

## Supplementary material

### A. English translation of technology acceptance questionnaire.

Below is the English translation of the originally Dutch questionnaire sent to participants to measure technology acceptance. The questionnaire is divided into 3 parts.

#### Part 1: Perceived ease of use

*Question 1 to 10 originate from the System Usability Scale, and are answered on a 5-point Likert scale (Strongly disagree, Somewhat disagree, Neutral, Somewhat agree, Strongly agree).*

1. I think that I would like to use the app frequently.
2. I found the app unnecessarily complex.
3. I thought the app was easy to use.
4. I think that I would need the support of a technical person to be able to use the app.
5. I found the various functions in the app were well integrated.
6. I thought there was too much inconsistency in the app.
7. I would imagine that most people would learn to use the app very quickly.
8. I found the app very cumbersome to use.
9. I felt very confident using the app.
10. I needed to learn a lot of things before I could get going with the system.

*Question 11 to 17 originate from user version of the Mobile Application Rating Scale (uMARS) Sections A through D, and are answered on 5-point Likert scales related to each individual question.*

11. Is the app interesting to use? Does it present its information in an interesting way?
12. Does the app allow you to customize the settings and preferences that you would like to (e.g. sound, content and notifications)?
13. Is the app content (visuals, language, design) appropriate for the target audience?
14. How accurately/fast do the app features and components work?
15. Is arrangement and size of buttons, icons, menus and content on the screen appropriate?
16. Is the app content correct, well written, and relevant to the goal/topic of the app?
17. Is the information within the app comprehensive but concise?

#### Part 2: Perceived usefulness

*Questions 18 to 23 originate from uMARS section F, and are answered on a 5-point Likert scale (1 – strongly disagree to 5 – Strongly agree).*

18. This app has increased my awareness of the importance of addressing the health behaviour.
19. This app has increased my knowledge/understanding of the health behaviour.
20. The app changed my attitudes toward improving this health behaviour.
21. The app has increased my intentions/motivation to address this health behaviour.
22. This app would encourage me to seek further help to address the health behaviour (if I needed it).
23. Use of this app will increase/decrease the health behaviour.

### Part 3: Attitude towards using

*Questions 24 to 27 originate from uMARS section E, and are answered on 5-point Likert scales adjusted to each question.*

24. Would you recommend this app to people who might benefit from it?
25. How many times do you think you would use this app in the next 12 months if it was relevant to you?
26. Would you pay for this app?
27. What is your overall (star) rating of the app?

#### **B. Start- and end-times of measurement period.**

Start of measurement period was defined as either 1) 24 hours after delivery or dispensing by pharmacy, or 2) 24 hours after the temperature logger first measured a temperature below 8 degrees Celsius in the case of delivery via mail. For participants where temperature loggers did not measure temperatures below 8 degrees Celsius, start of measurement period was defined as 24 hours after the temperature logger first measured a temperature below 15 degrees Celsius.

End of measurement period was defined as either 1) 12 weeks after the start of the measurement period, minus 24 hours or 2) the moment storage temperature exceeded 15 degrees Celsius for at least 12 hours consecutively without subsequent cooling below 15 degrees Celsius for at least 48 hours consecutively, minus 24 hours.

A forgiveness of 24 hours was incorporated to account for adjustment periods of temperature loggers for both start and end of measurement periods.

**C. Numbers of patients that stored bDMARDs in accordance with individual requirements for correct storage.**

Correct storage baseline vs intervention

	Baseline		
Intervention	Correct	Incorrect	<i>total</i>
Correct	8	11	19
Incorrect	1	28	29
<i>Total</i>	9	39	48

Correct storage, criterium 'no more than 2 hours consecutively below 0 degrees' baseline vs intervention

	Baseline		
Intervention	Correct	Incorrect	<i>total</i>
Correct	39	3	42
Incorrect	4	2	6
<i>Total</i>	43	5	48

