

## Sub-study 3: Does sending a pen with a follow-up questionnaire enhance return rates?

Material throughout this section has been reproduced from Fairhurst et al.<sup>1</sup>

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The text below includes minor additions and formatting changes to the original text.

### Introduction

RCTs are one of the key tools used to analyse the effectiveness of a new treatment. However, poor recruitment and retention rates pose a serious threat to RCTs as they can render the results of the trial inconclusive, prolong the duration of the trial and can even lead to the trial being closed down early.<sup>2</sup> Participants not completing follow-up data collection, can be very problematic for RCTs as it reduces power and, if differential between the arms, can introduce attrition bias.<sup>3</sup>

Various strategies have been deployed to help maximise retention in RCTs.<sup>4</sup> One such strategy is to include a pen when posting a follow-up questionnaire. This strategy is hypothesised to help improve retention response rates as it gives participants the means to complete the questionnaire while also making participants feel more inclined to return the questionnaire due to encouragement of positive reciprocal behaviour provided by the pen.<sup>5</sup> A SWAT aiming to investigate the impact of posting a pen with the 3-month follow-up participant questionnaire was embedded in the GYY trial.<sup>6</sup>

### Previous evidence

The TRIAL FORGE initiative has published an evidence pack on the use of sending a pen with a trial questionnaire and/or study materials on response rate (<https://www.trialforge.org/resource/evidence-pack-retention-adding-a-pen-ret3/>). Based on five prior RCTs,<sup>7-11</sup> they concluded that sending a pen probably increases retention and response rate (random effects meta-analysis pooled effect: increase in response rates of 1.9%, 95% CI 0.0% to 3.7%). We updated this meta-analysis with our results.

### Methods

#### Study design

This SWAT was a two-armed RCT embedded in the GYY trial.<sup>6</sup> The SWAT was registered with the Northern Ireland Network for Trials Methodology Research SWAT repository on 01/04/2019 (SWAT92;

<https://www.qub.ac.uk/sites/TheNorthernIrelandNetworkforTrialsMethodologyResearch/SWATSWARInformation/Repositories/SWATStore/>).

#### Participants

This study included participants allocated to the intervention arm of the GYY trial. Participants in the usual care arm of GYY were included in a different retention SWAT, namely the offer of a one-off GYY class at the end of their 12-month participation in the trial (described in the next section of this report). For logistical reasons, participants were randomised into the SWAT immediately after being randomised into the intervention arm of the main trial, but only those sent their 3-month questionnaire are actually included in this SWAT. Participants were not informed in advance that

they could be randomised into a SWAT to receive a pen with their 3-month questionnaire. This means that specific consent for the SWAT was not obtained; this was approved by the Research Ethics Committee as it was considered low risk. Written informed consent for the GYY main trial was obtained from all participants who took part.

### *Intervention*

The 3-month questionnaire was a 16-page booklet containing the following questions and standardised instruments: EQ-5D-5L,<sup>12</sup> PHQ-8,<sup>13</sup> GAD-7,<sup>14</sup> PROMIS-29,<sup>15</sup> UCLA 3-Item Loneliness scale<sup>16, 17</sup> and a direct loneliness question used in the English Longitudinal Study of Ageing,<sup>18</sup> questions asking about recent falls, bespoke health resource use items, and questions asking about participation in yoga over the previous 3 months. All participants in the GYY trial, who provided a valid mobile phone number and consented to be contacted via text message, were sent an SMS on the day the 3-month questionnaire was posted to them to pre-notify participants of its imminent arrival. Participants were also sent an unconditional GBP 5 with the questionnaire – this was in the form of cash (GBP 5 note) prior to the COVID-19 outbreak, and a shopping voucher thereafter. In addition, participants in the intervention group of the SWAT were sent a retractable ballpoint, black ink pen, branded with the GYY trial logo (Figure 1) with their 3-month follow-up postal questionnaire whereas the control group were not sent a pen with their 3-month questionnaire. Participants who did not return their 3-month questionnaire within two weeks were sent a postal reminder questionnaire; pens were not sent with reminder notices in either group. Telephone reminders, up to a maximum of three phone calls per participant, were additionally employed if the 3-month questionnaire had still not been returned within two weeks of the reminder questionnaire being sent.



