The influence of computer-based cognitive flexibility training on subjective cognitive well-being after stroke: A multi-center randomized controlled trial

Renate M. van de Ven1,*, Jaap M. J. Murre1, Jessika I. V. Buitenweg1, Dick J. Veltman2, Justine A. Aaronson3, Tanja C. W. Nijboer4,5, Suzanne J. C. Kruiper-Doesborgh6, Coen A. M. van Bennekom3, K. Richard Ridderinkhof1,7, Ben Schmand1,8

1 Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands, 2 Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands, 3 Department of Research and Development, Heliomare Rehabilitation Center, Wijk aan Zee, The Netherlands, 4 Center of Excellence in Rehabilitation Medicine, Brain Center Rudolf Magnus, University Medical Center Utrecht, and De Hoogstraat Rehabilitation, Utrecht, The Netherlands, 5 Department of Experimental Psychology, Utrecht University, Utrecht, the Netherlands, 6 Amsterdam Rehabilitation Research Center, Reade, Amsterdam, The Netherlands, 7 Amsterdam Brain & Cognition (ABC), University of Amsterdam, Amsterdam, The Netherlands, 8 Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

* r.m.vandeven@uva.nl

Abstract

Background

Stroke can result in cognitive complaints that can have a large impact on quality of life long after its occurrence. A number of computer-based training programs have been developed with the aim to improve cognitive functioning. Most studies investigating their efficacy used only objective outcome measures, whereas a reduction of subjective cognitive complaints may be equally important for improving quality of life. A few studies used subjective outcome measures but were inconclusive, partly due to methodological shortcomings such as lack of proper active and passive control groups.

Objective

The aim of the current study was to investigate whether computer-based cognitive flexibility training can improve subjective cognitive functioning and quality of life after stroke.

Methods

We performed a randomized controlled double blind trial (RCT). Adults (30–80 years old) who had a stroke 3 months to 5 years ago, were randomly assigned to either an intervention group (n = 38), an active control group (i.e., mock training; n = 35), or a waiting list control group (n = 24). The intervention and mock training consisted of 58 half-hour sessions within 12 weeks. The primary subjective outcome measures were cognitive functioning (Cognitive Failure Questionnaire), executive functioning (Dysexecutive Functioning Questionnaire), and quality of life (Stroke Impact Scale).
quality of life (Short Form Health Survey), instrumental activities of daily living (IADL; Lawton & Brody IADL scale), and participation in society (Utrecht Scale for Evaluation of Rehabilitation-Participation). Secondary subjective outcome measures were recovery after stroke, depressive symptoms (Hospital Anxiety Depression Scale—depression subscale), fatigue (Checklist Individual Strength—Fatigue subscale), and subjective cognitive improvement (exit list). Finally, a proxy of the participant rated the training effects in subjective cognitive functioning, subjective executive functioning, and IADL.

Results and conclusions
All groups improved on the two measures of subjective cognitive functioning and subjective executive functioning, but not on the other measures. These cognitive and executive improvements remained stable 4 weeks after training completion. However, the intervention group did not improve more than the two control groups. This suggests that improvement was due to training-unspecific effects. The proxies did not report any improvements. We, therefore, conclude that the computer-based cognitive flexibility training did not improve subjective cognitive functioning or quality of life after stroke.

Introduction
Up to 92% of stroke survivors report cognitive complaints in, for example, executive functioning, attention, memory, and processing speed [1]. These subjective complaints can be long lasting [2] and have been associated with lower return to work rates [3], higher mortality risk [4], worse cognitive functioning as measured by neuropsychological tasks, and depressive symptoms [5]. People who had a stroke reported lower quality of life compared to the general population [6, 7] and lower participation in social, vocational and leisure activities [8]. Improvement of emotional well-being during rehabilitation contributed to better health-related quality of life [9]. To foster this, it is important to improve subjective functioning.

Studies that investigate the efficacy of cognitive training typically focus on objective outcome measures. These, however, are not always in agreement with subjective measures [10]. In only about half of the studies a positive relationship between objective and subjective functioning in stroke patients was found [5]. Apparently, objective measures collected in the lab or in assessment rooms do not seem to reflect performance in daily living as perceived by the individual or their surroundings. It is, therefore, important to use both objective and subjective measures.

Impaired self-awareness of cognitive functioning after stroke may, however, make ratings of subjective functioning unreliable. To counteract this, subjective functioning can additionally be measured by asking a proxy of the stroke survivor. Ratings of subjective functioning have been found to differ significantly between stroke survivors and their proxies [11], although agreement between proxy and stroke survivor may be high for certain subjective measures, such as activities of daily living, and moderate to high for cognitive complaints [5] and quality of life ratings [12].

