@home for supporting the study, and the anonymous reviewers for providing valuable comments to improve the final version of the paper.

9. REFERENCES

- [1] Alankus, G., Lazar, A., May, M. and Kelleher, C. 2010. Towards customizable games for stroke rehabilitation. *Proc of CHI'10*, ACM, 2113–2122.
- [2] Ayoade, M. and Baillie, L. 2014. A novel knee rehabilitation system for the home. ACM Press, 2521–2530. DOI= <http://doi.org/10.1145/2556288.2557353>
- [3] Balaam, M., Rennick Egglestone, S., Hughes, A.-M., et al. 2010. Rehabilitation centred design. ACM Press, 4583. DOI= <http://doi.org/10.1145/1753846.1754197>
- [4] Balaam, M., Ricketts, I., Mawson, S., et al. 2011. Motivating mobility: designing for lived motivation in stroke rehabilitation. ACM Press, 3073. DOI= <http://doi.org/10.1145/1978942.1979397>
- [5] Bergego, C., Azouvi, P., Samuel, C., et al. 1995. Validation d'une échelle d'évaluation fonctionnelle de l'héminégligence dans la vie quotidienne: l'échelle CB. *Annales de réadaptation et de médecine physique*, Elsevier, 183–189.
- [6] Bloom, F.E., Lazerson, A., Hofstadter, L. and others. 1988. *Brain, mind, and behavior*. Freeman New York.
- [7] Burke, J.W., McNeill, M.D.J., Charles, D.K., Morrow, P.J., Crosbie, J.H. and McDonough, S.M. 2009. Optimising