**FIGURE 1** GYY logo-branded pen

### *Sample size*

No formal sample size calculation was undertaken as this was determined by the number of participants allocated to the intervention group of the main trial, which is typical for a SWAT. In this SWAT, the 240 participants allocated to the intervention arm in the main trial were randomised; this sample size was sufficient to have 80% power to detect an increase in response rates from 80% in the 'no pen' group to 93% in the 'pen' group assuming 10% of participants withdraw before the 3-month follow-up time point.

### *Randomisation*

Participants were randomised using simple randomisation and a 1:1 allocation ratio. The trial statistician, not otherwise involved in the recruitment or follow-up of participants, generated the allocation sequence using Stata v15.

### Blinding

Neither the statisticians analysing the data, nor the participants were blinded in this SWAT, as the nature of the intervention prevented the blinding of the latter to their allocation.

### Outcome measures

The primary outcome of this SWAT was the proportion of sent out 3-month follow-up questionnaires that were returned. Secondary outcomes were time taken to return the questionnaire, the proportion of participants who were sent a reminder to complete the questionnaire, and the completeness of the questionnaire. A full list of the outcomes measured in this SWAT are detailed in Table 1.

**TABLE 1** Outcome measures of the SWAT.

Outcome	Type	Definition
Proportion of 3-month questionnaires returned (primary)	Binary	The number of participants who returned their 3-month questionnaire divided by the number of participants who were sent this questionnaire.
Time taken to return 3-month questionnaire	Time to event	The number of days between the 3-month follow-up questionnaire being sent to the participant and being returned to York Trials Unit. This outcome is censored at 90 days for participants who do not return their 3-month questionnaire.
Reminder sent	Binary	The number of participants who were sent a reminder questionnaire to complete divided by the number of participants who were sent the 3-month questionnaire. Pens were not sent with the reminder questionnaires.
Number of items completed	Linear	The number of items completed in the questionnaire, if returned, out of a total of 78.

### Statistical Analysis

Outcomes are summarised by group and overall. For binary outcome measures, the count and proportion are reported and mean and standard deviation for number of completed items. For time to return, the median survival time (from the Kaplan–Meier survivor function) and its 95% CI are reported. Time to return was censored at 90 days (as participants were sent another follow-up questionnaire at 6 months post-randomisation) for participants who did not return their questionnaire.

Analyses were conducted under the principles of intention to treat (ITT) using two-tailed tests at the 5% significance level. Analyses were conducted in Stata v17. The primary outcome of 3-month questionnaire response was analysed using logistic regression adjusting for SWAT group allocation (“pen” or “no pen”), age, gender and an indicator variable for if the participant was allocated to receive an intervention (pen and/or GBP 5 versus neither) in a previous 2x2 factorial SWAT, which was undertaken at the recruitment stage of the GYY trial.<sup>19</sup> The treatment effect is presented as an odds ratio (OR) with associated 95% CI and p-value. The secondary outcomes were analysed as follows: time to return 3-month questionnaire by Cox Proportional Hazards model, with treatment effect presented as a hazard ratio (HR); whether a reminder was sent by logistic regression, with treatment effect presented as an OR; and number of completed items by linear regression, with

treatment effect presented as a mean difference. The models were adjusted as for the primary analysis.

Twenty-seven participants in the pen group were not sent a pen with their questionnaire due to an administrative error; per-protocol (PP) analyses were additionally conducted by removing these participants from the analysis models.