Efficacy studies that used subjective measures have been inconclusive. Based on a systematic review, computer-based functional retraining of executive functioning was found to result in improvements on several, but not all, subjective measures [13]. To summarize the results of the review, improvements were seen in subjective measures of symptom severity [14],

funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: CFQ, Cognitive Failure Questionnaire; CIS-F, Checklist Individual Strength-Fatigue; DEX, Dysexecutive Functioning Questionnaire; HADS-D, Hospital Anxiety Depression Scale—depression subscale; IADL, Instrumental Activities of Daily Living; SF, Short Form Health Survey; TAPASS, Training Project Amsterdam Seniors and Stroke; TICS, Telephone Interview for Cognitive Status; USER-P, Utrecht Scale for Evaluation of Rehabilitation-Participation; VAS, Visual Analog Scale.
attention [15–17], cognitive functioning [18], participation in social activities [19], and fatigue [15, 20]. No improvements were seen in life satisfaction and self-reported health index [21, 22]. Most studies did not show improvement on measures of depressive symptoms [14, 17, 23, 24], except for one study [25]. Subjective executive functioning improvements were seen after a general cognition training [19], but not after a working memory training [24]. The results were also mixed for subjective quality of life [14, 19] and subjective measures of Instrumental Activities of Daily Living [14, 25]. In the two studies that included proxies, proxies and participants agreed on a positive effect of the training of executive functioning and working memory, but they disagreed on a training effect for attention, as the proxy noted improvement after training that was not reported by the participant [17, 19].

Although these results seem promising, it is not possible to draw firm conclusions that cognitive training improves subjective functioning, because most studies lack proper control groups. Control groups are especially important with respect to subjective measures, because they are more sensitive to placebo effects than objective measures [26]. In the only study that included an active control group, subjective improvements were also seen in this group [19]. Further limitations were that most studies included only a small sample, training duration was generally short, and subjective functioning was commonly assessed with only one outcome measure. Inconsistent findings may also be explained because studies frequently used different outcome measures.

Due to the significance of subjective functioning and quality of life ratings of stroke survivors, we aimed to evaluate the effects of a computer-based cognitive flexibility training on subjective functioning after stroke while accounting for the above-reviewed methodological issues. The training consisted of 58 half-hour sessions focusing on attention, memory, and reasoning. Participants trained at home during 12 weeks. Stroke patients who received this training were compared to an active control group (which received a mock training) and to a waiting list control group. Cognitive flexibility is an executive function that is essential for many everyday life tasks. We, thus, expected that cognitive complaints would be reduced, and that societal participation and activities of daily living would increase to a greater degree after the computer-based training than after the mock training, and that there would be no change in the waiting list group.

Materials and methods

A detailed description of the design, training tasks, and outcome measures of this study has been published previously [27, 28].

Participants

Individuals who had a stroke 3 months to 5 years ago and were between 30 and 80 years old were recruited via Dutch rehabilitation centers and patient societies (April 2013—March 2015; last follow-up measurement in November 2015). Participants were selected who had cognitive impairments as testified by medical records, still had cognitive complaints at study entry, and were able to work with a computer. Individuals who had any disease other than stroke that results in severe cognitive impairments, had a history of substance abuse or addiction, or were incapable of executing the training or outcome measure tasks were excluded from the study (see [27] for a full description of the criteria). A schematic overview of the participant flow can be found in Fig 1. Participants included in this study are the same as those in Van de Ven et al (2017) [28].

A-priori sample size calculation suggested a sample size of at least 60 (3x 20) to be able to reveal large effects ($d = 0.80$) in univariate analyses with two groups. Based on an effect size of
Fig 1. Consolidated Standards of Reporting Trials (CONSORT) flow diagram. T2 = post-training; T3 = 4 weeks after training completion.

https://doi.org/10.1371/journal.pone.0187582.g001
$d = 0.56$ that was found in a previous training study [29] we strived for a sample of 120 (3 x 40) participants. With the expected attrition rate of 15%, we planned to stop inclusion at 138 participants or when the recruitment period was over.

Experimental design

The study was a prospective multicenter, double blind, randomized controlled study (RCT). Participants were randomly assigned evenly to one of three groups (intervention, mock training, waiting list) by software [30] that minimized the group differences in age, education, sex, time since stroke, scores on a cognitive screening [31], and level of computer experience. The minimization technique included elements of randomness into the minimization algorithm, to make the prediction to which group a certain participant would be allocated uncertain [32]. Other factors such as rehabilitation site were not considered during the randomization procedure. The groups were coded by the research coordinator. The participants and their proxies were not informed that one of the two training programs was considered to be a mock training, but they were told that we compared the effects of two training programs. The waiting list group was added during the course of the study, but participants before and after this addition did not differ on baseline characteristics.

Several questionnaires were administered online: at baseline (T0), after 6 weeks of training or waiting (T1), after training completion for the training groups or after 12 weeks of waiting for waiting list group (T2), and (for the training groups) 4 weeks after training completion (T3). Participants rated their impression of general recovery on a visual analogue scale (VAS) at T0 and T2. Neuropsychological assessment and MRI scanning were executed at T0 and T2, results of which will be reported elsewhere (see [28] for the results of the objective outcome measures). In compliance with the Helsinki Declaration, the study was approved by the ethical review board of the University of Amsterdam (i.e., Commissie Ethiek voor de Afdeling Psychologie; approved 17th of December 2012) and by the medical ethical review board of the VU University Medical Center, Amsterdam (i.e., the Medisch Ethische Toetsingscommissie Vrije Universiteit Medisch Centrum; approved 9th of July 2013, amendment approved 23rd of May 2014). The study was registered before study commencement as Training Project Amsterdam Seniors and Stroke (TAPASS) with the Central Committee on Research Involving Human Subjects Register NL4468502913 (www.toetsingonline.nl). Additionally, to fulfill the World Health Organization Registry criteria it was also registered with the Netherlands National Trial Register NTR5174. The authors confirm that all ongoing and related trials for this intervention are registered.

