


SWAT II0: Printing the primary outcome on Pink Paper versus standard paper to increase participant engagement to postal questionnaires (PEPPER)

Research Methods in Medicine & Health Sciences
2022, Vol. 3(2) 49–54
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/26320843221074344
journals.sagepub.com/home/rmm



Alexander Ooms¹ , Scott Parsons^{1,2}, Susan Dutton¹, Angela Garrett^{1,2}, Beth Fordham^{1,2}, Caroline Hing³, Sarah E Lamb^{2,4}, and Toby Smith^{2,5}

Abstract

Background: Missing data is a common issue in randomised controlled trials. There is a need to rigorously test means of participant retention. This embedded trial aims to examine the effect on postal response rates of printing a randomised controlled trial's primary outcome on pink versus white paper.

Methods: Our randomised Study Within A Trial (SWAT) was run within a behaviour-change intervention host trial for patients following hip or knee replacements. Participants were randomised to receive the host trial's primary outcome measure printed on either a sheet of pink or white paper within the 11 sheet (21 page) 6-month follow-up questionnaire. The SWAT's primary outcome was host trial primary outcome measure completion. Number of reminders sent, proportion of remaining questions completed and overall questionnaire returns were secondary outcomes.

Results: 176 participants were randomised: 88 received pink paper, 88 white paper. Host trial primary outcome measures were returned by 84.1% (74/88 participants) in the pink paper group and in 90.9% (80/88 participants) in the white paper group (risk ratio, 0.92 (95% CI 0.80, 1.06); $p = .24$). Reminders were sent to 48.9% (43/88 participants) in the pink paper group and in 30.7% (27/88 participants) in the white paper group (risk ratio 1.59 (95% CI 1.09, 2.33); $p = .01$). No other results were statistically significant.

Conclusion: Printing the primary outcome on pink paper does not increase data return. From this small randomised study, there is some evidence that it potentially decreases response and is more burdensome to collect postal data by increasing the necessity for reminders.

Introduction

Missing data on the primary outcome of a randomised controlled trial (RCT) risks introducing bias, reducing the sample size and statistical power of the trial, affecting the validity, reliability and generalisability of findings.^{1–5} The primary outcome is the most important piece of information to be collected in a RCT. Emphasis should therefore be placed on participants completing and returning these data as a priority.⁶

Human perception (attention) and consequent cognition (thoughts), emotion and behaviour can be influenced by colour. From a bottom-up aspect, neurophysiological models of perception suggest coloured objects are attended to for saliency more than grey-scale objects.⁷ From a top-down aspect, people are primed to attend to colours which match our beliefs and expectations.⁸ From these models, we surmise that automatic visual attention is drawn to coloured objects in preference to grey-scale. If people believe an

object to be in grey-scale, for example, a letter or questionnaire, then when it is not, attention is moved to examine

¹Oxford Clinical Trials Research Unit, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

²Centre for Rehabilitation Research in Oxford, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

³University of London St George's Molecular and Clinical Sciences Research Institute, London, UK

⁴College of Medicine and Health Sciences, University of Exeter, Exeter, Devon, UK

⁵Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

Corresponding author:

Alexander Ooms, Oxford Clinical Trials Research Unit, Botnar Research Centre, Windmill Road, Oxford OX3 7LD, UK.

Email: alexander.ooms@csm.ox.ac.uk

this mismatch. A study undertaken of 1.4 million customers, derived from an email host service company showed that using background colour in email messages can result in a higher percentage of emails being opened and read.⁹ Despite there being a large evidence-base on the use of colour in questionnaires, there has been limited research on colouring an individual questionnaire within a battery of outcome measures on the impact of response rates. Having just one outcome measure printed on coloured paper should make that measure more appealing than the rest of the questionnaire printed on white paper.

The objective of this Study Within A Trial (SWAT) embedded within the PEP-TALK RCT was to compare primary outcome response rates printing the primary outcome measure on coloured paper (pink) versus white paper.