## Results

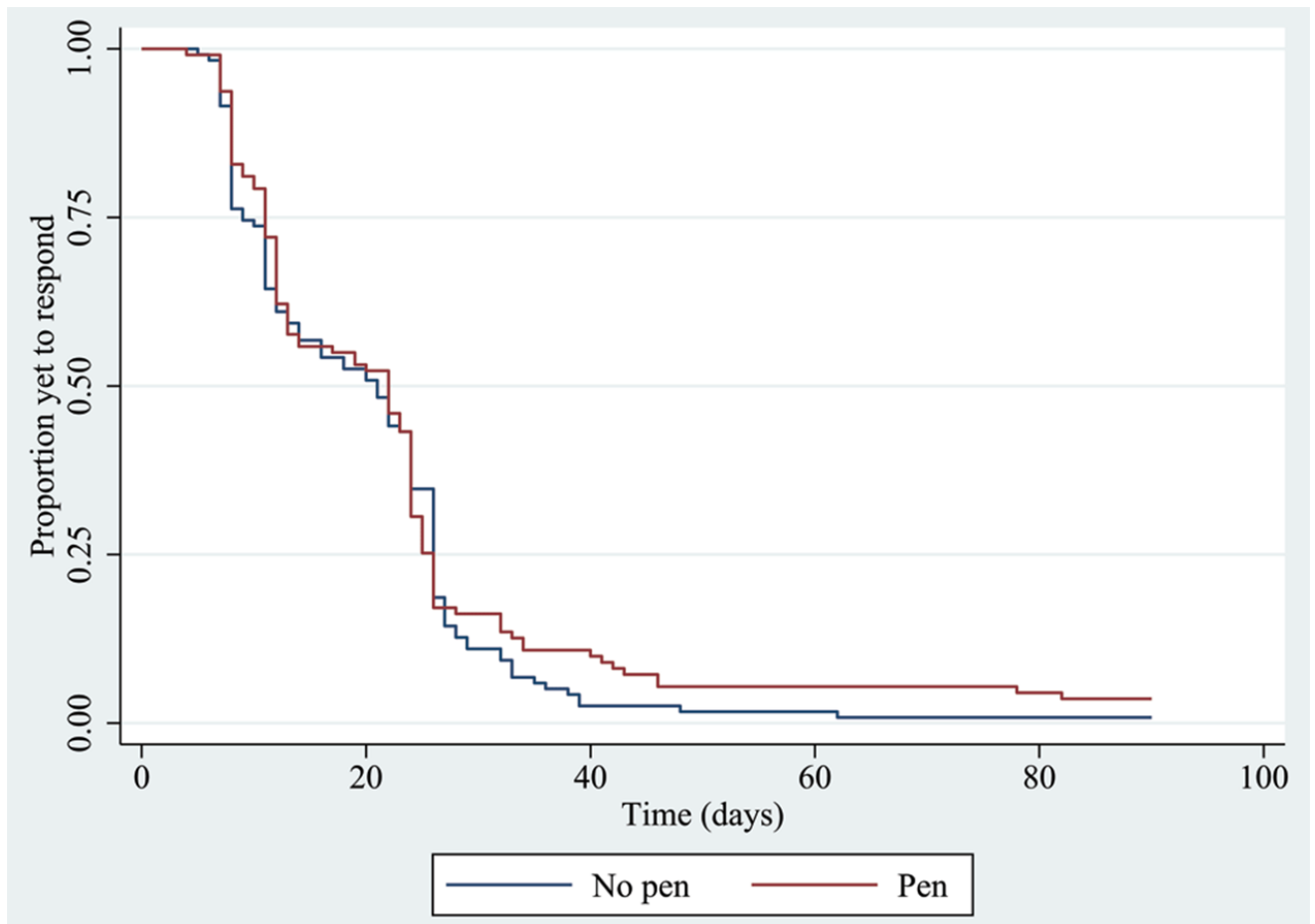
In total, 240 participants were randomised into the intervention arm of the main GYY trial, and 229 (95.4%) participants were sent their 3-month questionnaire and so were included in this SWAT (pen n=111; no pen n=118). The remaining 11 participants withdrew from the main trial before 3 months and so were not sent any follow-up questionnaires (6 (5.1%) from the pen group, and 5 (4.1%) from the no pen group). The questionnaires were mailed out between 20th January 2020 and 5th January 2022. Of participants sent a 3-month questionnaire, 144 (62.9%) were female (pen group n=66, 59.5%; no pen group n=78, 66.1%), the mean (SD) age was 73.2 (5.9) years (pen group 72.6 (5.5); no pen group 73.7 (6.2)), and 14 (6.1%) had been randomised to receive GBP 5 and/or a pen in the factorial recruitment SWAT (pen group n=7, 6.3%; no pen group n=7, 5.9%).

The proportion of participants who returned their 3-month questionnaire was similar in the two groups (pen n=107, 96.4%; no pen n=117, 99.2%) (Table 2). There was no evidence of a difference in return rates between the two groups (OR 0.23, 95% CI 0.02 to 2.19, p=0.20). The adjusted difference in proportions was -2.6 percentage points (95% CI -6.4 to 1.1).