Procedures

Participants were asked to select a proxy who was willing to complete questionnaires. The participant came to the University of Amsterdam to receive detailed information about the study, to provide informed consent, to undergo a neuropsychological assessment, and to rate their general recovery on the VAS (T0). The online questionnaires were completed at home by both the participant and the proxy. Depending on the group to which the participant was allocated, they either started training or waited for 12 weeks. The training was performed five times per week and consisted of a total of 58 half-an-hour sessions. Participants of the two training groups were contacted by phone by a neuropsychologist once every week or every two weeks to ask about their training experience and were sent an email as soon as they did not train for two days. The waiting list group was not contacted at all during this period, but received care as usual (which most often meant they did not receive any treatment during the study). After 12 weeks, both the participant and the proxy completed the questionnaires again (T2).
subset of the questionnaires was also completed by the participant after 6 weeks of training or waiting (T1) and in the training groups 4 weeks after training completion (T3). The waiting list group started the intervention training after they completed the questionnaires after the 12 weeks of waiting (T2).

**Intervention materials**

Both the intervention and the mock training were carried out by means of a professionally programmed website (www.braingymmer.com) tailored to older adults as well as stroke survivors. Each session lasted 30 minutes in which the tasks of that day were presented in a predefined order and feedback was provided immediately after each task (based on a three-star rating scale) and at the end of each session. The participants were not aware of the training protocol and tasks of the other intervention group. The cognitive flexibility training consisted of nine tasks in the domains of attention, memory, and reasoning. A new task of a different domain was presented after 3 minutes. With this fast task switching we aimed to trigger cognitive flexibility. Thus, each session contained 10 tasks. The difficulty of each task was adapted to the participant such that a higher level became available when one out of three stars was achieved; participants had been instructed to go to a higher level whenever two or three out of three tasks were obtained. The mock training consisted of four tasks which we considered to train executive functioning only minimally. Participants trained for 10 minutes per task and thus carried out three tasks per session. Task difficulty was not adaptive because participants were instructed to train at a constant level for one to two weeks and could only move to a higher level after this predetermined period. However, participants could (and occasionally did) move to higher levels—thereby disobeying the instructions—as these levels were still accessible.

**Outcome measures**

**Primary outcomes.** Subjective cognitive functioning was assessed with the Cognitive Failure Questionnaire [33], which consists of 25 questions completed on a five-point scale with a maximum total score of 100. The CFQ has good psychometric properties including test-retest reliability [34]. It was sensitive to change of a computerized working memory training for stroke survivors [18]. In addition, we created a proxy version that consisted of the same 25 questions asked to the participant.

The Dysexecutive Functioning Questionnaire [35] was used to measure subjective executive functioning. It consists of 20 questions on a five-point scale and the maximum total score is 80. DEX scores could be used to distinguish stroke patients with executive impairments from healthy individuals [36]. In addition, the proxy version of the DEX was used.

Subjective independence level in IADL were measured by the Lawton & Brody IADL scale [37]. The answers on the eight questions on a four- to six-point scale were added up to a total score that ranged from 0 (no problems at all) to maximum 22 (highly impaired). Questions that were answered with ‘not applicable’ or ‘never carried out myself in my life’, were replaced with the average of the remaining items completed by the participant. The IADL scale has good psychometric properties including test-retest reliability [38, 39]. The IADL scale has been sensitive to change after constraint-induced movement therapy in stroke survivors [40]. Furthermore, cognitive impairment before enrollment in stroke rehabilitation could predict IADL scores at 6 months follow-up [41]. A proxy version was created that consisted of the same eight questions asked to the participant.

Subjective quality of life was assessed by the Short Form Health Survey (SF-36) [42]. Total score of mental health was converted to a z-score and corrected for age and sex based on a
Dutch norm group [43]. The SF-36 has shown to be valid and reliable in a general population and chronic disease populations in the Netherlands [43].

Rating of participation in society was measured with the Utrecht Scale for Evaluation of Rehabilitation-Participation [44]. It is a 31 items questionnaire that assesses the frequency, restrictions, and satisfaction with participation in social, vocational, and leisure activities. Answers are given on a four- to six-point scale and were averaged and transformed to a 0–100 scale for the three subscales. The answers ‘not applicable’ were not used in the average. As the scores on the three subscales cannot be combined into a total score, we used the restriction subscale as the outcome measure. The USER-P had a good test-retest reliability in a sample that included stroke patients [45].

Scores of all questionnaires were recoded such that higher scores represent better performance. All questionnaires are commonly used in the stroke population for clinical evaluation and science.

Secondary outcomes. Subjective recovery after stroke was rated on a 100 mm VAS ranging from 0 (“I did not recover at all”) to 100 (“I recovered completely”).

