



## RESEARCH ARTICLE

**REVISED** Enclosing a pen in a postal questionnaire follow-up to increase response rate: a study within a trial [version 2; peer review: 2 approved]

Caroline Fairhurst <sup>1</sup>, Gillian Parkinson <sup>1</sup>, Catherine Hewitt <sup>1</sup>, Camila Maturana <sup>1</sup>, Laura Wiley <sup>1</sup>, Fiona Rose <sup>1</sup>, David Torgerson<sup>1</sup>, Jessica Hugill-Jones <sup>1</sup>, Alison Booth <sup>1</sup>, Laura Bissell<sup>2</sup>, Garry Tew <sup>3</sup>

<sup>1</sup>York Trials Unit, Department of Health Sciences, University of York, York, YO10 5DD, UK

<sup>2</sup>British Wheel of Yoga Qualifications, Sleaford, NG34 7RU, UK

<sup>3</sup>York St John University, York, YO31 7EX, UK

**V2** First published: 31 Oct 2022, 2:53  
<https://doi.org/10.3310/nihropenres.13324.1>  
 Latest published: 20 Apr 2023, 2:53  
<https://doi.org/10.3310/nihropenres.13324.2>

### Abstract

**Background:** Poor response rates to follow-up questionnaires can adversely affect the progress of a randomised controlled trial and the validity of its results. This embedded 'study within a trial' aimed to investigate the impact of including a pen with the postal 3-month questionnaire completed by the trial participants on the response rates to this questionnaire.

**Methods:** This study was a two-armed randomised controlled trial nested in the Gentle Years Yoga (GYG) trial. Participants in the intervention group of the GYG trial were allocated 1:1 using simple randomisation to either receive a pen (intervention) or no pen with their 3-month questionnaire (control). The primary outcome was the proportion of participants sent a 3-month questionnaire who returned it. Secondary outcomes were time taken to return the questionnaire, proportion of participants sent a reminder to return the questionnaire, and completeness of the questionnaire. Binary outcomes were analysed using logistic regression, time to return by Cox Proportional hazards regression and number of items completed by linear regression.

**Results:** There were 111 participants randomised to the pen group and 118 to the no pen group who were sent a 3-month questionnaire. There was no evidence of a difference in return rates between the two groups (pen 107 (96.4%), no pen 117 (99.2%); OR 0.23, 95% CI 0.02 to 2.19,  $p=0.20$ ). Furthermore, there was no evidence of a difference between the two groups in terms of time to return the questionnaire (HR 0.90, 95% CI 0.69 to 1.18,  $p=0.47$ ), the proportion of participants sent a reminder (OR 0.85, 95% CI 0.48 to 1.53,  $p=0.60$ ) nor the number of items completed (mean difference 0.51, 95% CI -0.04 to 1.06,

### Open Peer Review

Approval Status

	1	2
<b>version 2</b> (revision) 20 Apr 2023	 view	
<b>version 1</b> 31 Oct 2022	 view	 view

1. **Graeme MacLennan** , University of Aberdeen, Aberdeen, UK
2. **Valarie Smith** , Trinity College Dublin, Dublin, Ireland

Any reports and responses or comments on the article can be found at the end of the article.

p=0.07).

**Conclusion:** The inclusion of a pen with the postal 3-month follow-up questionnaire did not have a statistically significant effect on response rate.

### Keywords

study within a trial, pen, postal questionnaire, retention, randomised controlled trial, older people, multimorbidity

**Corresponding author:** Caroline Fairhurst ([caroline.fairhurst@york.ac.uk](mailto:caroline.fairhurst@york.ac.uk))

**Author roles:** **Fairhurst C:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Parkinson G:** Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing; **Hewitt C:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Maturana C:** Project Administration, Writing – Review & Editing; **Wiley L:** Project Administration, Writing – Review & Editing; **Rose F:** Project Administration, Writing – Review & Editing; **Torgerson D:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing; **Hugill-Jones J:** Project Administration, Writing – Review & Editing; **Booth A:** Project Administration, Writing – Review & Editing; **Bissell L:** Conceptualization, Funding Acquisition, Writing – Review & Editing; **Tew G:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This project is funded by the National Institute for Health and Care Research (NIHR) under its Health Technology Assessment (HTA) Programme (Grant Reference Number 17/94/36). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

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

**Copyright:** © 2023 Fairhurst C *et al.* 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.

**How to cite this article:** Fairhurst C, Parkinson G, Hewitt C *et al.* **Enclosing a pen in a postal questionnaire follow-up to increase response rate: a study within a trial [version 2; peer review: 2 approved]** NIHR Open Research 2023, 2:53 <https://doi.org/10.3310/nihropenres.13324.2>

**First published:** 31 Oct 2022, 2:53 <https://doi.org/10.3310/nihropenres.13324.1>

**REVISED Amendments from Version 1**

In response to the reviewers' very helpful comments, we have made the following amendments to the manuscript.