**TABLE 2** Summary of SWAT trial results (ITT analysis).

Pen	No pen	Overall		
<b>Returned 3-month questionnaire, n/Total (%)</b>			<b>OR (95% CI)</b>	<b>p-value</b>
107 / 111 (96.4)	117 / 118 (99.2)	224 / 229 (97.8)	0.23 (0.02, 2.19)	0.20
<b>Reminder sent, n/Total (%)</b>			<b>OR (95% CI)</b>	<b>p-value</b>
30 / 111 (27.0)	35 / 118 (29.7)	65 / 229 (28.4)	0.85 (0.48, 1.53)	0.60
<b>Time to response (days), median (95% CI)</b>			<b>HR (95% CI)</b>	<b>p-value</b>
22.0 (13.0, 24.0)	21.0 (13.0, 24.0)	22.0 (14.0, 23.0)	0.91 (0.69, 1.18)	0.47
<b>Number of completed items (if questionnaire returned), mean (SD)</b>			<b>Mean difference (95% CI)</b>	<b>p-value</b>
77.2 (1.4)	76.6 (2.6)	76.9 (2.1)	0.51 (-0.04, 1.06)	0.07

There was no evidence of a difference in the proportion of participants sent a reminder in each of the groups (pen n=30, 27.0%; no pen n=35, 29.7%; OR 0.85, 95% CI 0.48 to 1.53, p=0.60), nor in the time to return the questionnaire. The median time to return was 22 days in the pen group and 21 days in the no pen group (HR 0.91, 95% CI 0.69 to 1.18, p=0.47) (Figure 2).



**FIGURE 2** Kaplan–Meier survivor functions for time to return 3-month follow up questionnaire in pen retention SWAT

While the Grambsch and Therneau test<sup>20</sup> provided no evidence that the proportional hazards assumption had been violated (covariate-specific test for SWAT allocation  $p=0.56$ ; global test  $p=0.56$ ), the lines in the Kaplan–Meier curve for the time to return between the two groups cross one another, which can be an indication that the proportional hazards assumption is unsafe. Therefore, in post hoc sensitivity analyses, both a log-rank test and a generalised gamma accelerated failure time (AFT) model were conducted. These are, respectively, a simpler and more complex alternative to the Cox model that do not assume proportional hazards. These analyses did not indicate evidence of a difference between the two groups (log-rank test:  $\chi^2(1)=0.67$ ,  $p=0.41$ ; AFT model: time ratio 1.05, 95% CI 0.90 to 1.23,  $p=0.56$ ).

Among participants who returned a questionnaire, there was weak evidence of a difference in the number of items on the questionnaire completed between the two groups (mean (SD): pen 77.2 (1.4); no pen 76.6 (2.6), mean difference 0.51, 95% CI  $-0.04$  to 1.06,  $p=0.07$ ).

#### *Per-protocol analysis*

A total of 202 participants were included in the per-protocol analyses (pen  $n=84$ ; no pen  $n=118$ ). Among these, 129 (63.9%) were female (pen group  $n=51$ , 60.7%; no pen group  $n=78$ , 66.1%), the mean (SD) age was 73.7 (6.0) years (pen group 73.7 (5.8); no pen group 73.7 (6.2)), and 11 (5.5%) had been randomised to receive GBP 5 and/or a pen in the factorial recruitment SWAT (pen group

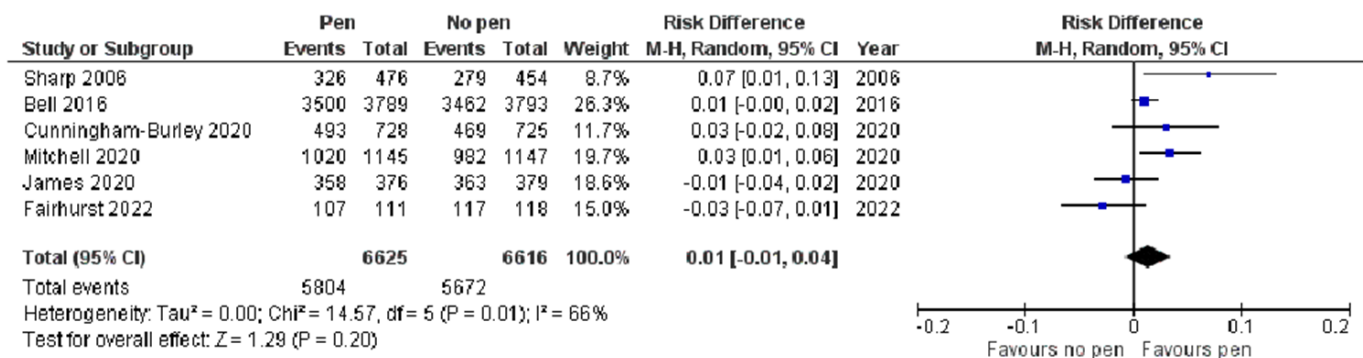
n=4, 4.8%; no pen group n=7, 5.9%). Results are provided in Table 3 and are similar to the ITT analysis.