Subjective cognitive improvement after training or the waiting period was assessed with a four-items exit questionnaire designed for this study. Improvements in overall cognition, attention, memory, and reasoning were rated on a five-point scale ranging from 1 (“I definitely did not improve”) to 5 (“I definitely improved”). Total score was based on the summation of these answers.

Depressive symptoms were assessed with the Hospital Anxiety Depression Scale - depression subscale [46] with a maximum score of 21. The HADS has shown to be a valid measure of emotional distress in people with traumatic brain injury [47].

Subjective level of fatigue was assessed with the Checklist Individual Strength- Fatigue subscale [48] which has a total score between 8 and 56. The test-retest reliability was high for the fatigue subscale in a general population of the Netherlands[49].

Again, scores were recoded such that higher scores represent better performance.

Training performance. Performance on the intervention training task was reflected by the levels and scores obtained in the domains attention, memory, and reasoning. The average was taken for the performance on the three training tasks that belonged to the same domain. The maximum total score for each domain was 2000. The mock training tasks did not belong to separate domains and thus the total score was based on the average of the four tasks, and the maximum total score was 900 (see [28] for detailed description of the training performance score).

Statistical analysis

The main analyses were performed on the data of all participants who started the study (i.e., intention to treat analyses). A repeated-measures MANOVA was performed on the total score on CFQ, total score on DEX, USER-P restriction score, SF-36 mental composite z-score, and total IADL score as dependent variables. Group (intervention, mock training, and waiting list control group) was the independent variable and time-points were before and after the 12 weeks waiting or training period (T0 and T2). In case of a significant time * group interaction effect, post-hoc univariate ANOVAs were performed on the difference score (T2-T0). The analyses were rerun with education, age, and time since stroke as covariates. The IADL scores were not normally distributed thus a Kruskal-Wallis test was used to confirm the outcome of the repeated-measures MANOVA.

Secondary analyses were performed in a similar way with scores on HADS-D, CIS-F, and recovery VAS as dependent variables. The exit questionnaire was only administered after
training completion. The total score was not normally distributed; thus the Kruskal-Wallis test was performed to evaluate group differences. The proxy version of the CFQ, DEX, and IADL were analyzed in the same way as the patient version.

In the training groups, Pearson’s correlation coefficients were computed to examine the relation between improvement on the training tasks and change (T2-T0) in the outcome measures. To determine the long-term effects of the training, a repeated-measures MANOVA was performed with CFQ and DEX as dependent variables; group (intervention versus active control) as independent variable; and T0, T2, and T3 as time-points. Post-hoc univariate ANOVAs were performed when the time effect was significant. Analyses were repeated with age, education, and time since stroke as covariates. Exploratory univariate ANOVAs were performed with difference score (T2–T0) from all outcome measures and group as independent variable.

Missing values were replaced by the method of last observation carried forward (or backward in case the baseline score was missing). In this way, 13.1% of the intention-to-treat data were imputed. Results reported are with outliers because there were no reasons to expect that extreme values were not a reflection of true scores of impaired participants. Analyses were rerun without outliers and whenever results differed they are reported. In these analyses, outliers in the (transformed) raw data were detected by Grubbs’ Extreme Studentized Deviation test [50] and were replaced with the nearest value of another participant in the overall sample. The main analyses were repeated with the participants who completed all follow-up measures and completed at least 50 training sessions (i.e., per protocol analyses).

The scores on the DEX, SF-36, USER-P, HADS, and recovery VAS were transformed because these scores were not normally distributed (see S1 File for formulas). All analyses were performed using SPSS version 22 or higher. P-values <.05 (two-tailed if not mentioned otherwise) were considered significant.

Results

Pre-training

After the recruitment period was over, 97 of the 223 potential participants who were screened, met all inclusion criteria and were included in the final analyses (see Fig 1 for participant flow-chart including drop-out reasons). At baseline, the three groups did not differ significantly in clinical and demographical variables (Table 1), except for fatigue where the intervention group reported higher levels of fatigue than the active control group (p = .02). The proxy reports on cognitive and dysexecutive functioning and independence in IADL did not differ significantly between the three groups.

Transfer effect of training

Subjective cognitive functioning (primary outcome measures). The reports by the participants on primary outcome measures revealed a significant time effect (F(5,90) = 64.44, p = .001, with partial eta squared effect size (η²p) = .20; see Table 2). There was no group * time interaction (F(10,182) = 1.35, p = .21, η²p = .07). Post-hoc univariate analyses revealed that the time effect was significant for CFQ (p < .001, η²p = .18) and DEX (p < .01, η²p = .08). Thus, all three groups improved in subjective cognitive and executive functioning, including the waiting list group. The time effect disappeared after correcting for age, education level, and time since stroke (F(5,87) = 1.30, p = .27, η²p = .07). However, these variables did not significantly explain any variance, suggesting that the model without these covariates is more valid. In view of the non-normal distribution of IADL scores, we reran the repeated-measures MANOVA without IADL and analyzed IADL with the non-parametric Kruskal-Wallis test, but this did not change the results.
Other subjective functioning (secondary outcome measures). Results of the secondary outcome measures did not show a significant time effect ($F(3,92) = 0.45, p = .72, \eta^2_p = .02$), or a time × group interaction ($F(6,186) = 0.40, p = .89, \eta^2_p = .01$; see Table 2). Thus, none of the groups improved with respect to depressive symptoms, fatigue, or recovery level. The overall cognitive improvement reported in the exit questionnaire did not differ significantly between groups either ($H = 3.07, p = .22$).

