PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effect of stamped reply envelopes and timing of newsletter delivery on response rates of mail survey: A randomised controlled trial in a prospective cohort study
AUTHORS	Wakabayashi, Chizuko ; Hayashi, Kunihiko; Nagai, Kazue; Sakamoto, Naoko; Iwasaki, Yoko

VERSION 1 - REVIEW

REVIEWER	Dr Katrina Lavelle Research Fellow School of Nursing , Midwifery & Social Work The University of Manchester United Kingdom I have no conflict of interest.
REVIEW RETURNED	31-Mar-2012

THE STUDY	There are several outcome measures. Identifying one primary outcome measure and hypothesis would help to clarify the main objective of this research (secondary outcomes can be included but it is helpful if they are identified as such). A consort diagram would really help to describe participants and methods more adequately. The actual process of randomisation needs more detail. Who undertook randomisation, where, using what software and unique identifiers. I would have expected to see a table of baseline characteristics of each group (with testing for significant differences between the groups) to check that the randomisation process worked properly. Not sure how useful it was to adjust for the effects of potential confounders as the randomisation process should have evenly distributed these between groups. Difficult to judge if abstract accurate as it refers to results which in the results section refers to Model 1 & 2 which are not in the tables. Last sentence in results section of abstract is not clear. No power calculation (based on primary hypothesis) or explanation for its absence. Although an estimation of cost analysis in discussion, the paper would be strengthened by an actual cost analysis of the presented data, particularly as cost is mentioned as a key issue several times in the paper. Cost per item returned and marginal costs could easily be derived (see Lavelle et al 2008 for example - see reference below). A key reference is missed; Edward et al's 2007 Cochrane review - reference is given below but check the Cochrane website to see if updated.

	Further justification for adjusting for 'potential confounding variables' needed from previous literature. Some of the explanatory variables would be correlated with each other e.g. menopausal status and age - how was multicolinearity checked for and dealt with? Lavelle K, Todd, C, Campbell M. Do postage stamps verses pre-paid envelopes increase responses to patient mail surveys?: A randomised controlled trial. BMC Health Services Research. 2008, 8: 113 Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R,Kwan I, Cooper R: Methods to influence response to postal questionnaires (Cochrane Methodology Review). The Cochrane Library 2007): Oxford [http://mrw.interscience.wiley.com/cochrane/clsysrev/articles/MR000008/frame.html].
RESULTS & CONCLUSIONS	No Model 1 or 2 in tables, figures or supplementary materials yet referred to in the results section. Consort diagram needed. Main conclusions warranted but cost estimation should be verified by actual cost analysis of presented data. Main conclusion should be discussed in light of Edwards et al 2007 Cochrane review - see reference above.

REVIEWER	Professor Elaine McColl Professor of Health Services Research Institute of Health & Society Newcastle University United Kingdom No conflicts of interest
REVIEW RETURNED	14-Apr-2012

THE STUDY	1. The design used here, where participants were simultaneously randomised to two interventions (business reply vs stamped return envelopes; newsletter with initial mailing vs newsletter with reminder) is a 2x2 factorial design; this should be specified explicitly. 2. It is not clear why reminders were sent to all participants, regardless of whether or not they had already responded. This is not likely to reflect normal practice in survey administration. 3. The analytical strategy seems over-complicated. It is unclear to me why a survival analysis was employed for the initial analysis of pre-reminder response rates. A straightforward reporting of % response rates and comparison (using chi-squared test) of whether the individual had responded or not at 6 weeks by return postage type (business reply vs stamped - i.e. groups 1 and 2 vs groups 3 and 4) and by inclusion of newsletter with inital letter (groups 1 and 3 vs groups 2 and 4) would be much simpler to present and understand. Similarly, for the post-reminder analysis, comparisons of the overall response rates (i.e. those responding by 12 weeks divided by all those initially receiving a questionnaire) and the post-reminder response rates (those responding between 6 and 12 weeks divided by those receiving a reminder) could be conducted

	and reported in the same way. At present, nowhere in the paper can I find a straightforward presentation of the response rates in the 4 groups, or for the two main effects in the factorial design, at 6 and 12 weeks respectively. 4. It is not clear why response rates at week 7 (rather than week 12) were used to compare the effect of the inclusion of the newsletter with the reminder. Bringing my 2nd and 3rd points together, a comparison of response (yes/no) at 12 weeks by groups (2&4 vs 1&3), with the denominator being those who had not responded by 6 weeks, would be the best approach to examining the effect of enclosing a newsletter with the reminder, since those who had responsed prior to the reminder would be disregarded. Some references are apparently missing, in the sense that they appear in the text as (ref) - see for example, page 3, lines 17 and 19.
RESULTS & CONCLUSIONS	The arguments in the discussion section regarding the cost-effectiveness of business-reply paid envelopes vs stamped envelopes are hard to follow, mainly because the relative costs of the two approaches are not given (here in the UK, the cost of a business reply envelope is slightly higher than that of a stamped envelope, but is only incurred if the envelope is used - it is not clear to me whether the same is true in Japan, or what the cost differential is).

