

## Methods

### Measurement instruments

#### *Technology Acceptance*

Acceptability of playing the serious game was based on all intervention participants who installed the game. Acceptability was determined using the Technology Acceptance Model [TAM].[1] This model postulates that ease of using a technology influences the perceived usefulness and the attitude toward using and together form the behavioural intention to use a technology which leads to actual use. Ease of use was measured using the System Usability Score [SUS, 10 Likert-scale items, item scores ranging from 1 to 5] questionnaire taken directly from the TAM.[1] Answers are transformed to a score between 0 and 100.[2] The SUS score is highly reliable [ $\alpha = 0.91$ ] and useful over a wide range of interface types.[3] Perceived usefulness of the game was operationalised as enjoyment and assessed using the playful experiences questionnaire [PLEXQ, 17 constructs of playfulness, each measured through three items].[4] The constructs nurture, fellowship, cruelty and subversion were removed as these did not align with the intention of the serious puzzle game. Constructs are taken together to distinguish a four-factor structure of playfulness: stimulative, pragmatic, momentary and negative experiences.[4] Perceived usefulness of the behavioural tasks) was assessed at three months with five statements on a five-point Likert scale (ranked 0 – 4).

### Secondary analyses

Secondary analyses included results at one month and a per-protocol analysis where all intervention participants who played the game for more than one hour during the study period were considered adherent to the protocol. Exposure-response analyses were also performed: total play time was plotted against the continuous outcomes (CQR, BMQ NCD, RADAI and HAQ) to determine regression coefficient. In addition playtime was plotted for both adherent and non-adherent intervention participants, based on the CQR, to determine whether there was a difference in average playtime between both groups.

P-values  $< 0.05$  were considered statistically significant. Statistical analyses were performed using Stata version 13.1.

## Results

### Acceptability of the serious game

Ease of use was scored an average of 66 out of 100 during the study. According to the adjective rating scale of Bangor et al.[27], this means usability can be regarded as ‘good’ [see Table S1].

Perceived usefulness of the game was considered to be the playfulness experiences of the serious game. Experiences did not differ much over the study and had an overall mean score of around 3 out of 5 with the exception of the negative experiences which scored around 2. Patients scored neutral to negative on the statements regarding the behavioural tasks.

**Table S1. Acceptability outcomes of the serious game at 1 and 3 months**

<i>Ease of use</i>	<b>1 month [n = 86]</b>	<b>3 months [n = 78]</b>
Average SUS score (0-100) <i>mean ± SD</i>	66 ± 16	66 ± 14
<b><i>Perceived usefulness of the game</i></b>		
Stimulative experience (1-5) <i>median [IQR]</i>	3.2 [2.8 – 3.5]	3.2 [2.6 – 3.6]
Pragmatic experience (1-5) <i>median [IQR]</i>	3.0 [2.7 – 3.4]	2.9 [2.6 – 3.2]
Momentary experience (1-5) <i>median [IQR]</i>	3.0 [2.6 – 3.4]	2.9 [2.6 – 3.3]
Negative experience (1-5) <i>median [IQR]</i>	2.3 [1.7 – 2.7]	2.0 [1.7 – 2.7]
<b><i>Perceived usefulness of the behavioural tasks</i></b>		
It was clear to me that the behavioural tasks were meant to remind me of my medication.		3 [2 – 3]
I find it agreeable to be reminded of my medication through performing a behavioural task.		1 [1 – 2]
Performing a behavioural task helps me in using my medication.		1 [0 – 1]
The push notifications reminding me to come and play were of great added value.		1 [1 – 2]
The serious puzzle game is of added value in the treatment of my rheumatoid arthritis.		1 [0 – 2]

**Abbreviations:** IQR – inter quartile range; no. – number; SD – standard deviation

### Study outcomes at one month

At one month 64% of the intervention participants were adherent compared to 53% of the control group and 95% confidence interval of the difference was -22% to 6% and not statistically significant (see Table S2).

The serious game did not show an effect on secondary medication outcomes at one month (see Table S2). Results were similar to the outcomes at three months (see Table 2 in the manuscript).