Methods

Trial design

A two-arm SWAT was undertaken with a randomisation allocation ratio of 1:1. The SWAT protocol was deposited at the SWAT Repository Store, SWAT 110 (<https://www.qub.ac.uk/sites/TheNorthernIrelandNetworkforTrialsMethodologyResearch/SWATSWARInformation/Repositories/SWATStore/>). The host trial protocol is publicly available.¹⁰ The PEP-TALK trial was an open, multicentre RCT that recruited inactive patients who had received a hip or knee replacement between April 2019 and March 2020 in the United Kingdom. Participants were randomised post-surgery, prior to hospital discharge, to either six, 30-min weekly group-based exercise sessions (control), or the same six weekly, group-based, exercise sessions each preceded by a 30-min cognitive behaviour approach discussion group and three follow-up telephone calls. The primary outcome measure of PEP-TALK was the University of California, Los Angeles (UCLA) Activity Score. No changes were made to the SWAT once started; however, the timings of randomisations were not performed as described. Randomisation into the sub-study was performed just before they were due to be sent their 6-month follow-up questionnaire. The study protocol stated we would randomise into the sub-study shortly after randomisation into the main trial. This deviation was for practical reasons.

Participants

All participants in the SWAT sub-study (named PEPPER) were consented and enrolled in the PEP-TALK trial. Participants included in PEPPER met the inclusion criteria for PEP-TALK. There were no additional inclusion criteria.

Intervention

The aim of the sub-study was to evaluate the effect on response rates of printing the host trial's primary outcome measure (UCLA activity score) at 6 months on pink paper, versus white paper in the printed questionnaire. The selection of pink paper was based on the findings of a meta-analysis which including 12 studies assessing response rate to coloured paper questionnaires. The findings suggested that of the most common colours studied (pink, green, blue, and yellow), pink had the greatest effect on response rate.¹¹ Accordingly pink was selected as the experimental paper colour in this trial.

Both groups' questionnaires were sent to trial participants 6-month post-randomisation. If participants did not return completed questionnaires, steps to address non-response specified in the PEP-TALK protocol¹⁰ were followed in the same way for both groups. If participants did not respond to the initial questionnaire, the trial team sent a reminder questionnaire. If there was no further response, attempts were made to telephone the participant on up to two occasions to remind them to complete the assessments. During the telephone calls, the participants were provided with the opportunity to complete the primary outcome and EQ-5D-5L questionnaires over the telephone along with any complications they may have experienced. The initial questionnaires for both groups received were identical except the primary outcome printed on pink paper for the intervention group.

Study outcomes

The primary outcome measure for PEPPER was UCLA Activity Score completion rate, defined as the raw number of UCLA activity scores returned.

Secondary outcome measures were

- UCLA activity score completion rate adjusting for the number questionnaires received per group
- The proportion of participants reminded to complete and return the questionnaire
- The proportion of the remaining questions in the questionnaire completed
- Overall return rate of 6-month data, irrespective of amount completed and method
- UCLA Activity Score completion rate for the compliant population, i.e. those who receive the intervention they were randomised to.

An additional outcome, not pre-specified in the original study protocol, of 6-month data return rates from initial contact only (i.e. discounting all information gathered through additional follow-up efforts) is analysed as part of this report. This outcome gives the most accurate estimate of

the effect of pink paper on return rates however is considered as supporting evidence only as further follow-up attempts are made in many clinical trials.

Randomisation and Blinding

Participants were randomised (1:1) using a non-stratified, random permuted block randomisation to receive one of the two interventions. Participants were randomised just prior to being sent their 6-month follow-up questionnaire. Participants who withdrew from the host trial's follow-up prior to reaching their 6-month follow-up were not randomised. Randomisations were performed by the host trial manager via a computer generated randomisation list that was held on a bespoke randomisation server maintained by the Oxford Clinical Trial Research Unit. They were also responsible for delivering the SWAT intervention.

Trial participants were blinded to the nature and objectives of the PEPPER sub-study. The trial team was not blinded to the allocation.

Statistical Methods

No formal power calculation was undertaken, as is usual in SWAT research.¹² The sample size was constrained by the number of participants in the host trial.

All analyses were on an intention-to-treat basis, i.e. analysed according to their allocated treatment irrespective of the intervention received. A chi-square test was used to assess all outcomes, absolute differences, risk ratios (RR) and associated 95% confidence intervals (CI) were reported. Analyses were undertaken using R version 4.1.0 (2021-05-18).¹³

Results

Study population, recruitment and compliance

A total of 224 participants were randomised in the PEP-TALK trial. Of these, 176 (78.6%) were randomised into PEPPER (88 pink paper/88 white paper) (Figure 1). Forty-eight participants were not included in PEPPER; 19 were missed due to having already passed the 6-month time-point