Proxy reports. Participants who had a proxy report on at least one of the two time-point were included in the analyses ($n_{intervention} = 30, n_{active control} = 31, n_{waiting list} = 19$). The above-mentioned significant time effect seen in the participants report was not replicated in the proxy ratings ($F(3,75) = 0.37, p = .77, \eta^2_p = .02$; see Table 2) and the group × time interaction remained non-significant ($F(6,152) = 1.38, p = .23, \eta^2_p = .05$).

Relation between improvement on training task and outcome measures. Difference from baseline (T0) to end of training (T2) in the above-mentioned subjective outcome measures was compared to improvement in training tasks in the two training groups. Thirty-six participants who started the intervention training and 33 who started the mock training were included in these comparisons. Even though participants clearly improved on training tasks,
### Table 2. Mean (standard deviation) and MANOVA of the outcome measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>Pre-training</th>
<th>Post-training</th>
<th>Δd</th>
<th>Fp-value</th>
<th>ηp²</th>
<th>d</th>
<th>Fp-value</th>
<th>ηp²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFQ</td>
<td>Intervention group (n = 38)</td>
<td>34.2 (13.8)</td>
<td>31.2 (13.8)</td>
<td>-3.0</td>
<td>0.1</td>
<td>0.84</td>
<td>-3.0</td>
<td>0.1</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Active control group (n = 35)</td>
<td>32.4 (13.6)</td>
<td>29.6 (13.6)</td>
<td>-2.8</td>
<td>0.1</td>
<td>0.79</td>
<td>-2.8</td>
<td>0.1</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Waiting list group (n = 24)</td>
<td>32.3 (13.5)</td>
<td>29.4 (13.5)</td>
<td>-2.9</td>
<td>0.1</td>
<td>0.81</td>
<td>-2.9</td>
<td>0.1</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pre-training</td>
<td>34.2 (13.8)</td>
<td>31.2 (13.8)</td>
<td>-3.0</td>
<td>0.1</td>
<td>0.84</td>
<td>-3.0</td>
<td>0.1</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Post-training</td>
<td>32.4 (13.6)</td>
<td>29.6 (13.6)</td>
<td>-2.8</td>
<td>0.1</td>
<td>0.79</td>
<td>-2.8</td>
<td>0.1</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Δd</td>
<td>Pre-training</td>
<td>Post-training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recovery</td>
<td>57.7 (20.5)</td>
<td>56.1 (20.4)</td>
<td>-1.6</td>
<td>0.1</td>
<td>0.78</td>
<td>-1.6</td>
<td>0.1</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>VAS</td>
<td>57.7 (20.5)</td>
<td>56.1 (20.4)</td>
<td>-1.6</td>
<td>0.1</td>
<td>0.78</td>
<td>-1.6</td>
<td>0.1</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>CFQ</td>
<td>39.4 (11.7)</td>
<td>38.4 (11.6)</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>DEX</td>
<td>39.4 (11.7)</td>
<td>38.4 (11.6)</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>IADL</td>
<td>3.5 (3.1)</td>
<td>3.1 (3.0)</td>
<td>-0.4</td>
<td>0.1</td>
<td>0.70</td>
<td>-0.4</td>
<td>0.1</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>SF-36</td>
<td>-0.8 (1.1)</td>
<td>-0.8 (1.1)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.98</td>
<td>0.0</td>
<td>0.0</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>USER-P</td>
<td>75.3 (16.6)</td>
<td>74.0 (16.6)</td>
<td>-1.3</td>
<td>0.1</td>
<td>0.80</td>
<td>-1.3</td>
<td>0.1</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>57.7 (20.5)</td>
<td>56.1 (20.4)</td>
<td>-1.6</td>
<td>0.1</td>
<td>0.78</td>
<td>-1.6</td>
<td>0.1</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>CFQ</td>
<td>39.4 (11.7)</td>
<td>38.4 (11.6)</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>DEX</td>
<td>39.4 (11.7)</td>
<td>38.4 (11.6)</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
<td>-1.0</td>
<td>0.1</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>IADL</td>
<td>3.5 (3.1)</td>
<td>3.1 (3.0)</td>
<td>-0.4</td>
<td>0.1</td>
<td>0.70</td>
<td>-0.4</td>
<td>0.1</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>SF-36</td>
<td>-0.8 (1.1)</td>
<td>-0.8 (1.1)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.98</td>
<td>0.0</td>
<td>0.0</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>USER-P</td>
<td>75.3 (16.6)</td>
<td>74.0 (16.6)</td>
<td>-1.3</td>
<td>0.1</td>
<td>0.80</td>
<td>-1.3</td>
<td>0.1</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>13.6 (3.5)</td>
<td>13.6 (3.5)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.99</td>
<td>0.0</td>
<td>0.0</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Cognitive improvement</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>13.6 (3.5)</td>
<td>13.6 (3.5)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.99</td>
<td>0.0</td>
<td>0.0</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Proxy</td>
<td>13.6 (3.5)</td>
<td>13.6 (3.5)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.99</td>
<td>0.0</td>
<td>0.0</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Note: All scores are total scores where lower scores represent better performance except for SF-36, USER-P, cognitive improvement, and recovery VAS (mm) where higher score reflects better performance. Δ = difference score between pre- and post-measurement recoded in such a way that higher difference score represents improvement; d = Cohen's d (effect size); F was based on Pillai's Trace; H was based on Kruskal-Wallis test; Results were not affected by excluding outliers. Bold values are considered significant and survived Bonferroni-Holm adjustment where appropriate. ηp² = partial eta squared (effect size); CFQ = Cognitive failure questionnaire; DEX = Dysexecutive Functioning Questionnaire; IADL = Instrumental Activities of Daily Living; Fatigue subscale; HADS D = Hospital Anxiety Depression Scale—Depression; n.a. = not applicable; Proxy = analyses based on Kruskal-Wallis test; n = number of participants.
correlations between training improvement and change in subjective functioning were weak ($r$ ranging from -.28 to .31). Only the correlation of improvement in training tasks with improvement of cognitive functioning (CFQ) in the active control group was significant ($r = .31$, $p = .04$, one-tailed). Overall, the results of proxy reports were similar except for the correlation of improvement in training tasks with cognitive improvement (CFQ), which was weak and not significant ($r = -.06$, $p = .36$).

