Increasing response rates to postal questionnaires: systematic review

Phil Edwards, Ian Roberts, Mike Clarke, Carolyn DiGuiseppi, Sarah Pratap, Reinhard Wentz, Irene Kwan

Abstract

**Objective** To identify methods to increase response to postal questionnaires.

**Design** Systematic review of randomised controlled trials of any method to influence response to postal questionnaires.

**Studies reviewed** 292 randomised controlled trials including 258 315 participants

**Intervention reviewed** 75 strategies for influencing response to postal questionnaires.

**Main outcome measure** The proportion of completed or partially completed questionnaires returned.

**Results** The odds of response were more than doubled when a monetary incentive was used (odds ratio 2.02; 95% confidence interval 1.79 to 2.27) and almost doubled when incentives were not conditional on response (1.71; 1.29 to 2.26). Response was more likely when short questionnaires were used (1.86; 1.55 to 2.24). Personalised questionnaires and letters increased response (1.16; 1.06 to 1.28), as did the use of coloured ink (1.39; 1.16 to 1.67). The odds of response were more than doubled when the questionnaires were sent by recorded delivery (2.21; 1.51 to 3.25) and increased when stamped return envelopes were used (1.26; 1.13 to 1.41) and questionnaires were sent by first class post (1.12; 1.02 to 1.23). Contacting participants before sending questionnaires increased response (1.54; 1.24 to 1.92), as did follow up contact (1.44; 1.22 to 1.70) and providing non-respondents with a second copy of the questionnaire (1.41; 1.02 to 1.94). Questionnaires designed to be of more interest to participants were more likely to be returned (2.44; 1.99 to 3.01), but questionnaires containing questions of a sensitive nature were less likely to be returned (0.92; 0.87 to 0.98). Questionnaires originating from universities were more likely to be returned than questionnaires from other sources, such as commercial organisations (1.31; 1.11 to 1.54).

**Conclusions** Health researchers using postal questionnaires can improve the quality of their research by using the strategies shown to be effective in this systematic review.

**Introduction**

Postal questionnaires are widely used to collect data in health research and are often the only financially viable option when collecting information from large, geographically dispersed populations. Non-response to postal questionnaires reduces the effective sample size and can introduce bias. As non-response can affect the validity of epidemiological studies, assessment of response is important in the critical appraisal of health research. For the same reason, the identification of effective strategies to increase response to postal questionnaires could improve the quality of health research. To identify such strategies we conducted a systematic review of randomised controlled trials.

**Methods**

**Identification of trials**

We aimed to identify all randomised controlled trials of strategies to influence the response to a postal questionnaire. Eligible studies were not restricted to medical surveys and included any questionnaire topic in any population. Studies in languages other than English were included. Strategies requiring telephone contact were included, but strategies requiring home visits by investigators were excluded for reasons of cost. We searched 14 electronic bibliographical databases (table 1). Two reviewers independently screened each record for eligibility by examining titles, abstracts, and keywords. Records identified by either reviewer were retrieved. We searched the reference lists of relevant trials and reviews, and two journals in which the largest number of eligible trials had been published (Public Opinion Quarterly and American Journal of Epidemiology). We contacted authors of eligible trials and reviews to ask about unpublished trials. Reports of potentially relevant trials were obtained, and two reviewers assessed each for eligibility. We estimated the sensitivity of the combined search strategy (electronic searching and manual searches of reference lists) by comparing the trials identified by using this strategy with the trials identified by manually searching journals. We used ascertainment intersection methods to estimate the number of trials that may have been missed during screening.2
Data extraction and outcome measures

Two reviewers independently extracted data from eligible reports by using a standard form. Disagreements were resolved by a third reviewer. We extracted data on the type of intervention evaluated, the number of participants randomised to intervention or control groups, the quality of the concealment of participants’ allocation, and the types of participants, materials, and follow up methods used. Two outcomes were used to estimate the effect of each intervention on response: the proportion of completed or partially completed questionnaires returned after the first mailing and the proportion returned after all follow up contacts had been made. We wrote to the authors of reports when these data were missing or the methods used to allocate participants were unclear (for example, where reports said only that participants were “divided” into groups).