- engagement for stroke rehabilitation using serious games. *International Journal of Computer Graphics: The Visual Computer* 25, 12: 1085–1099.
- [8] Buxbaum, L.J., Ferraro, M.K., Veramonti, T., et al. 2004. Hemispatial neglect Subtypes, neuroanatomy, and disability. *Neurology* 62, 5: 749–756. DOI= <http://doi.org/10.1212/01.WNL.0000113730.73031.F4>
- [9] Chow, P., Chen, C., Cheong, A., et al. 2014. Factors and Trade-Offs With Rehabilitation Effectiveness and Efficiency in Newly Disabled Older Persons. *Archives of physical medicine and rehabilitation* 95, 8: 1510–1520.
- [10] Dalmaijer, E.S., Stigchel, S.V. der, Nijboer, T.C.W., Cornelissen, T.H.W. and Husain, M. 2014. CancellationTools: All-in-one software for administration and analysis of cancellation tasks. *Behavior Research Methods*: 1–11. DOI= <http://doi.org/10.3758/s13428-014-0522-7>
- [11] Flores, E., Tobon, G., Cavallaro, E., Cavallaro, F.I., Perry, J.C. and Keller, T. 2008. Improving patient motivation in game development for motor deficit rehabilitation. ACM Press, 381. DOI= <http://doi.org/10.1145/1501750.1501839>
- [12] Forer, S., Granger, C.V. and others. 1987. Functional independence measure. *Buffalo, The Buffalo General Hospital State University of New York at Buffalo*.
- [13] Graziadio, S., Davison, R., Shalabi, K., et al. 2014. Bespoke video games to provide early response markers to identify the optimal strategies for maximizing rehabilitation. ACM Press, 20–24. DOI= <http://doi.org/10.1145/2554850.2554953>
- [14] Halligan, P.W., Cockburn, J. and Wilson, B.A. 1991. The behavioural assessment of visual neglect. *Neuropsychological Rehabilitation* 1, 1: 5–32. DOI= <http://doi.org/10.1080/09602019108401377>
- [15] Hernandez, H.A., Ye, Z., Graham, T.C.N., Fehlings, D. and Switzer, L. 2013. Designing Action-based Exergames for Children with Cerebral Palsy. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1261–1270. DOI= <http://doi.org/10.1145/2470654.2466164>
- [16] Jamieson, M., McGee-Lennon, M., Cullen, B., Brewster, S. and Evans, J. 2015. Issues Influencing the Uptake of Smartphone Reminder Apps for People with Acquired Brain Injury. *Proc. of SIGACCESS'15*, ACM, 339–340. DOI= <http://doi.org/10.1145/2700648.2811368>
- [17] Jehkonen, M., Ahonen, J.-P., Dastidar, P., Koivisto, A.-M., Laippala, P. and Vilkki, J. 1998. How to detect visual neglect in acute stroke. *The Lancet* 351, 9104: 727–728. DOI= [http://doi.org/10.1016/S0140-6736\(05\)78497-X](http://doi.org/10.1016/S0140-6736(05)78497-X)
- [18] Knoche, H., Hald, K., Tamsen, D. and Holm Jespersen, L. 2015. Tracking rehabilitative progress with Fitts and starts: Performance measures in a tablet game for hemi-spatial neglect patients. *Proc. Pervasive Health'15*, IEEE, 217–220. DOI= <http://doi.org/10.4108/icst.pervasivehealth.2015.259066>
- [19] Koh, G.C.-H., Chen, C.H., Petrella, R. and Thind, A. 2013. Rehabilitation impact indices and their independent predictors: a systematic review. *BMJ Open* 3, 9: e003483. DOI= <http://doi.org/10.1136/bmjopen-2013-003483>
- [20] Lezak, M.D. 2004. *Neuropsychological assessment*. Oxford University Press.
- [21] Linehan, C., Waddington, J., Hodgson, T.L., Hicks, K. and Banks, R. 2014. Designing games for the rehabilitation of functional vision for children with cerebral visual impairment. ACM Press, 1207–1212. DOI= <http://doi.org/10.1145/2559206.2581219>
- [22] Mattingley, J.B., Husain, M., Rorden, C., Kennard, C. and Driver, J. 1998. Motor role of human inferior parietal lobe revealed in unilateral neglect patients. *Nature* 392, 6672: 179–182. DOI= <http://doi.org/10.1038/32413>
- [23] Millán, M. and Dávalos, A. 2006. The Need for New Therapies for Acute Ischaemic Stroke. *Cerebrovascular Diseases* 22, 1: 3–9. DOI= <http://doi.org/10.1159/000092327>
- [24] Patient@home. Retrieved from <http://www.en.patientathome.dk/>
- [25] Rabosky, D.L., Grundle, M., Anderson, C., et al. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5, 7: 701–707.
- [26] Rabuffetti, M., Farina, E., Alberoni, M., et al. 2012. Spatio-temporal features of visual exploration in unilaterally brain-damaged subjects with or without neglect: results from a touchscreen test. *PLoS one* 7, 2: e31511.
- [27] Reitan, R.M. and Wolfson, D. 1985. *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation*. Reitan Neuropsychology.
- [28] Robertson, I.H., Mattingley, J.B., Rorden, C. and Driver, J. 1998. Phasic alerting of neglect patients overcomes their spatial deficit in visual awareness. *Nature* 395, 6698: 169–172. DOI= <http://doi.org/10.1038/25993>
- [29] Romano, J.G., Schulz, P., Kenkel, S. and Todd, D.P. 2008. Visual field changes after a rehabilitation intervention: Vision restoration therapy. *Journal of the neurological sciences* 273, 1: 70–74.
- [30] Rorden, C. and Karnath, H.-O. 2010. A simple measure of neglect severity. *Neuropsychologia* 48, 9: 2758–2763. DOI= <http://doi.org/10.1016/j.neuropsychologia.2010.04.018>
- [31] Ryan, R.M. and Deci, E.L. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 1: 68–78. DOI= <http://doi.org/10.1037/0003-066X.55.1.68>
- [32] Schenkenberg, T., Bradford, D.C. and Ajax, E.T. 1980. Line bisection and unilateral visual neglect in patients with neurologic impairment. *Neurology* 30, 5: 509–509. DOI= <http://doi.org/10.1212/WNL.30.5.509>
- [33] Silverstein, S.M., Light, G. and Palumbo, D.R. 1998. The Sustained Attention Test: a measure of attentional disturbance. *Computers in Human Behavior* 14, 3: 463–475. DOI= [http://doi.org/10.1016/S0747-5632\(98\)00017-X](http://doi.org/10.1016/S0747-5632(98)00017-X)
- [34] Strauss, E., Sherman, E.M. and Spreen, O. 2006. *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press, USA.
- [35] Uzor, S., Baillie, L., Skelton, D. and Fairlie, F. 2011. Identifying barriers to effective user interaction with rehabilitation tools in the home. In *Human-Computer Interaction-INTERACT 2011*. Springer Berlin Heidelberg, 36–43.
- [36] Vanier, M., Gauthier, L., Lambert, J., et al. 1990. Evaluation of left visuospatial neglect: norms and discrimination power of two tests. *Neuropsychology* 4, 2: 87.
- [37] Wade, D.T., Wood, V.A. and Hower, R.L. 1985. Recovery after stroke--the first 3 months. *Journal of Neurology, Neurosurgery & Psychiatry* 48, 1: 7–13. DOI= <http://doi.org/10.1136/jnnp.48.1.7>