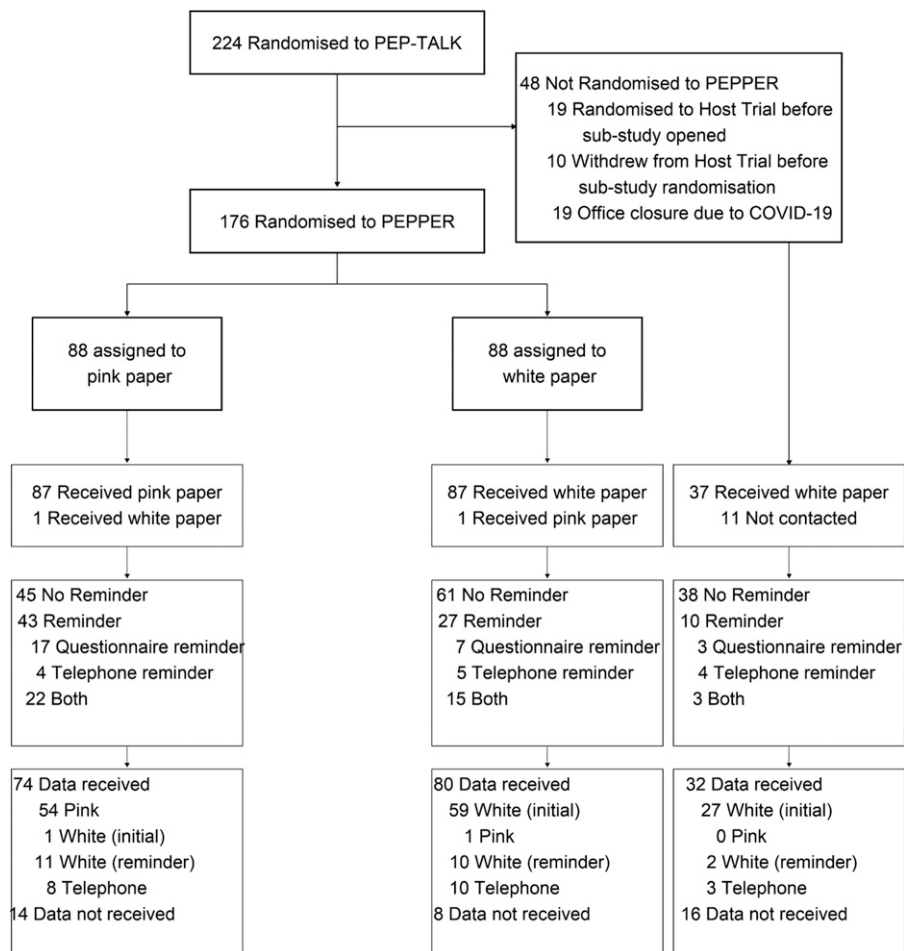
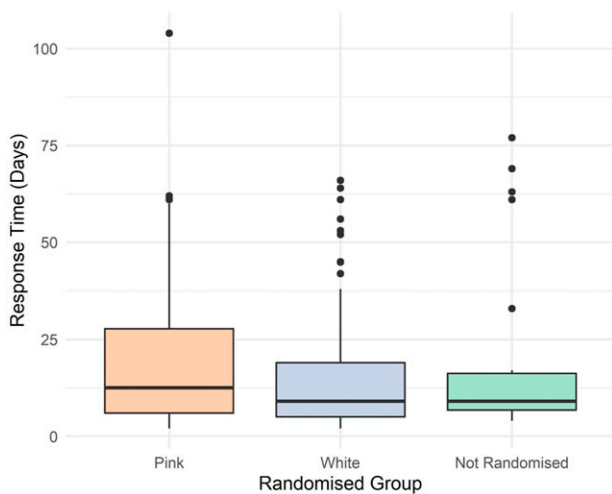


Figure 1. Flow diagram of participants.

Table 1. Baseline characteristics.

	Pink (<i>n</i> = 88)	White (<i>n</i> = 88)	Not randomised (<i>n</i> = 48)
Age, years	67.9 (9.3)	68.6 (8.0)	68.8 (8.8)
Sex			
Male	30 (34.1)	32 (36.4)	21 (43.8)
Female	58 (65.9)	56 (63.6)	27 (56.2)
BMI, kg/m ²	30.8 (5.9)	30.6 (5.6)	31.5 (5.6)
Operation type			
Hip replacement	48 (54.5)	39 (44.3)	21 (43.8)
Knee replacement	40 (45.5)	49 (55.7)	27 (56.2)
CCI	2.8 (1.3)	2.9 (1.0)	3.1 (1.7)

Data are *n* (%) for categorical variables and mean (SD) for continuous outcomes unless stated otherwise. BMI = body mass index; CCI = Charlson Comorbidity Index.