Interventions were classified and analysed within distinct strategies to increase response. In trials with factorial designs, interventions were classified under two or more strategies. When interventions were evaluated at more than two levels (for example, highly, moderately, and slightly personalised questionnaires), we combined the upper levels to create a dichotomy. To assess the influence of a personalised questionnaire on response, for example, we compared response to the least personalised questionnaire with the combined response for the moderately and highly personalised questionnaires.

Data analysis and statistical methods

We used Stata statistical software to analyse our data. For each strategy, we estimated pooled odds ratios in a random effects model. We calculated 95% confidence intervals and two sided P values for each outcome. Selection bias was assessed by using Egger’s weighted least squares regression and Begg’s rank correlation test and funnel plot. Heterogeneity among the trials’ odds ratios was assessed by using a χ² test at a 5% significance level. In trials of monetary incentives, we specified a priori that the amount of the incentive might explain any heterogeneity between trial results. To investigate this, we used regression to examine the relation between response and the current value of the incentive in US dollars. When the year of the study was not known, we used the average delay between year of study and year of publication for other trials (three years). We also specified a priori that, in trials of questionnaire length, the number of pages used might explain any heterogeneity between trial results, and to investigate this, the odds of response were regressed on the number of pages.

Results

We identified 292 eligible trials including a total of 258,315 participants that evaluated 75 different strategies for increasing response to postal questionnaires. The average number of participants per trial was 1091 (range 39-10,047). The trials were published in 251 reports—80 (32%) in medical, epidemiological, or health related journals, 58 (23%) in psychological, educational, or sociological journals, 105 (42%) in marketing, business, or statistical journals, and 8 (3%) in engineering journals or dissertations, or they had not yet been published (see Appendix A).