**Per-protocol analyses.** The main analyses were rerun without participants who dropped out before the T2 measurement (18 participants) or who completed less than 50 training sessions (one intervention and one active control participant). Analyses were based on 77 participants ($n_{\text{intervention}} = 28$, $n_{\text{active control}} = 29$, $n_{\text{waiting list}} = 20$). The 20 participants who dropped out or did not complete the training protocol were not significantly different in baseline subjective functioning from these 77 participants. The only exception was that they were slightly more restricted in their participation (USER-P: $t(95) = 1.97$, $p = .05$) than participants who followed the protocol.

Results from the per-protocol analyses of primary and secondary outcome measures were similar to the intention-to-treat analyses (see S1 Table), thus suggesting that participants who followed the study protocol did not improve more than those who did not.

**Follow-up.** Follow-up measurements were not performed after the waiting period. Consequently, the following analyses refer to the two training groups. Both groups improved significantly over time ($F(4, 68) = 5.85$, $p < .001$, $\eta^2_p = .26$) in the CFQ ($p < .001$, $\eta^2_p = .21$) and the DEX ($p < .01$, $\eta^2_p = .09$; see Fig 2). Post-hoc pairwise comparisons revealed that follow-up scores were significantly better than baseline scores (T0), but did not significantly differ from the immediate training effect (T2). Thus, scores remained stable after training completion.

There was no significant group $\times$ time interaction ($F(4,68) = 1.25$, $p = .30$, $\eta^2_p = .07$). Results from the per-protocol analyses were similar. The significant time effect disappeared after
correcting for age, education, and time since stroke. However, none of these covariates explained a significant amount of variance, suggesting that the analysis without covariates is more valid. There was no significant group difference in subjective cognitive improvement (based on the exit list) reported at follow-up \( (p = .06 \text{ in intention-to-treat and } p = .34 \text{ in the per-protocol analyses}) \).

**Exploratory analyses.** To confirm that the lack of training effect was not due to insufficient statistical power, we explored the data (post-hoc) with univariate ANOVAs on difference score \( (T2-T0) \). The only significant group difference was found for CFQ \( (F(2,94) = 3.31, p = .04, \eta^2 = .07) \) where the active control group improved significantly more than the intervention \( (p = .05) \) and waiting list group \( (p = .02) \). However, this group difference would not remain significant after adjustment for multiple testing. There were no significant group differences when both training groups were pooled and compared to the waiting list control group.

**Discussion**

The aim of this study was to examine the effects of computer-based cognitive flexibility training on subjective cognitive functioning after stroke. Results indicate that the computer training did not result in larger improvements than the effect of care as usual. Improvements in all three groups were seen only in subjective cognitive functioning (as measured with the CFQ) and subjective executive functioning (DEX) which remained stable 4 weeks after training completion. These improvements were most likely due to training-unspecific effects such as test-retest effects, placebo effects, or the Hawthorne effect (i.e., the effect of merely participating in a scientific study).

The results are in line with the objective cognitive improvements that were seen in all three groups \[28\]. Similar to the current report, a time effect was found on three out of five objective executive functioning measures and three out of seven objective cognitive domain scores. This suggests that both the objective and subjective measures were in agreement.

Our results replicate the findings by Spikman et al. (2010), who also found time effects on subjective executive functioning without a superior improvement of the intervention group compared to an active control group. Moreover, Ponsford et al. (1988) found improvements in subjective attention after a waiting period before the start of the training. Thus, training-unspecific factors such as spontaneous recovery and the Hawthorne effect may already result in positive changes.