**Figure 2.** Time taken for 6-month questionnaires to be returned.

when the SWAT opened, 19 because of the COVID-19 pandemic and March 2020 lockdown, when questionnaires were not being sent by post, and 10 due to early withdrawal from the host trial. Recruitment to PEPPER began 17 December 2019 and ended 05 August 2020, approximately 4 months after the final participant was recruited to PEP-TALK on 27 March 2020. There was a pause in PEPPER randomisations between 18 March 2020 and 09 July 2020, caused by COVID-19-associated trial office closure.

In total, 186 participants (including those not randomised into PEPPER) supplied data at 6 months giving an overall response rate of 83.0%. Of those not recruited to PEPPER who were sent questionnaires, 86.5% (32/37) responded. There were 90.9% (80/88) responses in the white group and 84.1% (74/88) in the pink group.

Both randomised groups and those not randomised into PEPPER had comparable baseline characteristics (Table 1). There were approximately 10% more hip replacements in the pink paper group (54.5%) compared to white (44.3%) or not randomised (43.8%).

Of the 224 randomised into PEP-TALK, 44 (19.6%) responded either via telephone or reminder questionnaire. Almost half (48.9%) of the participants in the pink group were sent some form of reminder. This was higher than the 27 (30.7%) in the white group.

The pink group took slightly longer to respond (median 12; interquartile range (IQR): 6–28), compared to the white, (median 9; IQR: 5–19), and not randomised groups (Figure 2).

Outcome results

Table 2 presents a summary of the analysed sub-study results. The analysis of adjusted UCLA Activity Score completion rate could not be performed as all participants returned data on the primary outcome if they returned the 6-month questionnaire. The analysis on the compliant population was also not performed as there was insufficient distinction between the primary intention-to-treat and compliant populations for this to be meaningful. All participants randomised to PEPPER received at least one postal questionnaire. No PEPPER participants were only contacted by telephone call and there was only one crossover in each group.

There was no statistically significant difference between the two randomised groups for the primary outcome response rate. For the 88 participants randomised to both groups, there were 84.1% (74/88) and 90.9% (80/88) returns for the pink and white groups respectively, RR 0.92, 95% CI (0.83, 1.03), $p = .17$. There was a statistically significantly higher number of participants sent a reminder in the pink group, RR 1.59, 95% CI (1.09, 2.33), $p = .01$, suggesting a higher risk of needing to contact participants again after sending out the initial questionnaire. The result from Figure 2 in combination with the increased need to send reminders to the pink group indicates that participants in the white group were easier to follow-up and provided their responses more promptly. No other result was statistically significant.

Table 2. PEPPER outcome results table.

	Pink	White	Absolute difference	Risk ratio (95% CI)	p-value
UCLA completion rate (primary outcome)	74/88	80/88	−0.07	0.92 (0.83, 1.03)	.17
Reminders sent	43/88	27/88	0.18	1.59 (1.09, 2.33)	.01
All secondary PROMs answered	43/88	48/88	−0.06	0.90 (0.67, 1.19)	.45
Six-month questionnaire returned	74/88	80/88	−0.07	0.92 (0.83, 1.03)	.17
Returns from initial questionnaire ^a	55/88	60/88	−0.06	0.92 (0.74, 1.14)	.17

There were no harms caused by this SWAT.

^aNot pre-specified. PROM=Patient Reported Outcome Measure.

Discussion

The results from the PEPPER study indicate that there was no evidence that printing the primary outcome measure on pink paper was superior to white paper on a follow-up questionnaire in terms of primary outcome measure response rate. There was a statistically significant difference in the number of participants needing to be sent a reminder in the pink paper group. The results indicate that these respondents also took longer to respond. These findings suggest that the pink paper group were more difficult to follow-up and required more effort from the trial team to collect their data. Therefore, it is possible that receiving a pink primary outcome had the opposite effect of the hypothesised reaction. This led to participants needing to be contacted further as colour did not stimulate a response, but discouraged one.