All tests for selection bias were significant (P<0.05) in five strategies: monetary incentives, varying length of questionnaire, follow up contact with non-respondents, saying that the sponsor will benefit if participants return questionnaires, and saying that society will benefit if participants return questionnaires. Tests were not possible in 15 strategies where fewer than three trials were included. The method of randomisation was not known in most of the eligible trials. Where information was available, the quality of the concealment of participants’ allocation was poor in 30 trials and good in 12 trials. The figure shows the pooled odds ratios and 95% confidence intervals for...
<table>
<thead>
<tr>
<th>Strategy</th>
<th>No of trials (No of participants)</th>
<th>Odds ratio (95% CI)</th>
<th>P value for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incentives</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monetary incentive v no incentive</td>
<td>49 (46 474)</td>
<td>2.02 (1.79 to 2.27)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Incentive with questionnaire v incentive on return</td>
<td>10 (13 713)</td>
<td>1.71 (1.29 to 2.26)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Non-monetary incentive v no incentive</td>
<td>45 (44 708)</td>
<td>1.19 (1.11 to 1.28)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Length</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shorter v longer questionnaire</td>
<td>40 (40 669)</td>
<td>1.86 (1.55 to 2.24)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Appearance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown envelope v white</td>
<td>2 (5311)</td>
<td>1.52 (0.67 to 3.44)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Coloured ink v standard</td>
<td>1 (3540)</td>
<td>1.39 (1.16 to 1.67)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Folder or booklet v stapled pages</td>
<td>2 (1845)</td>
<td>1.17 (0.94 to 1.45)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>More personalised v less personalised</td>
<td>38 (39 210)</td>
<td>1.16 (1.06 to 1.28)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Identifying feature on return v none</td>
<td>7 (4014)</td>
<td>1.08 (0.78 to 1.51)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Coloured questionnaire v white</td>
<td>8 (14 797)</td>
<td>1.06 (0.99 to 1.14)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Delivery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded delivery v standard</td>
<td>6 (2127)</td>
<td>2.21 (1.51 to 3.25)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Stamped returned envelope v business reply or franked</td>
<td>14 (38 259)</td>
<td>1.26 (1.13 to 1.41)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Questionnaire sent to work address v home address</td>
<td>2 (1140)</td>
<td>1.16 (1.09 to 1.22)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>First class outward mailing v other class</td>
<td>1 (7370)</td>
<td>1.12 (1.02 to 1.23)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Pre-paid return envelope v not pre-paid</td>
<td>4 (4904)</td>
<td>1.09 (0.71 to 1.76)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Stamped outward envelope v franked</td>
<td>6 (13 964)</td>
<td>0.95 (0.88 to 1.03)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Commemorative stamp v ordinary stamp</td>
<td>4 (5238)</td>
<td>0.92 (0.78 to 1.09)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Contact</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precontact v no precontact</td>
<td>28 (28 785)</td>
<td>1.54 (1.24 to 1.92)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Follow up v no follow up</td>
<td>12 (16 740)</td>
<td>1.44 (1.22 to 1.70)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Postal follow up including questionnaire v postal follow up excluding questionnaire</td>
<td>6 (8310)</td>
<td>1.41 (1.02 to 1.94)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Mention of follow up contact v none</td>
<td>6 (6553)</td>
<td>1.04 (0.91 to 1.18)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Precontact by telephone v post</td>
<td>2 (1375)</td>
<td>0.99 (0.70 to 1.36)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More interesting v less interesting questionnaire</td>
<td>2 (2151)</td>
<td>2.44 (1.99 to 3.01)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>User friendly questionnaire v standard</td>
<td>1 (3540)</td>
<td>1.46 (1.21 to 1.75)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Factual questions only v factual and attitudinal</td>
<td>1 (1280)</td>
<td>1.34 (1.01 to 1.77)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>More relevant questions first v other items first</td>
<td>1 (5817)</td>
<td>1.23 (1.10 to 1.37)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Demographic items first v other items first</td>
<td>2 (1040)</td>
<td>1.04 (0.81 to 1.34)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>‘Don’t know’ boxes included v not included</td>
<td>1 (1360)</td>
<td>1.03 (0.82 to 1.29)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Sensitive question included v no sensitive question</td>
<td>8 (19 851)</td>
<td>0.92 (0.87 to 0.98)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Most general question first v last</td>
<td>1 (2000)</td>
<td>0.89 (0.67 to 0.96)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Origin</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University sponsorship or as source v other organisation</td>
<td>13 (20 428)</td>
<td>1.31 (1.11 to 1.54)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Sent by more senior or well known person v less senior or less well known</td>
<td>4 (2584)</td>
<td>1.13 (0.95 to 1.35)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Ethnically unidentifiable/white name v other name</td>
<td>2 (1800)</td>
<td>1.11 (0.91 to 1.36)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explanation for not participating requested v not requested</td>
<td>1 (1240)</td>
<td>1.32 (1.05 to 1.66)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Appeal stresses benefit to respondent v other</td>
<td>6 (9332)</td>
<td>0.96 (0.92 to 1.02)</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Appeal stresses benefit to sponsor v other</td>
<td>7 (9708)</td>
<td>1.01 (0.86 to 1.19)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Appeal stresses benefit to society v other</td>
<td>8 (10 088)</td>
<td>1.00 (0.84 to 1.20)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Response deadline given v no deadline</td>
<td>4 (4340)</td>
<td>1.00 (0.84 to 1.20)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Instructions given v not given</td>
<td>1 (2000)</td>
<td>0.89 (0.74 to 1.06)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Choice to opt out from study given v none</td>
<td>3 (3045)</td>
<td>0.76 (0.65 to 0.89)</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>

Effects on questionnaire response of 40 strategies where combined trials included over 1000 participants

The 40 different strategies in which the combined trials included more than 1000 participants.

Table 2 may be used to translate odds ratios into response rates from different baseline rates. At least one strategy in each category was found to influence response. For example, when incentives were used the odds of response were more than doubled when money was the incentive (odds ratio 2.02; 95% confidence interval 1.79 to 2.27) and were almost doubled when incentives were not conditional on response (1.71; 1.29 to 2.26). The length of questionnaires influenced response: short questionnaires made response more likely (1.86; 1.55 to 2.24). The use of coloured ink as opposed to blue or black ink increased response (1.39; 1.16 to 1.67) as did making questionnaires and letters more personal (1.16; 1.06 to 1.28). When recorded delivery was used the odds of response were more than doubled (2.21; 1.51 to 3.25), and they were increased when stamped return envelopes were used (1.54; 1.24 to 1.92), as did follow up contact (1.44; 1.22 to 1.70) and providing non-respondents with a second copy of the questionnaire (1.41; 1.02 to 1.94). Questionnaires designed to be of more interest to participants were more likely to be returned (2.44; 1.99 to 3.01), but questionnaires containing questions of a sen-
increase response to postal questionnaires. The pooled marketing, business, or statistical; and two fifths were educational, or health related; one quarter were psychological, medical surveys: one third were medical, epidemiological review.