The time effect on subjective cognitive functioning contrasts with a study using a similar training (though lasting only 8 weeks), which did not result in time or transfer effects \[51\]. In addition, the absent transfer effect to subjective cognitive functioning differs from the results of a working memory training were subjective cognitive functioning improved more in the experimental group than in passive control group \[52\]. We cannot rule out, however, that the apparent transfer effect is due to placebo effects, because the study lacked an active control group. Nevertheless, our exploratory univariate ANOVA replicated this latter finding. Namely, we found that subjective cognitive functioning in our active control group improved more than the waiting list group (and intervention group). This may be explained by the fact that our mock training was more adaptive than we had planned, due to participants who managed to achieve higher levels than was allowed based on the training protocol. Therefore, it may have been an effective training after all. However, the group difference in our study was non-significant after correction for multiple testing and should therefore be interpreted with caution.

The lack of improvement in the other questionnaires is in agreement with several studies that also failed to find improvements in IADL \[14\], depressive symptoms \[14, 17, 23, 24\], and...
health related quality of life [21–23, 51]. Spikman et al. (2010) did find improvements in reported quality of life, but this improvement did not differ between the intervention and the active control group. This suggests that effects on general mental health and quality of life are absent or nonspecific, such as placebo effects. Perhaps more frequent and or longer training duration is required to result in far transfer effects in the more general quality of life ratings.

De Luca et al. (2014) did find larger improvements of IADL and depressive symptoms in the intervention group than in the control group, which received care as usual. The training program used was not described in detail, thus it is unclear which elements could have resulted in their positive findings. De Luca et al. included participants who had a severe brain injury 3–6 months earlier, whereas our sample did not report many IADL impairments. Perhaps computer-based training is more effective in improving subjective functioning in a more severely affected population. However, the improvements in IADL found by De Luca et al. might also have been due to spontaneous recovery as improvements were also seen in the control group and their statistical comparison of group differences would not survive adjustment for multiple testing.

Proxies in our study did not report any improvements. This contrasts with the improvements found by Spikman et al. (2010) in both their treatment groups. They included participants with specific executive functioning complaints before the start of study, whereas we included participants with unspecified cognitive complaints. The participants and proxies in Spikman et al. indeed reported worse executive functioning at study entry than the proxies and participants in our study. Thus, more severely affected stroke participants might benefit more from the training. Because Spikman et al. did not include a passive control group, their results may also be due to placebo effects or a regression towards the mean.

There are some limitations to our study that may have affected our results. A general limitation of subjective measures is reactive measurement. Whenever people start to focus on their way of functioning, they may notice more cognitive failures, even though they may have made the same errors before. This would result in increased cognitive dysfunction reports over time, which may have masked any potential training effect.

The relationship between the proxy and the participant (e.g. whether they were partners or siblings) was unknown. It could be that the proxy did not spend enough time in proximity of the participant to notice any changes. The limited reliability of the questionnaires could also have played a role. Sometimes, for unknown reasons, a proxy completed the questionnaire twice at the same time-point, which was possible because the link to the questionnaire remained active, and these scores did not always agree. Because the questionnaires were completed online, it could have taken the proxies more effort to ask for clarifications than when they were completed face-to-face. However, it is unlikely that this led to different answers, as the questionnaires were formulated carefully. Moreover, the test-retest reliability in the proxies of the waiting list group who completed the questionnaires at both time-points (n = 16) was acceptable (DEX: r = .87, CFQ: r = .67, IADL: r = .95). Not every participant had a proxy, which may have caused selection bias. Nevertheless, participants with a proxy did not differ on baseline characteristics or time effects from those without a proxy.

Several participants did not complete the follow-up questionnaires (four weeks after training completion, T3). The substituted missing values in the intention-to-treat analyses may have influenced the results. In the intention-to-treat analyses there was a trend-wise significant cognitive improvement (based on the exit list), but this was not confirmed in the per-protocol analyses. The remainder of the follow-up results did not differ between the intention-to-treat and per-protocol analyses, suggesting that the effects on executive functioning and cognitive functioning (based on the CFQ) were valid.
Although not statistically significant, the percentage of participants who received cognitive rehabilitation during the study period differed between groups. This may have biased the training effect. Nevertheless, explorative analyses without participants who received cognitive rehabilitation during the study period did not change the results.

Generally, the effect sizes of change over time were small, which is in agreement with two large RCTs of computer-based training in healthy adults and older adults [53, 54]. One might argue that such small effects are not clinically relevant and not noticeable in daily living.

Possibly, the duration of our training was not sufficient to influence daily living and quality of life. Nevertheless, Westerberg et al. (2007) found improvements on cognitive functioning after 5 weeks of training, whereas Wentink et al. (2016) did not find an effect on the same outcome measure after 8 weeks of training. The relation between training duration and its effect on subjective functioning remains unclear.

The training did not include information on how training improvements could be used in daily living. Patients may benefit from advice on how to apply in daily life what they have learned in these more abstract training tasks. Nevertheless, the training aimed to improve core aspects of executive functioning. When clinically relevant improvements are made, they should generalize to untrained tasks such as daily life activities.