Many factors are associated with trial attrition rates, including study centre and age.¹⁴ Data collection for this trial was impacted by COVID-19 restrictions. This made data collection more difficult, possibly affecting the results, although this effect should be the same across both randomised groups. Repeated use of this SWAT design would improve external validity as the populations of subsequent host trials could vary widely. Alternative colours could be explored as well as using coloured web page backgrounds for trials with virtual follow-up. Some people with certain types of deuteranomaly (colour-blindness) like to use pink/pastel paper when reading.¹⁵ The estimated incidence of deuteranomaly is 4.5%.¹⁶ Whilst there is some rationale to think participants with these conditions would be more likely to respond to the pink paper, a treatment-effect estimate for these groups cannot be ascertained from this data as the sample size is not large enough and we did not collect information on deuteranomaly as part of this study. Given the low cost of the intervention and applicability to a wide range of host trials, it is therefore recommended that this SWAT could be repeated in many other trials. Following this, these results maybe combined in a meta-analysis to give a more accurate treatment-effect estimate.

The participant population was limited by the eligibility criteria of the host trial, possibly making the results not

generalisable to more diverse populations. There was also a relatively small sample size. For an intervention where one would expect to see a modest treatment effect, recruiting 176 participants is likely not sufficient to get an accurate estimate of the true treatment effect. The trial was blinded to participants because informing participants that they are in a study examining questionnaire response and completion rates may confound the results, however the study team were not blinded. This may have introduced an unconscious bias but this seems unlikely as the trial team followed steps outlined in the host trial protocol when pursuing this data. The trial also did not randomise all host trial participants as it was implemented part way through recruitment.

In conclusion, based on the findings of this study, printing the primary outcome on pink paper does not increase response rates. From this small randomised study, there is some evidence to suggest that it potentially decreases response rate and makes it more burdensome to collect postal data by increasing the necessity for reminders to be sent.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This SWAT was funded by the PROMETHEUS project, University of York (Grant number: MR/R013748/1).

Host trial registration

ISRCTN29770908. This SWAT is registered with the Northern Ireland Methodology Hub's SWAT repository. (<https://www.qub.ac.uk/sites/TheNorthernIrelandNetworkforTrialsMethodologyResearch/SWATSWARInformation/Repositories/SWATStore/>, accessed October 2021)

ORCID iD

Alexander Ooms  <https://orcid.org/0000-0003-1877-8323>

References

1. Edwards P, Roberts I, Clarke M, et al. Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002; 324(7347): 1183.
2. Brueton 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.
3. Fewtrell MS, Kennedy K, Singhal A, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? *Arch Disease Childhood* 2008; 93(6): 458–461.
4. Schulz KF and Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781–785.
5. Waller R. Principles of exposure measurement in epidemiology. *Occup Environ Med* 1994; 51(11): 790. Protocol Hughes-Morley and Torgerson Improving retention using text message prompts v3 05 v3 05.03.2016 Page 8 of 10.
6. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *New Engl J Med* 2012; 364(14): 1355–1360.
7. Koch C, Ullman S Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol* 1985; 4: 219–227.
8. Folk CL, Remington RW, Johnston JC. Involuntary covert orienting is contingent on attentional control settings. *J Exp Psychol Hum Perception Perform* 1992; 18: 1030–1044. DOI: [10.1037/0096-1523.18.4.1030](https://doi.org/10.1037/0096-1523.18.4.1030).
9. Zviran M, Te'eni D, Gross Y Does color in email make a difference? *Commun ACM* 2006; 49: 94–99.
10. Smith TO, Parsons S, and Fordham B, et al. Behaviour change physiotherapy intervention to increase physical activity following hip and knee replacement (PEP-TALK): study protocol for a pragmatic randomised controlled trial. *BMJ Open* 2020; 10: e035014. DOI: [10.1136/bmjopen-2019-035014](https://doi.org/10.1136/bmjopen-2019-035014).
11. Hartley J, Rutherford A The effects of using coloured paper to boost response-rates to surveys and questionnaires. *J Tech Writing Commun* 2003; 33(1): 29–40.
12. Treweek S, Bevan S, Bower P, et al. Trial forge guidance 1: what is a study within a trial (SWAT)? *Trials* 2018; 19: 139. DOI: [10.1186/s13063-018-2535-5](https://doi.org/10.1186/s13063-018-2535-5).
13. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019.
14. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun* 2018; 11: 156–164. DOI: [10.1016/j.conctc.2018.08.001](https://doi.org/10.1016/j.conctc.2018.08.001). Published
15. Royal National Institute of Blind People. *How to Choose Colour and Contrast for Printed Materials that Benefits People with Sight Problems*, <http://scdg.org.uk/wp-content/uploads/2019/02/Colour-information-with-header-sheet-2.pdf> (2015, accessed 10 February 2021).
16. NHS. *Colour Vision Deficiency*, <https://www.nhs.uk/conditions/colour-vision-deficiency/> (accessed 10 February 2021).