than twice as many trials as any previously published in the literature on research surveys over the past 40 years. Our review, which was based on a systematic search of published and unpublished literature in English and other languages, includes more than one trial. For trials of monetary incentives the heterogeneity among the results was significant (P < 0.00001). In regression analysis, the relation was positive between the size of the incentive and the odds of response:

\[ \text{Log(Odds Ratio)} = \text{Log(OR)} = 0.69 \left[ \text{SE 0.05} \right] + 0.084 \left[ \text{SE 0.03} \right] \times \text{Log(Amount in US$)} \]

The model predicts, for example, that the odds of response with a $1 incentive will be twice that with no incentive. The model also predicts that the marginal benefit will diminish, in terms of increasing the odds of response, for each additional $1 increase in the amount given (for example, the odds of response with a $15 incentive will be only 2.5 times that with no incentive). For trials examining the effect of questionnaire length, heterogeneity between results was apparent on inspection of the forest plot (P < 0.00001). In regression analysis a relation was found between the number of pages used and the odds of response:

\[ \text{Log(Odds Ratio)} = \text{Log(OR)} = 0.36 \left[ \text{SE 0.21} \right] - 0.64 \left[ \text{SE 0.15} \right] \times \text{Log(Pages in short)} + 0.32 \left[ \text{SE 0.18} \right] \times \text{Log(Pages in long)} \]

The model predicts, for example, that the odds of response with a single page will be twice that with three pages.

Discussion

Several reviews and meta-analyses of strategies to increase response to postal questionnaires have been published in the literature on research surveys over the past 40 years. Our review, which was based on a systematic search of published and unpublished literature in English and other languages, includes more than twice as many trials as any previously published reviews. The trials identified were not restricted to intervention or control groups, and we have not examined whether the results of these trials differ systematically from the included trials. Tests for selection bias were significant in five strategies. Although these results may be due to true heterogeneity between trial results rather than bias in the selection of trials, we cannot rule out the possibility of selection bias having an effect on the results.

Identification and inclusion of all relevant trials

Identifying and including all relevant trials in systematic reviews reduces random error in meta-analyses and, because ease of identification of trials is associated with the size of treatment effects, complete ascertainment may reduce bias. We estimate that our search strategy retrieved nearly all eligible trials (estimated sensitivity 95%; 84% to 99%) and that we missed very few relevant records during screening. We excluded some trials because we could not confirm that participants had been randomly allocated to intervention or control groups, and we have not examined whether the results of these trials differ systematically from the included trials. Tests for selection bias were significant in five strategies. Although these results may be due to true heterogeneity between trial results rather than bias in the selection of trials, we cannot rule out the possibility of selection bias having an effect on the results.

Methodological quality of trials

Inadequate allocation concealment can bias the results of clinical trials. In our review, information on allocation concealment was unavailable for most of the included trials. If they had inadequate concealment, this may have biased the results, which is unlikely in this context because the researchers making the allocations would find it difficult to predict propensity to respond to a questionnaire.

Heterogeneity among trial results

We found substantial heterogeneity among the results of trials in half of the strategies, and for these it may be inappropriate to combine results to produce a single estimate of effect. Before undertaking the analyses we developed hypotheses concerning underlying differences in the trials of monetary incentives and length of questionnaire that might explain heterogeneity. Regression analyses identified relations between response and amounts of incentive and between response and questionnaire length. These models explain some of the heterogeneity. For other strategies, variation between trial interventions and populations is likely to explain some of the heterogeneity. For example, among trials evaluating non-monetary