A sentence has been added to the Intervention section to explain that all participants, in both the intervention and control group, were sent an unconditional GBP 5 with questionnaires at all follow-up timepoints in the host trial as a thank you for their continuing participation. A justification for the absence of blinding of the trial statistician is now included in the Blinding section. We have added into the Methods section that we present the treatment effect for the primary outcome in the form of an adjusted absolute difference in proportions, as well as an odds ratio, and clarified that these are both estimated from the primary logistic regression model. We have also moved the description of the post hoc sensitivity analyses from the Results section to the Methods. We have changed the Kaplan-Meier graph to display the proportion responding and amended the scale of the x-axis (with increments every 7 days and labels every 14). Further detail on the distribution of the 'number of items completed' outcome is included in the Results section, and we have introduced a post hoc Wilcoxon rank-sum test since there was evidence that the assumptions of the linear regression might not be met. We have also expanded on recommendations for future meta-analyses to correct for baseline risk in the Discussion section.

Other general edits have been made throughout the paper to improve clarity and readability.

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

Randomised controlled trials (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>1</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>2</sup>.

Various strategies have been deployed to help maximise retention in RCTs<sup>3</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>4</sup>. A study within a trial (SWAT) aiming to investigate the impact of posting a pen with the 3-month follow-up participant questionnaire was embedded in the Gentle Years Yoga (GYG) trial<sup>5</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>6-10</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 shall update this meta-analysis with our results.

## Methods

### Study design

This SWAT was a two-armed RCT embedded in the GYG trial that aims to investigate the impact of the offer of participation in a 12-week Yoga programme on the health-related quality of life of older adults with multimorbidity in England and Wales<sup>5</sup>. This study is being conducted by the York Trials Unit (YTU), University of York (recruitment complete and trial in follow-up at the time of writing; ISRCTN13567538, registered 18/03/2019 <https://www.isrctn.com/ISRCTN13567538>). 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/SWATSWAR-Information/Repositories/SWATStore/>). The GYG trial, and its embedded sub-studies, was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment (HTA) Programme (ref 17/94/36) and received approval from the North East–York Research Ethics Committee on 24/04/2019 (19/NE/0072), and the Health Research Authority.

### Participants

This study included participants allocated to the intervention arm of the GYG trial. Participants in the usual care arm of GYG were included in a different retention SWAT, namely the offer of a one-off GYG class at the end of their 12-month participation in the trial. This SWAT will be reported separately. 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 GYG 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>11</sup>, PHQ-8<sup>12</sup>, GAD-7<sup>13</sup>, PROMIS-29<sup>14</sup>, UCLA 3-Item Loneliness scale<sup>15,16</sup> and a direct loneliness question used in the English Longitudinal Study of Ageing, and questions asking about recent falls, health resource use, and participation in yoga over the previous 3 months. All participants in the GYG 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. This action formed part of the host trial's approach to promoting

continuing participation and was standard practice with each follow-up questionnaire across both arms of the GYY trial. 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.

**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, for example, 80% in the ‘no pen’ group to 93% in the ‘pen’ group assuming 10% of participants withdraw before the 3-month follow-up timepoint.



**Figure 1.** GYY logo-branded SWAT pen.

**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 (RRID: SCR\_012763). Stata is a proprietary software but an open-access alternative in which the sequence could have been generated is Google Sheets (RRID:SCR\_017679).

**Blinding**

The nature of the intervention prevented the blinding of participants to their allocation. Nor were the statisticians analysing the data blinded, as the risk associated with this was deemed to be low as the outcomes are objective and were prespecified, along with the analyses, in the host trial protocol.

**Outcomes**

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.

**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% confidence interval (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.

**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.

Analyses were conducted in Stata v17 (RRID: SCR\_012763). An open-access alternative that can perform an equivalent function to Stata for analysis is R, a free software environment for statistical computing and graphics (RRID: SCR\_001905). 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 2×2 factorial SWAT, which was undertaken at the recruitment stage of the GYY trial<sup>17</sup>. The treatment effect is estimated from the logistic regression model and presented as an odds ratio (OR) and adjusted absolute difference in proportions, 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.