The sample of this study consisted of a heterogenous group of individuals who had a stroke 3 months to 5 years ago. Possibly, the heterogeneity of the sample may have resulted in a large variety of training effects. Previous studies did, however, find improvements in both post-acute stroke samples and chronic stroke samples [18, 19]. Furthermore, time since stroke did not explain a significant amount of variance in any of our analyses. Therefore, we do not assume that this may have affected our results. Future studies are needed to evaluate whether there is an essential time window after stroke in which computer-based retraining is most effective.

One of the strengths of this study is that the sample size was sufficient to detect clinically relevant effects. Lack of statistical power of our main multivariate analyses cannot explain the absence of group differences, because exploratory univariate analyses did not reveal any group difference either, except for CFQ. Second, this study examined everyday life functioning and quality of life. Even though the results are not positive, we hope that future studies continue to include these measures to examine cognitive functioning outside the lab. Another strength is that by including two control groups, we were able to rule out training-unspecific effects such as placebo effects and Hawthorne effect.

Our recommendation for future studies is to measure twice at baseline to avoid reactive measurement and to include objective measures of everyday life functioning. Whenever proxy ratings are included in a study, it is important to register the relationship between proxy and participant and the frequency of contact between the proxy and the participant during the study. Finally, it is important to include proper control groups as subjective measures are prone to placebo effects.

Our study provides insight into whether a computer-based cognitive training, as commercially available, can improve subjective cognitive functioning. Only a few studies have examined the effects of such training on subjective measures of daily living. It is important to include such outcome measures, because subjective functionality in daily living affects well-being after stroke. We did find improvements over time in subjective executive functioning and cognitive functioning. Even though these effects were found in all groups, suggesting these are due to training-unspecific effects, this could be beneficial for a better state of mind. Nevertheless, in line with several other training studies, our results did not support the effectiveness of commercial computer-based brain training programs above the effects of care as usual. Our results do not, however, imply that computer-based training programs can never work. More
specific and (even) more intense training and prolonged training may have beneficial effects. Future studies may determine whether training programs tailored to the individual are effective. Our general conclusion is that further research is needed before claims about the effectiveness of general brain training can be regarded as evidence-based.

**Supporting information**

S1 Table. Mean (standard deviation) and MANOVA of the outcome measures of per-protocol analyses. (PDF)

S1 File. Subjective brain training effects after stroke. (PDF)

S2 File. Protocol approved by medical ethical committee_TAPASS_V1.1. (PDF)

S3 File. TAPASS_Subj_CVA_minimal data excl demo info. (SAV)

S4 File. CONSORT 2010 checklist_TAPASS_Subjective. (PDF)

**Acknowledgments**

The authors would like to thank all the participants and their relatives for participating in the study; the students for collecting the data and coaching the participants; Dezzel Media for making Braingymmer available for our study; and the mental health care institutions for helping with patient recruitment (Heliomare Wijk aan Zee, Reade Amsterdam, De Hoogstraat Revalidatie Utrecht, De Trappenberg Huizen, Adelante Zorggroep Hoensbroek, Revant Lindenhoft Goes, and Academic Medical Centre Amsterdam, The Netherlands).

**Author Contributions**

**Conceptualization:** Renate M. van de Ven, Jaap M. J. Murre, Justine A. Aaronson, Coen A. M. van Bennekom, Ben Schmand.

**Data curation:** Renate M. van de Ven, Jessika I. V. Buitenweg.

**Formal analysis:** Renate M. van de Ven, Jaap M. J. Murre, K. Richard Ridderinkhof.

**Funding acquisition:** Jaap M. J. Murre.

**Investigation:** Renate M. van de Ven, Jessika I. V. Buitenweg.

**Methodology:** Renate M. van de Ven, Jaap M. J. Murre, Jessika I. V. Buitenweg.

**Project administration:** Renate M. van de Ven.

**Resources:** Renate M. van de Ven, Jaap M. J. Murre, Jessika I. V. Buitenweg, Dick J. Veltman, Justine A. Aaronson, Tanja C. W. Nijboer, Suzanne J. C. Kruiper-Doesborgh, Coen A. M. van Bennekom, K. Richard Ridderinkhof, Ben Schmand.

**Software:** Renate M. van de Ven, Jessika I. V. Buitenweg.

**Supervision:** Jaap M. J. Murre, Ben Schmand.

**Validation:** Renate M. van de Ven, Jessika I. V. Buitenweg.
Visualization: Renate M. van de Ven.

Writing – original draft: Renate M. van de Ven, Jaap M. J. Murre, Dick J. Veltman, Ben Schmand.

Writing – review & editing: Renate M. van de Ven, Jaap M. J. Murre, Jessika I. V. Buitenweg, Dick J. Veltman, Justine A. Aaronson, Tanja C. W. Nijboer, Suzanne J. C. Kruiper-Doesborgh, Coen A. M. van Bennekom, K. Richard Ridderinkhof, Ben Schmand.

References


32. Saghaei M. An Overview of Randomization and Minimization Programs for Randomized Clinical Trials. PLOS ONE | https://doi.org/10.1371/journal.pone.0187582 November 16, 2017 17 / 18