<table>
<thead>
<tr>
<th>Baseline rate (%)</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
<th>2.25</th>
<th>2.50</th>
<th>2.75</th>
<th>3.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>16</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>11</td>
<td>16</td>
<td>20</td>
<td>24</td>
<td>27</td>
<td>30</td>
<td>33</td>
<td>36</td>
<td>38</td>
<td>41</td>
<td>43</td>
</tr>
<tr>
<td>30</td>
<td>18</td>
<td>24</td>
<td>30</td>
<td>35</td>
<td>39</td>
<td>43</td>
<td>46</td>
<td>49</td>
<td>52</td>
<td>54</td>
<td>56</td>
</tr>
<tr>
<td>40</td>
<td>25</td>
<td>33</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>54</td>
<td>57</td>
<td>60</td>
<td>63</td>
<td>65</td>
<td>67</td>
</tr>
<tr>
<td>50</td>
<td>33</td>
<td>43</td>
<td>50</td>
<td>56</td>
<td>60</td>
<td>64</td>
<td>67</td>
<td>69</td>
<td>71</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>60</td>
<td>43</td>
<td>53</td>
<td>60</td>
<td>65</td>
<td>69</td>
<td>72</td>
<td>75</td>
<td>77</td>
<td>79</td>
<td>80</td>
<td>82</td>
</tr>
<tr>
<td>70</td>
<td>48</td>
<td>58</td>
<td>65</td>
<td>70</td>
<td>74</td>
<td>76</td>
<td>79</td>
<td>81</td>
<td>82</td>
<td>84</td>
<td>85</td>
</tr>
<tr>
<td>80</td>
<td>54</td>
<td>64</td>
<td>70</td>
<td>74</td>
<td>78</td>
<td>80</td>
<td>82</td>
<td>84</td>
<td>85</td>
<td>87</td>
<td>88</td>
</tr>
<tr>
<td>90</td>
<td>60</td>
<td>69</td>
<td>75</td>
<td>79</td>
<td>82</td>
<td>84</td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
<td>90</td>
</tr>
<tr>
<td>100</td>
<td>67</td>
<td>75</td>
<td>80</td>
<td>83</td>
<td>86</td>
<td>88</td>
<td>89</td>
<td>90</td>
<td>91</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>110</td>
<td>74</td>
<td>81</td>
<td>85</td>
<td>88</td>
<td>89</td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>93</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>120</td>
<td>82</td>
<td>87</td>
<td>90</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>130</td>
<td>90</td>
<td>93</td>
<td>95</td>
<td>96</td>
<td>97</td>
<td>97</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

Table 2 Conversion of odds ratios to response rates from different baseline rates

The model predicts, for example, that the odds of response with a $1 incentive will be 2.5 times that with no incentive. The model also predicts that the marginal benefit will diminish, in terms of increasing the odds of response, for each additional $1 increase in the amount given (for example, the odds of response with a $15 incentive will be only 2.5 times that with no incentive). For trials examining the effect of questionnaire length, heterogeneity between results was apparent on inspection of the forest plot (P < 0.00001). In regression analysis a relation was found between the number of pages used and the odds of response:

\[ \text{Log(Odds Ratio)} = \text{Log(OR)} = 0.69 \left[ \text{SE 0.05} \right] + 0.084 \left[ \text{SE 0.03} \right] \times \text{Log(Amount in US$)} \]

The model predicts, for example, that the odds of response with a single page will be twice that with three pages.

Discussion

Several reviews and meta-analyses of strategies to increase response to postal questionnaires have been published in the literature on research surveys over the past 40 years. Our review, which was based on a systematic search of published and unpublished literature in English and other languages, includes more than twice as many trials as any previously published reviews. The trials identified were not restricted to intervention or control groups, and we have not examined whether the results of these trials differ systematically from the included trials. Tests for selection bias were significant in five strategies. Although these results may be due to true heterogeneity between trial results rather than bias in the selection of trials, we cannot rule out the possibility of selection bias having an effect on the results.

Identification and inclusion of all relevant trials

Identifying and including all relevant trials in systematic reviews reduces random error in meta-analyses and, because ease of identification of trials is associated with the size of treatment effects, complete ascertainment may reduce bias. We estimate that our search strategy retrieved nearly all eligible trials (estimated sensitivity 95%; 84% to 99%) and that we missed very few relevant records during screening. We excluded some trials because we could not confirm that participants had been randomly allocated to intervention or control groups, and we have not examined whether the results of these trials differ systematically from the included trials. Tests for selection bias were significant in five strategies. Although these results may be due to true heterogeneity between trial results rather than bias in the selection of trials, we cannot rule out the possibility of selection bias having an effect on the results.