*Post hoc* sensitivity analyses were conducted for the time to return the 3-month questionnaire outcome following indications that the Proportional Hazards assumption was violated for the Cox model. The analyses included both a log-rank test and a generalized gamma accelerated failure time (AFT) model, which are, respectively, a simpler and more complex alternative to the Cox model that do not assume proportional hazards. There was also evidence that the standardised residuals for the linear regression of number of items completed were not normally distributed and so the assumptions of this approach were questionable. Therefore, a *post hoc* Wilcoxon rank-sum test was performed, which does not make any assumption about the distribution.

27 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 20<sup>th</sup> January 2020 and 5<sup>th</sup> 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%).

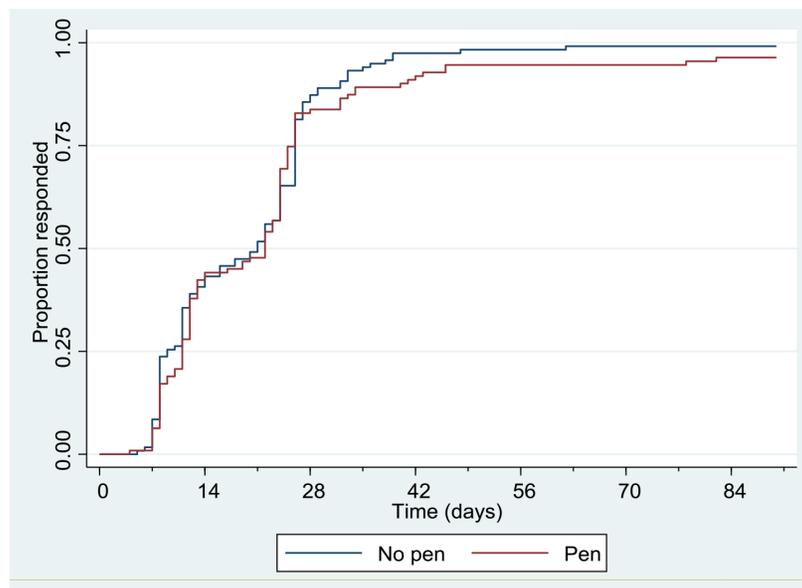
The proportion of participants who returned their 3-month questionnaire was similar in the two groups (pen n=107, 96.4%; no pen n=117, 99.2%) (Table 2). There was no evidence of a difference in return rates between the two groups (OR 0.23, 95% CI 0.02 to 2.19, p=0.20). The adjusted difference in proportions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

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).

While the Grambsch and Therneau<sup>18</sup> test provided no evidence that the proportional hazards assumption had been violated

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

Results					
	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



**Figure 2. Kaplan–Meier survivor functions for time to return 3-month follow up questionnaire.**

(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. The log-rank test and generalized gamma AFT model conducted as post hoc sensitivity 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,  $p=0.07$ ). However, there was evidence of a ceiling effect with this outcome, with 64.6% ( $n=148$ ) of returned forms having 100% completion (response to all 78 items). Additionally, there was a difference in the variance between the two groups with a larger SD in the no pen group than the pen group. This was caused by a small number of participants completing noticeably fewer items in the pen group than the no pen group (range 62–78 compared to 72–78). The standardised residuals from the linear regression demonstrated deviation from normality and so a post hoc Wilcoxon rank-sum test was performed, which indicated some evidence of a difference ( $p=0.09$ ).