**TABLE 3** Summary of SWAT trial results (PP analysis).

Pen	No pen	Overall		
<b>Returned 3-month questionnaire, n/Total (%)</b>			<b>OR (95% CI)</b>	<b>p-value</b>
82 / 84 (97.6)	117 / 118 (99.2)	199 / 202 (98.5)	0.42 (0.04, 4.91)	0.49
<b>Reminder sent, n/Total (%)</b>			<b>OR (95% CI)</b>	<b>p-value</b>
22 / 84 (26.2)	35 / 118 (29.7)	57 / 202 (28.2)	0.85 (0.45, 1.59)	0.60
<b>Time to response (days), median (95% CI)</b>			<b>HR (95% CI)</b>	<b>p-value</b>
13.0 (12.0, 20.0)	21.0 (13.0, 24.0)	16.0 (13.0, 22.0)	0.99 (0.75, 1.33)	0.97
<b>Number of completed items (if questionnaire returned), mean (SD)</b>			<b>Mean difference (95% CI)</b>	<b>p-value</b>
77.0 (1.5)	76.6 (2.6)	76.8 (2.2)	0.37 (-0.26, 0.99)	0.25

### Meta-analysis

Details of the included studies are as follows. Bell et al.<sup>7</sup> evaluated the use of adding a pen to the 60-month questionnaire in a trial of screening for the prevention of fractures in women aged 70–85 years; in Cunningham-Burley et al.,<sup>8</sup> the pen was added to the 14-week questionnaire in a slip-prevention trial among NHS staff (mean (SD) age 43 (11.3) years); James et al.<sup>9</sup> enclosed the pen in the 12-month questionnaire in a falls prevention trial in older people (65 years+); Mitchell et al.<sup>10</sup> investigated pens for the 14-week questionnaire in an orthopaedic trial (mean (SD) age 69 (8.9) years); and Sharp et al.<sup>11</sup> embedded the pen SWAT in a cervical screening trial in women (mean (SD) age 34 (10.4) years) at their next follow-up (12, 18, 24, or 30 months). A random effects meta-analysis conducted using RevMan 5.3 indicated that the pooled effect across the six included studies was a risk difference, favouring use of a pen, of 1% (95% CI -1% to 4%, p=0.20; Figure 3). An  $I^2$  value of 66% indicates moderate to large heterogeneity.



**FIGURE 3** Meta-analysis of inclusion of a pen on questionnaire return rates

### Discussion

The results of this trial do not indicate any demonstrable benefit of including a trial-branded pen with the postal 3-month questionnaire in the GYY trial. Indeed, a slightly higher response rate was observed in the no pen arm, albeit this required a marginally higher proportion of participants to be sent a reminder notice than in the pen group. The scope for improvement in the return rate for the

questionnaire was extremely limited given that, in the no pen group, all but one participant who was sent a questionnaire returned it. Furthermore, because of the high rate of return in the control group, the trial was severely underpowered to be able to detect a difference and so we would not have expected any statistically significant results.

In the meta-analysis, two trials were observed to have a negative effect, ours and James et al.<sup>9</sup> in both of these, the overall response rate was over 95%, whereas response rates averaged 78% among the four positive trials. This may explain some of the heterogeneity observed, and further evidence the limited potential for improvement when the response rates are already high.

Follow-up in GYY straddled the outbreak of the COVID-19 pandemic. A quarter of the 3-month questionnaires were sent out prior to COVID-19 having any real presence in our daily lives (all in January 2020), the next 3-month follow-ups were only due in December 2020 or later (up to January 2022). An exploratory, post hoc examination of the data suggests response rates were higher, across both the pen and no pen groups, in the follow-ups sent during the pandemic (97.7% and 100%, respectively) than those sent before (91.7% and 96.9%, respectively). This may be a chance finding, or it is possibly a direct consequence of the pandemic. Participants, particularly given their age, were likely to be adhering to social isolation guidelines and so may have had more time at home to complete the questionnaire. Additionally, it is feasible that news coverage of the pandemic could have increased awareness and respect in the population of the importance of research, trials and data, thus leading to greater engagement in the trial. The continually high response rates might additionally be attributed to the age group of participants, with many likely to be out of full-time employment or retired, hence able to more easily allocate time to completing and returning questionnaires, despite their reasonable length (the 3-month questionnaire was 16 pages long).