Methodological quality of trials

Inadequate allocation concealment can bias the results of clinical trials. In our review, information on allocation concealment was unavailable for most of the included trials. If they had inadequate concealment, this may have biased the results, which is unlikely in this context because the researchers making the allocations would find it difficult to predict propensity to respond to a questionnaire.

Heterogeneity among trial results

We found substantial heterogeneity among the results of trials in half of the strategies, and for these it may be inappropriate to combine results to produce a single estimate of effect. Before undertaking the analyses we developed hypotheses concerning underlying differences in the trials of monetary incentives and length of questionnaire that might explain heterogeneity. Regression analyses identified relations between response and amounts of incentive and between response and questionnaire length. These models explain some of the heterogeneity. For other strategies, variation between trial interventions and populations is likely to explain some of the heterogeneity. For example, among trials evaluating non-monetary
incentives, the types of incentive used are very different, ranging from donations to charity to free key rings. Much of the heterogeneity between results may disappear when subgroups of trials are analysed. Further exploratory subgroup analyses may show important sources of variation—for example, according to methodological quality, questionnaire topic, age of the study, or type of population. In this review, our aim was to identify eligible trials systematically, critically appraise them, and present the relevant data. We did not intend to produce single effect estimates for every strategy. For many statistically heterogeneous strategies the direction of the effects is the same. For these strategies we cannot be sure about the size of the effect, but we can be reasonably confident that there was an effect on response.

Conclusions

Researchers can increase response to postal questionnaires by using the strategies shown to be effective in this systematic review. Some strategies will require additional materials or administrative time, whereas others can be implemented at little extra cost.

We have presented odds ratios for methodological reasons, but the practical implications of the odds ratio for a strategy may be difficult to interpret without knowing the response at baseline without the strategy. If the size of the effect that would be expected if a specific strategy were used is an important consideration for researchers, the data used in this review may be accessed through the Cochrane Library, where they will be updated regularly.

We thank Peter Sandercock, Iain Chalmers, and Catherine Peckham for their help and advice with the study.

Contributors: MC, CDG, PE, and IR contributed to study design, record screening, reviewing reports, data extraction, and drafting the report. PE and IR analysed the data. SF and IR contributed to data searches and data extraction. RW conducted the electronic searches. PE and IR are guarantors.

Funding: The study was supported by a grant from the BUPA Foundation and a Nuffield Trust short term fellowship.

Competing interests: None declared.

Appendix A: References to reviewed studies.


Asch DA. Use of a coded postcard to maintain anonymity in a highly sensitive mail survey: cost, response rates, and bias. Epidemiology 1996;7:530-1.


Binner PM, Barton DL. Justifying the enclosure of monetary incentives in mail survey cover letters. Psychol Marketing 1997;14:153-62.


Borm J, Sluiter JM, van der Beek AJ. Effectiveness of personal contact in randomized mail surveys: a randomized trial and meta-analysis. J Acquir Immun Defic Syndr Hum Retrovirol 1995;13:401.


Dillman DA, Sinich MA, Clark JR. Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. Public Opin Q 1993;57:289-304.


Downmeyer CJ. Experimentation on threatening appeals in the follow-up letters of a mail survey. Doctoral dissertation, University of Santa Clara, California, 1989.


Godwin K. The consequences of large monetary incentives in mail surveys of executives. Public Opin Q 1979; 43:387-98.


Greer TV, Lobina R. Effects of source and paper color on response rates in mail surveys. Ind Marketing Manage 1994; 23:37-47.


Larsson I. Increasing the rate of returns in mail surveys. A methodological study. Didactics Sociology 1970; 72:45-70.


Papers


Matteson ME. Type of transmittal letter and questionnaire colour as two variables influencing response rates in a mail survey. J Appl Psychol 1974;59:50-5.


Windsor J. What can you ask about? The effect on response to a postal screening of asking about two potentially sensitive questions. *Epidemiol Community Health* 1995;50:241-5.
Woodthorpe BR, Valcarce RW. Relative effectiveness of personalized and form covering letters in initial and follow-up mail surveys. *Psychol Rep* 1985;57:735-44.
(Accepted 3 December 2001)