#### 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.

#### Meta-analysis

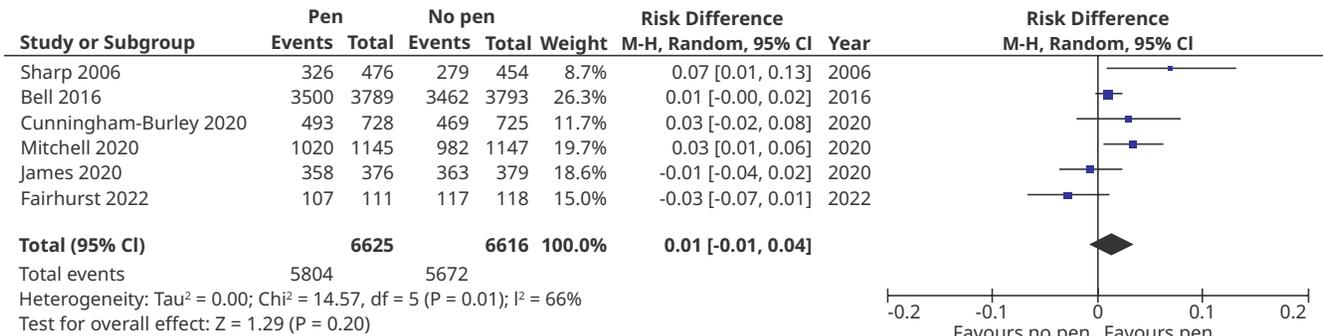
Details of the included studies are as follows. Bell *et al.* (2016)<sup>6</sup> evaluated the use of adding a pen to the 60-month questionnaire in a trial of screening for the prevention of fractures in women aged 70–85 years; in Cunningham-Burley *et al.* (2020)<sup>7</sup>, the pen was added to the 14-week questionnaire in a slip-prevention trial among NHS staff (mean (SD) age 43 (11.3) years); James *et al.* (2020)<sup>8</sup> enclosed the pen in the 12-month questionnaire in a falls prevention trial in older people (65 years+); Mitchell *et al.* (2020)<sup>9</sup> investigated pens for the 14-week questionnaire in an orthopaedic trial (mean (SD) age 69 (8.9) years); and Sharp *et al.* (2006)<sup>10</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 (RRID: SCR\_003581) 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.

#### 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

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

Results				
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



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

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.* (2021); in both of these, the overall response rate was over 95%, whereas response rates averaged 78% among the four positive component trials. This may explain some of the heterogeneity observed, and further evidence the limited potential for improvement when the response rates are already high. To address these limitations, future meta-analyses could include factors to correct for baseline response rates, allowing the impact of consistently high response rates to be accounted for.

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 generalizable 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 timepoint 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 timepoint (further follow-ups in GYY were conducted at 6 and 12 months).

## Conclusion

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.

## Data availability

### Underlying data

OSF: Underlying data for ‘Enclosing a pen in a postal questionnaire follow-up to increase response rate: a Study within a Trial’. <https://doi.org/10.17605/OSF.IO/26MPA>

This project contains the following underlying data:

Data file 1: GYY\_retention\_SWAT\_csv\_data.csv

Data file 2: GYY\_retention\_SWAT\_Stata\_data.dta

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

### Analysis code

OSF: Code to replicate the completed analyses in ‘Enclosing a pen in a postal questionnaire follow-up to increase response rate: a Study within a Trial’. <https://doi.org/10.17605/OSF.IO/26MPA>

Code file: GYY\_retention\_SWAT\_analysis.do

Analysis code, revised following peer-review, can be found here: <https://doi.org/10.17605/OSF.IO/9BGW6>

Code file: GYY\_retention\_SWAT\_analysis\_revised.do

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

### Reporting guidelines

OSF: CONSORT checklist for ‘Enclosing a pen in a postal questionnaire follow-up to increase response rate: a Study within a Trial’. <https://doi.org/10.17605/OSF.IO/26MPA>

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

## Acknowledgements

We would like to acknowledge the contributions of Professor Tim Rapley (Department of Social Work, Education and Community Wellbeing, Northumbria University), Belen Corbacho Martin, Helen Tilbrook (York Trials Unit, University of York) and Jenny Howsam (British Wheel of Yoga Qualifications), who were co-applicants on the initial GYY trial grant proposal, and Lesley Ward (Department of Sport, Exercise and Rehabilitation, Northumbria University) for her contributions to the trial.

## References

- Gul RB, Ali PA: **Clinical trials: the challenge of recruitment and retention of participants.** *J Clin Nurs.* 2010; **19**(1–2): 227–233. [PubMed Abstract](#) | [Publisher Full Text](#)
- Nunan D, Aronson J, Bankhead C: **Catalogue of bias: attrition bias.** *BMJ Evid Based Med.* 2018; **23**(1): 21–22. [PubMed Abstract](#) | [Publisher Full Text](#)
- Brueeton VC, Tierney JF, Stenning S, *et al.*: **Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis.** *BMJ Open.* 2014; **4**(2): e003821. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Whiteside K, Flett L, Mitchell AS, *et al.*: **Using pens as an incentive for trial recruitment of older adults: An embedded randomised controlled trial [version 1; peer review: 2 approved].** *F1000Res.* 2019; **8**: 315. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tew GA, Bissell L, Corbacho B, *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. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bell K, Clark L, Fairhurst C, *et al.*: **Enclosing a pen reduced time to response to questionnaire mailings.** *J Clin Epidemiol.* 2016; **74**: 144–150. [PubMed Abstract](#) | [Publisher Full Text](#)
- Cunningham-Burley R, Roche J, Fairhurst C, *et al.*: **Enclosing a pen to improve**