VERSION 1 – AUTHOR RESPONSE

Reviewer: Dr. Katrina Lavelle

1) There are several outcome measures. Identifying one primary outcome measure and hypothesis would help to clarify the main objective of this research (secondary outcomes can be included but it is helpful if they are identified as such).

As pointed out by the reviewer, we changed the Measurement of Methods as "'Primary outcome measure was cumulative response proportion at 12 weeks after mailing out the questionnaire. Secondary outcome measure was cumulative response at the end of 6 weeks after initial mailing, just before delivering the reminder letters."

2) A consort diagram would really help to describe participants and methods more adequately.

We appreciate the valuable comment. We changed 'Sample and Data' for 'Participants' and 'Intervention', and included setting and eligibility criteria, according to the CONSORT guideline. We made a new figure as a CONSORT diagram and included it in Results and we removed table 1 instead.

3) The actual process of randomisation needs more detail. Who undertook randomisation, where, using what software and unique identifiers.

We added the process of randomization in methods as follows. "When the participants registered at baseline survey, the sequential unique 7-digit ID numbers were assigned randomly. According to the ID numbers, participants were allocated to the four groups. The allocated group number for each participant was the remainder when the ID number was divided by four.''

4) I would have expected to see a table of baseline characteristics of each group (with testing for significant differences between the groups) to check that the randomisation process worked properly. Not sure how useful it was to adjust for the effects of potential confounders as the randomisation process should have evenly distributed these between groups.

As the reviewer pointed out, we added descriptions about comparisons of characteristics between four groups in the Methods and Results. We made a new table for comparison of characteristics.

There were no variables that showed statistically significant differences between groups. We could assume that the randomisation process worked properly. Although there were slight differences among four groups in age at the survey (p=0.081) and menopausal status (p=0.066), no significant differences were found between any pairs of groups

5) Difficult to judge if abstract accurate as it refers to results which in the results section refers to Model 1 & 2 which are not in the tables.
We removed the expressions of Model-1 and Model-2. We apologize for the confusing expressions.

6) Last sentence in results section of abstract is not clear.
We changed the results section of abstract and of body text extensively, according to the comments of Reviewer-2. “The cumulative response at twelve weeks were 58.3% for Group-1, 54.1% for Group-2, 60.5% for Group-3, and 56.7% for Group-4, and these proportions significantly differed among the groups (p=0.001). The odds of the response was higher for stamped envelopes than for business-reply envelopes (OR [95%CI] = 1.10 [1.00-1.21]). The odds was higher for newsletter delivery with initial mailing than for with reminder letters (1.18 [1.07-1.29]). The response in first six weeks for stamped envelope groups was significantly higher than for business-reply envelope groups (p=0.047). Although the response in six weeks for women received the newsletter with initial mailing was lower than for women who did not, the proportions did not differ significantly (p=0.291).”

7) No power calculation (based on primary hypothesis) or explanation for its absence.
We added the explanation for its absence in Intervention of Methods. “We did not perform any power calculations based on primary hypothesis of this study.”

8) Although an estimation of cost analysis in discussion, the paper would be strengthened by an actual cost analysis of the presented data, particularly as cost is mentioned as a key issue several times in the paper. Cost per item returned and marginal costs could easily be derived (see Lavelle et al 2008 for example - see reference below).
We appreciate your advice. We showed postal costs for business-reply envelopes and for stamped envelopes according to two scenarios in a new table (table 3), referring the BMC Health Services Research paper.

9) A key reference is missed; Edward et al's 2007 Cochrane review - reference is given below but check the Cochrane website to see if updated.
We appreciate for the reference. We included the updated review paper.

10) Further justification for adjusting for 'potential confounding variables' needed from previous literature.
We added three papers including the Cochrane review. We added the sentence, “These variable included factors previously studied10,11,12 and reproductive health-related issues in women”, in the measurements of methods.

10. Edwards PJ, Roberts I, Clarke MJ, Diguiseppi C, Wentz R , et al. Methods to increase response to postal and electronic questionnaires. Cochrane Database Syst Rev 2009 Jul 8; (3): MR000008.
11. Barton J, Bain C, Hennekens CH, Rosner B, Belanger C, et al.: Characteristics of respondents and non-respondents to a mailed questionnaire. Am J Public Health 1980; 70 (8): 823-5.
12.Steffen AD, Kolonel LN, Nomura AM, Nagamine FS, Monroe KR, et al. The effect of multiple mailings on recruitment: The multiethnic cohort. Cancer Epidemiol Biomakers Prev 2008; 17(2): 447-54.