**Table S2. Study outcomes at one month**

	<b>Control group (N=108)</b>	<b>Intervention group (N=95)</b>	<b>Group difference [95% CI]</b>
Adherent (N, %)*	56 (53)	59 (64)	-8% [-22 – 6]
CQR continuous (mean, SD)	74 ± 11	74 ± 11	0.4 [-2.8 – 3.6]
BMQ-Specific NCD score (mean, SD)	5.3 ± 4.5	4.8 ± 4.7	-0.5 [-1.8 – 0.8]

**Abbreviations:** 95% CI – 95% confidence interval; n – number; SD – standard deviation; CQR – Compliance Questionnaire on Rheumatology; BMQ – Beliefs about Medication questionnaire. \* Percentage of the total number of participants excluding missing data.

### Per protocol analysis

Of the 110 intervention participants that started the study, 87 participants (79%) installed the game and 70 participants (75%) played the game for at least an hour and were eligible for the per protocol analysis.

The per protocol analysis showed no differences between control and intervention group on medication or clinical outcomes (see Table S2). Pill count adherence was much higher (around 96%) as compared to adherence scored by the Compliance Questionnaire on Rheumatology (around 50%).

**Table S3. Per protocol analysis on medication outcomes at three months**

	<b>Control group (N=101)</b>	<b>Intervention group (N=70)</b>	<b>Group difference [95% CI]</b>
<b>Medication outcomes</b>			
Adherent (N, %)*	55 (54)	42 (62)	-7% [-22 – 8]
CQR continuous (mean, SD)	75 ± 12	72 ± 11	2.9 [-0.6 – 6.4]
Pill count# (mean, SD)	95 ± 16	97 ± 9	-2.0% [-9.7 – 5.7]
BMQ-Specific NCD score (mean, SD)	4.8 ± 4.2	5.3 ± 4.7	-0.5 [-1.9 – 0.8]
<b>Clinical outcomes</b>			
RADAI score (median, IQR)	2.5 [1.2 – 4.0]	2.5 [1.5 – 4.1]	0.0 [-0.9 – 0.8]
HAQ score (median, IQR)	0.8 [0.3 – 1.4]	0.6 [0.3 – 1.4]	-0.1 [-0.5 – 0.2]

**Abbreviations:** 95% CI – 95% confidence interval; n – number; SD – standard deviation; CQR – Compliance Questionnaire on Rheumatology; BMQ NCD – Beliefs about Medication questionnaire necessity-concerns differential; RADAI – Rheumatoid Arthritis Disease Activity Index; HAQ – Health Assessment Questionnaire; IQR – interquartile range. \* Percentage of the total number of participants excluding missing data. # N=21 for the control group and N=22 for the intervention group

### Exposure-response analyses

Playtime of intervention participants was plotted against continuous outcomes CQR (figure S1), BMQ NCD (figure S2) and HAQ and RADAI (figure S3) and regression coefficients were fitted. None of the continuous outcomes showed a relation with playtime. Playtime was also plotted in a boxplot for both nonadherent and adherent intervention participants as categorised by the CQR (figure S4). Median playtime and interquartile range do not differ between nonadherent and adherent participants. Both exposure-response analyses showed there is no relation between playtime and study outcomes.