- response rate to postal questionnaire: an embedded randomised controlled trial [version 1; peer review: 2 approved].** *F1000Res.* 2020; **9**: 577. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. James S, Parker A, Cockayne S, *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 [version 2; peer review: 2 approved].** *F1000Res.* 2021; **9**: 623. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  9. Mitchell AS, Cook L, Dean A, *et al.*: **Using pens as an incentive for questionnaire return in an orthopaedic trial: an embedded randomised controlled retention trial [version 2; peer review: 1 approved, 1 approved with reservations].** *F1000Res.* 2021; **9**: 321. [Publisher Full Text](#)
  10. Sharp L, Cochran C, Cotton SC, *et al.*: **Enclosing a pen with a postal questionnaire can significantly increase the response rate.** *J Clin Epidemiol.* 2006; **59**(7): 747–54. [PubMed Abstract](#) | [Publisher Full Text](#)
  11. Herdman M, Gudex C, Lloyd A, *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. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  12. Kroenke K, Strine TW, Spitzer RL, *et al.*: **The PHQ-8 as a measure of current depression in the general population.** *J Affect Disord.* 2009; **114**(1–3): 163–73. [PubMed Abstract](#) | [Publisher Full Text](#)
  13. Spitzer RL, Kroenke K, Williams JBW, *et al.*: **A brief measure for assessing generalized anxiety disorder: the GAD-7.** *Arch Intern Med.* 2006; **166**(10): 1092–7. [PubMed Abstract](#) | [Publisher Full Text](#)
  14. Fischer F, Gibbons C, Coste J, *et al.*: **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. [PubMed Abstract](#) | [Publisher Full Text](#)
  15. Hughes ME, Waite LJ, Hawkey LC, *et al.*: **A short scale for measuring loneliness in large surveys: results from two population-based studies.** *Res Aging.* 2004; **26**(6): 655–72. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  16. Russell DW: **UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure.** *J Pers Assess.* 1996; **66**(1): 20–40. [PubMed Abstract](#) | [Publisher Full Text](#)
  17. Fairhurst C, Roche J, Bissell L, *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 [version 2; peer review: 2 approved].** *F1000Res.* 2022; **10**: 326. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  18. Grambsch PM, Therneau TM: **Proportional hazards tests and diagnostics based on weighted residuals.** *Biometrika.* 1994; **81**(3): 515–526. [Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 20 April 2023

<https://doi.org/10.3310/nihropenres.14559.r29183>

© 2023 MacLennan G. This is an open access peer review report 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.



**Graeme MacLennan** 

Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, Scotland, UK

No further comments from me.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Clinical trials statistician

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Reviewer Report 20 February 2023

<https://doi.org/10.3310/nihropenres.14450.r29002>

© 2023 Smith V. This is an open access peer review report 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.



**Valarie Smith** 

<sup>1</sup> School of Nursing and Midwifery, Trinity College Dublin, Dublin, Ireland

<sup>2</sup> School of Nursing and Midwifery, Trinity College Dublin, Dublin, Ireland

This is a well-written paper that adds to the body of evidence on the use of a pen for the return of follow-up questionnaires in a trial. The study is presented as a SWAT and is methodologically sound. I enjoyed reading the paper, and I liked that the authors updated the previous meta-

analysis on this topic to incorporate their own results. I only have two queries regarding the report:

1. The authors note the statistician was not blinded. Why was this? An explanation would help.
2. Also in the description of the intervention, I wasn't clear where the 5GBP came in. Did all participants of the main host trial (both groups) receive this, or was this specific to this (or another) embedded SWAT? Again a line or two for reader clarity would be helpful here.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** RCTs, SWATs, trial methodology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 04 Apr 2023

**Caroline Fairhurst**

We would like to thank the reviewer for their thorough and constructive feedback. We outline below our responses and the changes made to the manuscript, which we believe have really improved the paper.

We have added further details to the Blinding section to explain the justification for the absence of blinding of the trial statistician in this SWAT, as this was considered low risk as the outcomes were subjective and were prespecified in the host trial protocol.