The strength of this study was that it was a randomised trial; however, since it was conducted in a population of older adults with multimorbidity, and particularly during the COVID-19 pandemic, findings may not be generalisable to other populations or contexts. This trial already implemented several retention strategies including sending an SMS to participants a few days before their postal questionnaire arrived, including an unconditional GBP 5 'thank you' payment, and reminder questionnaires and phone calls. All of these may have lessened the potential benefit of the addition of a pen with the mail out. Also, the incentive was tested at a reasonably early time point in the trial (3 months), when engagement in the trial might still be expected to be high; perhaps an increased benefit would have been seen at a later time point (further follow-ups in GYY were conducted at 6 and 12 months).

In summary, this SWAT suggests that enclosing a pen in a questionnaire mail out may not be an effective method to increase response rates in a trial of older adults with multimorbidity, particularly when other initiatives are in place, such as a prenotification SMS, an unconditional financial incentive, and a robust reminder procedure as was the case in this trial. Nevertheless, this SWAT adds to the growing evidence base of the effect of sending a pen out to trial participants on the rate of retention. Current pooled evidence suggests pens may still offer an effective incentive for improving response rates.

## References

1. Fairhurst C, Parkinson G, Hewitt C, Maturana C, Wiley L, Rose F, et al. Enclosing a pen in a postal questionnaire follow-up to increase response rate: a study within a trial. NIHR Open Research. 2022;2.

2. Gul RB, Ali PA. Clinical trials: the challenge of recruitment and retention of participants. *J Clin Nurs*. 2010;19(1-2):227-33.
3. Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. *BMJ Evid Based Med*. 2018;23(1):21-2.
4. Brueton VC, Tierney JF, Stenning S, Meredith S, Harding S, Nazareth I, et al. Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis. *BMJ Open*. 2014;4(2):e003821.
5. Whiteside K, Flett L, Mitchell A, Fairhurst C, Cockayne S, Rodgers S, et al. Using pens as an incentive for trial recruitment of older adults: An embedded randomised controlled trial. *F1000Res*. 2019;8:315.
6. Tew GA, Bissell L, Corbacho B, Fairhurst C, Howsam J, Hugill-Jones J, et al. Yoga for older adults with multimorbidity (the Gentle Years Yoga Trial): study protocol for a randomised controlled trial. *Trials*. 2021;22(1):269.
7. Bell K, Clark L, Fairhurst C, Mitchell N, Lenaghan E, Blacklock J, et al. Enclosing a pen reduced time to response to questionnaire mailings. *J Clin Epidemiol*. 2016;74:144-50.
8. Cunningham-Burley R, Roche J, Fairhurst C, Cockayne S, Hewitt C, Iles-Smith H, et al. Enclosing a pen to improve response rate to postal questionnaire: an embedded randomised controlled trial. *F1000Res*. 2020;9:577.
9. James S, Parker A, Cockayne S, Rodgers S, Fairhurst C, Torgerson DJ, et al. Including a pen and/or cover letter, containing social incentive text, had no effect on questionnaire response rate: a factorial randomised controlled Study within a Trial. *F1000Res*. 2020;9:623.
10. Mitchell AS, Cook L, Dean A, Fairhurst C, Northgraves M, Torgerson DJ, et al. Using pens as an incentive for questionnaire return in an orthopaedic trial: an embedded randomised controlled retention trial. *F1000Research*. 2021;9.
11. Sharp L, Cochran C, Cotton SC, Gray NM, Gallagher ME, group T. Enclosing a pen with a postal questionnaire can significantly increase the response rate. *J Clin Epidemiol*. 2006;59(7):747-54.
12. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-36.
13. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord*. 2009;114(1-3):163-73.
14. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092-7.
15. Fischer F, Gibbons C, Coste J, Valderas JM, Rose M, Leplege A. Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. *Qual Life Res*. 2018;27(4):999-1014.
16. Hughes ME, Waite LJ, Hawkey LC, Cacioppo JT. A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. *Res Aging*. 2004;26(6):655-72.
17. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. *J Pers Assess*. 1996;66(1):20-40.
18. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. *Proc Natl Acad Sci U S A*. 2013;110(15):5797-801.
19. Fairhurst C, Roche J, Bissell L, Hewitt C, Hugill-Jones J, Howsam J, et al. A 2x2 randomised factorial SWAT of the use of a pen and small, financial incentive to improve recruitment in a randomised controlled trial of yoga for older adults with multimorbidity. *F1000Res*. 2021;10:326.
20. Grambsch PM, Therneau TM. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika*. 1994;81(3):515-26.