11) Some of the explanatory variables would be correlated with each other e.g. menopausal status and age - how was multicolinearity checked for and dealt with?

Yes. Those variables were correlated and we had done stepwise model selection analyses. As you pointed out, there are multicolinearity problems and the variables such as participation of previous survey seem to be not confounding factors but intermediate variables. Therefore, we showed the results of odds ratios only for age-adjusted, and removed the results of multivariate adjusted odds ratios. There was slight difference among four groups in age at the survey (p=0.081) and it is worthwhile the effects of interventions was examined by age-adjusted odds ratios. The effect of menopausal status was significant even after adjusted by age.

12) No Model 1 or 2 in tables, figures or supplementary materials yet referred to in the results section.
We removed the words of Model-1 and Model-2. Again, we apologize for the confusing expressions.

13) Consort diagram needed.
As the reviewer pointed out, we made a new figure as a CONSORT diagram and put it in Results (figure 1).

14) Main conclusions warranted but cost estimation should be verified by actual cost analysis of presented data.
We added the postal costs for business-reply envelopes and for stamped envelopes according to two scenarios as a new table (table 3)

15) Main conclusion should be discussed in light of Edwards et al 2007 Cochrane review - see reference above.
We refer the paper along with ref.3(Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, et al. Increasing response rates to postal questionnaires: systematic review. BMJ 2002; 324(7347):1183.) in discussion.

16) No reference is made to ethics processes in the article.
We added following descriptions about ethics process in Methods. “The baseline survey includes 49,927 responses from participants in Japan. Among them, 14,844 women signed an informed consent form and participated in the follow-up survey. Institutional review boards of Gunma University and the National Institute of Public Health reviewed and approved the JNHS study protocol.”

Reviewer: Professor Elaine McColl
1) The design used here, where participants were simultaneously randomised to two interventions (business reply vs stamped return envelopes; newsletter with initial mailing vs newsletter with reminder) is a 2x2 factorial design; this should be specified explicitly.
We appreciate the comment. As the reviewer pointed out, we changed the Design of Abstract to “A randomized controlled trial in a prospective cohort study with a 2x2 factorial design of two interventions, types of reply envelope and timing of newsletter delivery.” We also added the phrase ‘a 2x2 factorial design’ in Objectives and Intervention of Methods in body text. We added “The participants were simultaneously randomised to two interventions; type of return envelope (business reply vs stamped return envelopes) and timing of Newsletter delivery (newsletter with initial mailing vs newsletter with reminder).” in measurements section. Instead, we removed “Two relevant test variables in this analysis were the type of return envelope and the timing of the newsletter delivery. --- In contrast, if the newsletter was mailed with the reminder letter, then the variable of newsletter_reminder was coded as 1; otherwise (i.e., newsletters were already received with initial letters), it was coded as 0.”(page 2, line 56 to page 3, line 13) from Measurements section.

2) It is not clear why reminders were sent to all participants, regardless of whether or not they had already responded. This is not likely to reflect normal practice in survey administration.
The letter included not only reminding message for non-responders but also thank-you message for

women who had already responded. We do send the thank-you and reminder letter as usual practice, especially when the duration between initial mailing and reminder mailing is a short period of time. It is so helpful to reduce administrative troubles of distinguishing responders and non-responders in a large study population.

3) The analytical strategy seems over-complicated. It is unclear to me why a survival analysis was employed for the initial analysis of pre-reminder response rates. A straightforward reporting of % response rates and comparison (using chi-squared test) of whether the individual had responded or not at 6 weeks by return postage type (business reply vs stamped - i.e. groups 1 and 2 vs groups 3 and 4) and by inclusion of newsletter with inital letter (groups 1 and 3 vs groups 2 and 4) would be much simpler to present and understand. Similarly, for the post-reminder analysis, comparisons of the overall response rates (i.e. those responding by 12 weeks divided by all those initially receiving a questionnaire) and the post-reminder response rates (those responding between 6 and 12 weeks divided by those receiving a reminder) could be conducted and reported in the same way. At present, nowhere in the paper can I find a straightforward presentation of the response rates in the 4 groups, or for the two main effects in the factorial design, at 6 and 12 weeks respectively.
We appreciate your valuable advice for the analytical strategy. I changed the statistical analysis as follows; “Characteristics of participants were compared between groups using ANOVA and chi-square test to check the relevance of randomization process. Before examining main effect of two interventions, type of return envelope (business reply vs stamped) and timing of newsletter provision (with initial mailing vs. with reminder) on cumulative response proportion at 6 and 12 weeks after initial mailing, the interaction of these two interventions was tested by logistic regression model. The main effects of the interventions were tested by chi-square test. In order to examine the factors affecting the responses in 12 weeks after initial mailing, logistic regression models were used to estimate unadjusted and age-adjusted odds ratios (ORs) and their 95% confidence intervals (95%CIs).”