The Intervention section has been amended to clarify the use of the GBP 5 incentive. This was not specific to a SWAT but formed part of the host trial's approach to encouraging

continuing participation. As such, the inclusion of GBP 5 was used at each follow-up timepoint across both the intervention and control arms.

**Competing Interests:** No competing interests

Reviewer Report 23 January 2023

<https://doi.org/10.3310/nihropenres.14450.r28906>

© 2023 MacLennan G. This is an open access peer review report 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.



**Graeme MacLennan** 

<sup>1</sup> Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, Scotland, UK

<sup>2</sup> Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, Scotland, UK

This manuscript describes the results of a randomised SWAT to evaluate the inclusion of a pen to with a questionnaire to improve response the Gentle Years Yoga trial. My review is primarily statistical, I have only minor comments and suggestions:

- Sample size, suggest that authors make clear that the 80% control response is the based on the 20% expected attrition in GYY (if indeed it was, could be a coincidence!).
- Methods/Results, the primary outcome is reported on the adjusted absolute difference scale also, this is not mentioned in the methods, I take it this was estimated from the logistic regression model? I ask because the upper bound of that CI is not possible given the control proportion.
- Methods/results, the description of the post hoc analysis would be better in the methods section. As an aside, rather than the PH assumption being violated, I think the the more reasonable assumption is that underlying DGM is a HR of 1, but accept the belt and braces approach.
- Number of items completed, there are ceiling effects here and clear difference in variance, probably caused by radically fewer items completed by one or two people in the control group. Did you investigate this further? Would the primary outcome completion at 3 months not have been of more interest?
- Meta-analysis, as aside, I agree with the point in discussion about high response rates in the controls groups, which points towards more sophisticated MA requirements in future (correcting for baseline risk).

#### Suggestions

- In this instance the time-to-event outcome might be better plotted as the "failure", i.e. proportion responding, rather than yet to respond, this makes more intuitive sense. KM plot, the ticks on the time axis could be weekly or fortnightly, again more intuitive for this

time scale. This is just a suggestion, can be ignored.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Clinical trials statistician

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 04 Apr 2023

**Caroline Fairhurst**

We would like to thank the reviewer for their thorough and constructive feedback. We outline below our responses and the changes made to the manuscript, which we believe have really improved the paper.

The SWAT was not specifically powered to detect a certain difference and so the Sample Size section simply presents an example of the difference that the trial would be powered to detect. The 80% control rate is not necessarily based on the expected attrition assumed in the sample size calculation for the host trial as this was based on participants who provide a valid response to the EQ-5D-5L at one or more of the assessment timepoints in the host trial (3, 6, and 12 months), not necessarily at a single timepoint.

We have added into the Methods section that we present the treatment effect for the primary outcome in the form of an adjusted absolute difference in proportions, as well as an odds ratio, and clarified that these are both estimated from the primary logistic regression model. We have also moved the description of the post hoc sensitivity analyses from the Results section to the Methods. We have changed the Kaplan-Meier graph as

suggested; we agree that the proportion responding is a more appropriate measure to display given the context of the outcome and a scale with increments every 7 days (labels every 14) is more intuitive and readable for the user.

We did not specify the completion of the primary outcome (from the host trial, the EQ-5D-5L) as a binary outcome for this SWAT. This was because the EQ-5D-5L is a short, standardised instrument and was the first section in the questionnaire and, as such, we expected that if a questionnaire was returned then this section would be well completed. Therefore, this analysis would have been virtually identical to the primary analysis of whether or not a questionnaire was returned. We opted to consider the number of completed items in total as, in addition to several standardised instruments, this also included questions asking about recent falls, health resource use, and participation in yoga over the previous 3 months, which may not have been completed as thoroughly. The reviewer is correct that there is evidence of a ceiling effect with this outcome, and that the difference in variance (the standard deviation was smaller in the 'pen' group than the 'no pen' group) was due to fewer items completed by a small number of the respondents in the 'no pen' group than the 'pen' group. On reflection, we agree that this warranted further investigation. Therefore, we have added in more details about the distribution of this outcome into the Results, and now include a post hoc Wilcoxon rank-sum test, which is appropriate when the assumption of linear regression that the residuals are normally distributed is questionable.

We have also expanded on recommendations for future meta-analyses to correct for baseline risk in the Discussion section.

**Competing Interests:** No competing interests