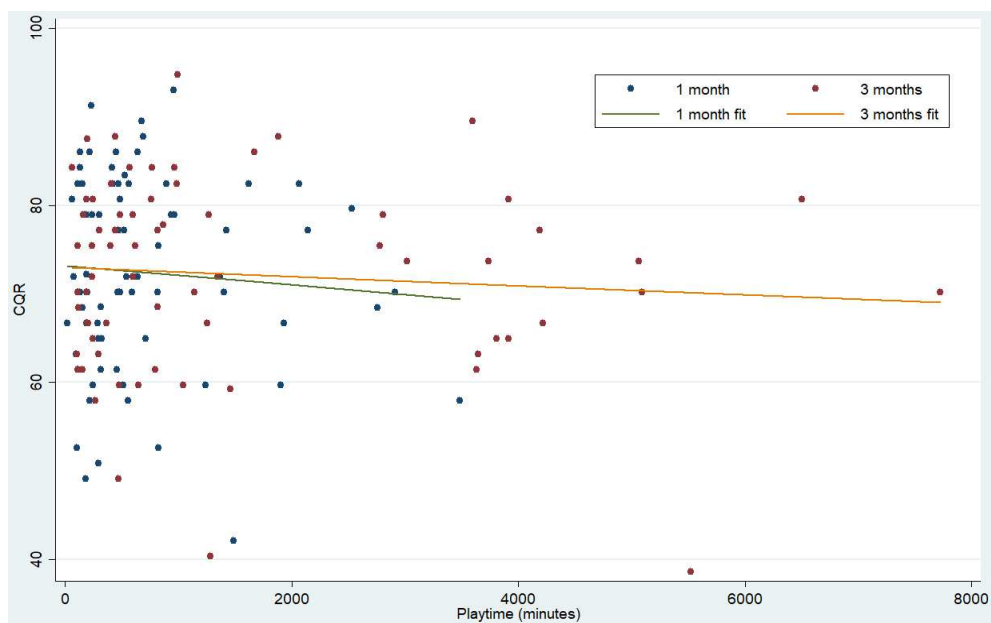


Figure S1. CQR continuous score versus playtime and plotted regression line at 1 and 3 months

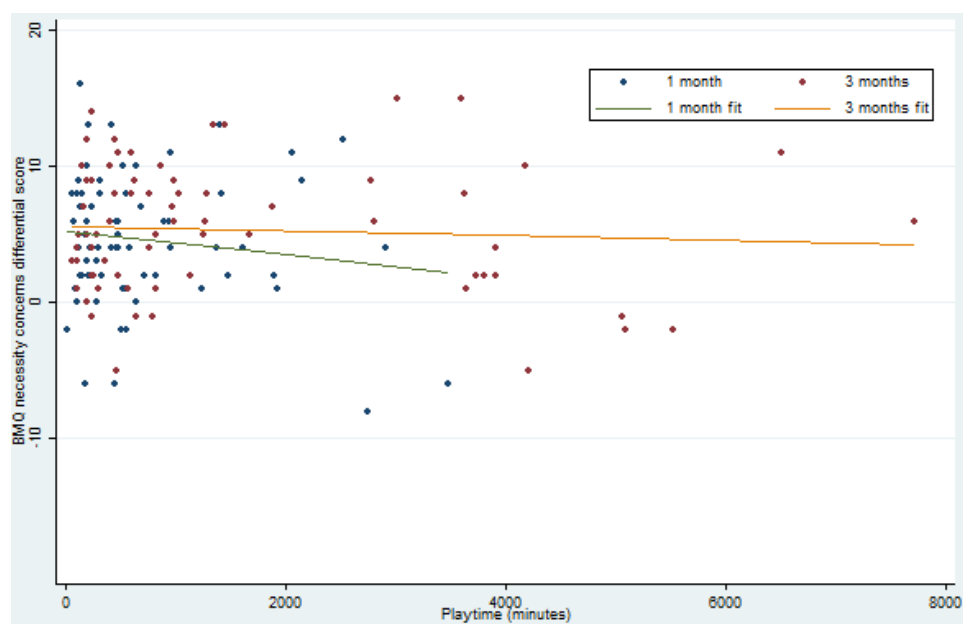


Figure S2. BMQ NCD-score versus playtime and plotted regression line at 1 and 3 months