4) It is not clear why response rates at week 7 (rather than week 12) were used to compare the effect of the inclusion of the newsletter with the reminder. Bringing my 2nd and 3rd points together, a comparison of response (yes/no) at 12 weeks by groups (2&4 vs 1&3), with the denominator being those who had not responded by 6 weeks, would be the best approach to examining the effect of enclosing a newsletter with the reminder, since those who had responsed prior to the reminder would be disregarded.
Thank you for the comment. We changed the analytical strategy as mentioned above (point# 3). The pre- and post- analyses were confusing.

5) Some references are apparently missing, in the sense that they appear in the text as (ref) - see for example, page 3, lines 17 and 19.
We apologize for the confusing expressions, [ref]s. The [ref.] did not indicate a reference paper, but a reference category in logistic regression models. For example, the category of ‘register nurse [ref.]’ was set a reference category to estimate odds ratios of other categories, “public health nurse”, “midwife”, and “licensed nurse”. We removed all the [ref.]s from the text.

6) The arguments in the discussion section regarding the cost-effectiveness of business-reply paid envelopes vs stamped envelopes are hard to follow, mainly because the relative costs of the two approaches are not given (here in the UK, the cost of a business reply envelope is slightly higher than that of a stamped envelope, but is only incurred if the envelope is used - it is not clear to me whether the same is true in Japan, or what the cost differential is).
We added the postal costs for business-reply envelopes and for stamped envelopes according to two scenarios as a new table (table 3). I hope the table helps the augment on the postal costs of mailing survey in Japan.

Sincerely yours,
Kunihiko Hayashi, PhD
Professor of Epidemiology and Biostatistics
School of Health Sciences, Gunma University

VERSION 2 – REVIEW

REVIEWER	Elaine McColl Director, Newcastle CTU Newcastle University United Kingdom No conflicts of interest
REVIEW RETURNED	24-Jul-2012

THE STUDY	In general,m the authors have responded well to the feedback on their initial submission. However, I recommend attention to the following points. 1 - please show the 6 week response patterns in the CONSORT diagram 2 - I find the revised sentence 'We did not perform any power calculations based on the primary hypothesis of this study' a little confusing. I suggest 'Sample size was determined by the size of the available cohort. With an expected number of 3469 (i.e. 6,938/2) per group, and a reference response rate of 60%, for 80% power and 5% significance, the detectable difference in response rate was +/- 3.3%'. 3 - While I agree with the recommendation of the other reviewer regarding the presentation of the table (now Table 2) of baseline data, I disagree with the recommendation of performing statistical tests regarding baseline comparability. Amongst triallists, this is not recommended practice - see for example Senn S (1994). Testing for baseline balance in clinical trials. Statistics in Medicone, 13 (17), 1715-26. I therefore recommend removal of the left-most column in table 2, and the commentary on the statistical significance of the baseline comparison on page 3.

VERSION 2 – AUTHOR RESPONSE

> Reviewer: Elaine McColl
> In general,m the authors have responded well to the feedback on their
> initial submission. However, I recommend attention to the following points.
> 1 - please show the 6 week response patterns in the CONSORT diagram

As pointed out by the reviewer, we added the 6 week responses in the diagram.

> 2 - I find the revised sentence 'We did not perform any power calculations
> based on the primary hypothesis of this study' a little confusing.
> I suggest 'Sample size was determined by the size of the available cohort.
> With an expected number of 3469 (i.e. 6,938/2) per group, and a
> reference response rate of 60%, for 80% power and 5% significance, the

> detectable difference in response rate was +/- 3.3%'.

We appreciate the valuable suggestion. We replaced the sentence about sample size according to the suggestion.

> 3 - While I agree with the recommendation of the other reviewer regarding
> the presentation of the table (now Table 2) of baseline data, I disagree
> with the recommendation of performing statistical tests regarding baseline
> comparability. Amongst triallists, this is not recommended practice - see
> for example Senn S (1994). Testing for baseline balance in clinical trials.
> Statistics in Medicone, 13 (17), 1715-26. I therefore recommend removal
> of the left-most column in table 2, and the commentary on the statistical
> significance of the baseline comparison on page 3.

According to your recommendation, we removed the column for p-value from the table and the sentence, “Although there were slight differences among four groups in age at the survey (p=0.081) and menopausal status (p=0.066), no significant differences were found between any pairs of groups” on page 3.

Thank you again for the valvaluable comments.

Sincerely yours,
Kunihiko Hayashi, PhD
Professor of Epidemiology and Biostatistics
School of Health Sciences,
Gunma University