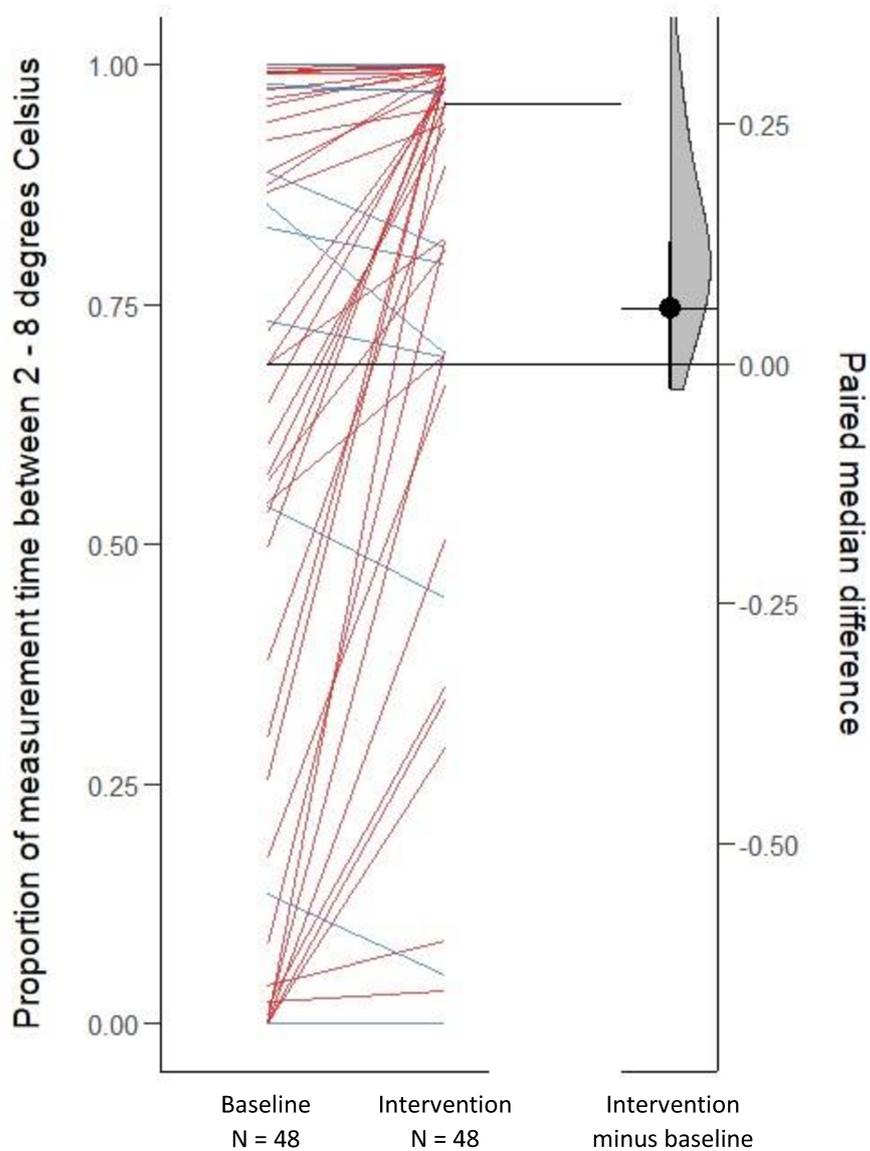
Correct storage, criterium 'no more than 2 hours consecutively above 25 degrees' baseline vs intervention

	Baseline		
Intervention	Correct	Incorrect	<i>total</i>
Correct	47	1	48
Incorrect	0	0	0
<i>Total</i>	0	0	48

Correct storage, criterium 'no more than 48 hours in total outside 2-8 degrees' baseline vs intervention

	Baseline		
Intervention	Correct	Incorrect	<i>total</i>
Correct	8	12	20
Incorrect	1	27	28
<i>total</i>	9	29	48

**D. Individual baseline and intervention measurements on proportion of measurement time within 2 – 8 degrees Celsius.**



Graph depicting the individual difference in proportion of measurement time between 2 – 8 degrees Celsius between baseline and intervention measurements. Red lines depict participants that improved from baseline to intervention, blue lines depict participants that declined from baseline to intervention period.