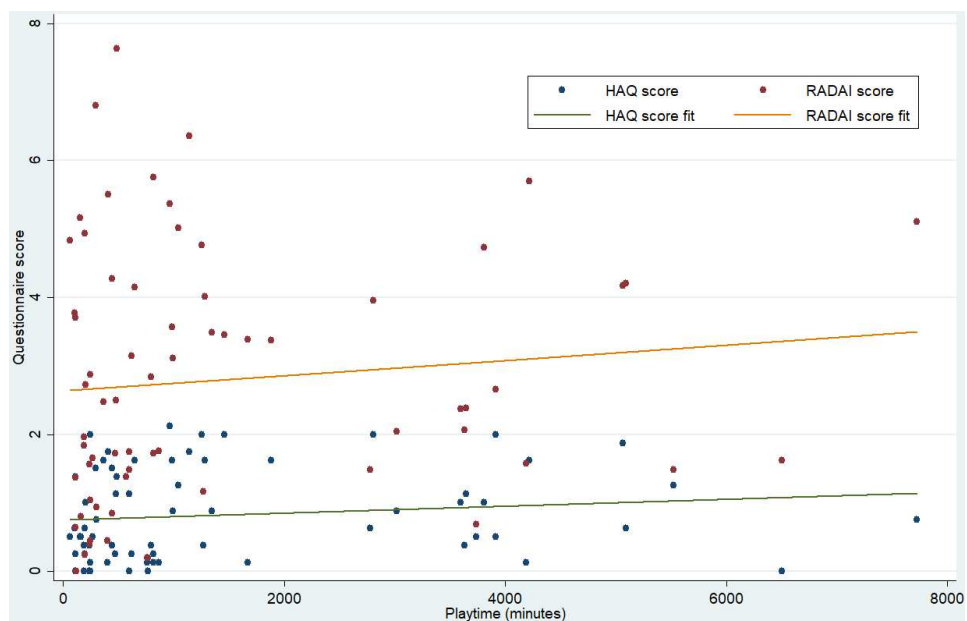


Figure S3. HAQ and RADAI score versus playtime and plotted regression line at 3 months

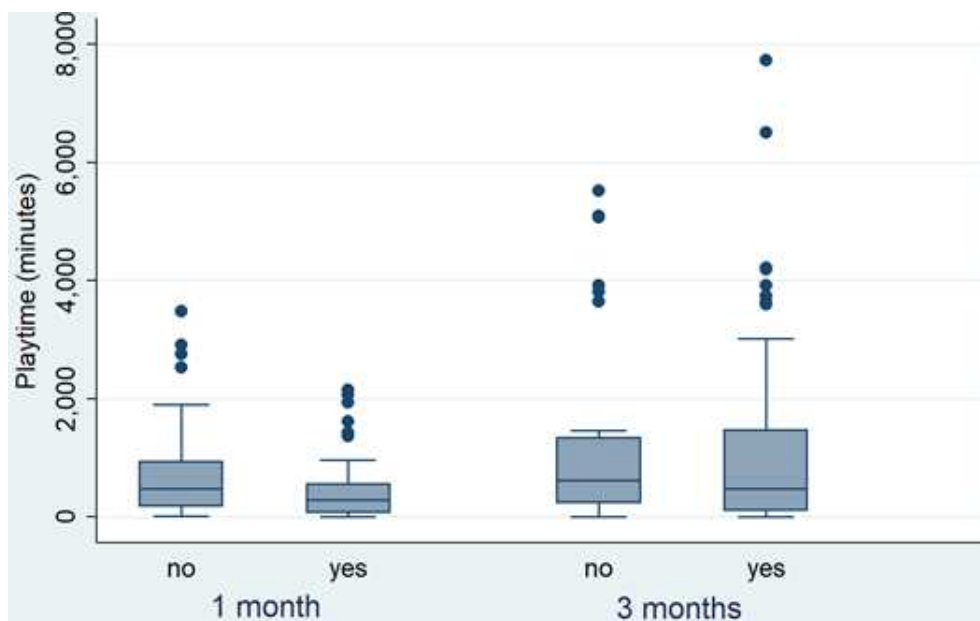


Figure S4. Boxplots of playtime for nonadherent (no) and adherent (yes) intervention participants at 1 and 3 months

## References

1. Davis FD. A Technology Acceptance Model For Emperically Testing New End-User Information Systems: Theory And Results. Massachusetts Institute of Technology; 1985.
2. Brooke J. SUS - A Quick and Dirty Usability Scale. In: Jordan, P.W., Thomas B, Weerdmeester, B.A., & McClelland IL, editors. Usability Eval Ind. Taylor and Francis Ltd; 1996. p. 189–194.
3. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. *Int J Hum Comput Interact*. 2008;24[6]:574–594.
4. Boberg M, Karapanos E, Holopainen J, Lucero A. PLEXQ: Towards a playful experiences questionnaire. *CHI Play 2015 - Proc 2015 Annu Symp Comput Interact Play*. 2015;381–392.