Author's response to reviews

Title: Response rates to a mailed survey of representative cancer patients: incentive and length effects.

Authors:

Bridget J. Kelly (bkelly@rti.org)
Taressa K. Fraze (tfraze.fraze@gmail.com)
Robert C. Hornik (rhornik@asc.upenn.edu)

Version: 2 Date: 31 December 2009

Author's response to reviews: see over
Dear Dr. Norton,

Please consider the revised version of the following manuscript, entitled: *Response rates to a mailed survey of a representative sample of cancer patients: incentive and length effects.*

The paper has undergone major revisions to address the concerns of the four reviewers. Some data tables have been deleted and others added in order to satisfy reviewer comments. Specific details about how each reviewer comment was addressed are included in the pages that follow.

All authors of this research paper have directly participated in the planning, execution, or analysis of the study.

The content of this manuscript has not been copyrighted or published previously and is not now under consideration for publication elsewhere. The contents will not be copyrighted, submitted, or published elsewhere while acceptance by this journal is under consideration.

We appreciate your consideration of this revised manuscript.

Thank you,

Bridget

Bridget Kelly, PhD, MPH
RTI Research Analyst
RTI International
701 13th Street, NW, Ste. 750
Washington, DC 20005-3967
phone: (202) 728-2098
fax: (202) 974-7855
Reviewer's report
Reviewer 1
Reviewer: Wendy Demark-Wahnefried

Major concerns:
Data on cancer stage should be available for cases in the cancer registry, and these data should have been made available, described within the paper (this manuscript appears to be lacking the traditional Table 1, where the sample is described in terms of cancer-type and stage, age, race, and gender (at least for CRC cases), and finally used in the analysis. The authors speculate that poor response rates in CRC patients may be due to higher proportions of individuals with distant disease….indeed, they should base this statement on data from this study, and furthermore, it really is unfortunate that thoughtful sampling was not pursued apriori, i.e., a careful plan to stratify the sample on important variables, such as stage, age, gender and race so that the data that emanated from this study could be stronger. While this is not possible at this point in time, gathering the stage data from the cancer registry and conducting appropriate analyses is indeed possible and should be done.

Response: The stage data is available and is included in table 4. We have also added a descriptive table for respondents (see Table 1), which includes stage data. To the point about careful sampling, this study was developed as a pilot effort to determine how best to sample the population. The results of this study were used to create a sampling plan for the larger study (n=2072). The results of this study (which demonstrated that a smaller incentive and a long questionnaire were viable) also established that stage 4 and African-American respondents responded at a lower rate than others. As the result the full study oversampled by African-American race and cancer stage 4. Language to explain this has been added to the methods section.

Likewise, many cancer registries record disease status and are able to categorize whether patients are alive with no evidence of disease (NED), are living with progressive or recurrent cancer, are dead, and also whether they have only been diagnosed with this cancer, or have multiple primaries. No mention was made in the methods to make sure the sample was limited to those who were alive, NED and devoid of other cancers – was it? If the state registry does not uniformly track these data, was any attempt made to run the cases against the state mortality tapes? (this is much more important to do upfront than to present any sort of data on estimated mortality rates and then speculate that the reason for non-response is that the survey was mailed to deecedents).

Response: This comment appears to raise two distinct issues – whether our sampling plan might have been responsive to other information in the registry, and whether our estimates of mortality to correct the response rate estimates might have been based on case-specific mortality information rather than global estimates. As to who was sampled: we were anxious to get to the field with the mailed questionnaire as soon as possible after diagnosis to maximize accuracy of memory and so made use of the preliminary information from the registry as soon as it was available to define the sample. Information that was eventually available from the registry or other sources was not available to us at the time we were developing the sample for this study. In addition
wanted a sample that fully represented the diagnosed population so did not consider restricting it by other information. As to the method we chose to correct our response rates for likely mortality, we were not confident about the completeness of the registry mortality information and we were not free (given restrictions on our use of the identifying information provided with the registry names) to use external sources of information. In addition we considered the SEER and age-specific mortality data sets used to estimate likely mortality unbiased sources for this analysis.

Last, but not least, it is concerning that DCIS cases were included in this sample and that a letter actually went out stating that the person was contacted because they had cancer….I am hopeful that the letter included verbiage regarding in situ disease (please confirm since this is an IRB issue)? For the patients who denied that they had cancer, it is abundantly important to present the stage data.

Response: Our recruitment letter indicated that the PCR was the source of their name and indicated that it was possible that they did not have cancer. This letter was approved by the Institutional Review Board at the University of Pennsylvania. The exact language is as follows, “Your name was selected from a list of names provided to us by the PCR. We have enclosed a brochure which describes the registry in detail. We are only interested in people who have been diagnosed with cancer. If you have not been diagnosed with cancer, you are not eligible to participate in this study. If you have any questions about the cancer registry’s policies, please contact the PCR directly at 1-800-272-1850.”

2. The AAPOR formulae presented for response rate seem much more relevant for telephone surveys than mailed surveys.

Response: These are standard formulae that are applied to mailed surveys according to AAPOR. We used the definitions for “AAPOR’s Standard Definitions: Mail Surveys of Specifically Named Persons,” which includes specific instructions for how to handle “returned questionnaires,” “eligible, no returned questionnaire,” and other circumstances specific to mailed surveys.

3. Independent analyses are used to compare breast cancer cases to prostate cancer cases and then CRC cases to prostate cancer cases instead of conducting these analyses simultaneously.

Response: We agree with the reviewer that it is more appropriate conduct these analyses simultaneously. We have re-run the regression with colon cancer as a reference category and breast and prostate cancers entered as dummy variables. Table 2 has been updated and appropriate text added to the methods section.

4. The dollar amounts that have been used in previous studies that were presented in the introduction should be translated to today’s dollar value so that the reader can put this information in perspective to the amounts used in the current survey.

Response: Amounts have been adjusted for inflation using a calculator from the U.S. Bureau of Labor Statistics. (See page X).

5. One could speculate that there is an effect by gender regarding survey length
and incentive amount. Granted, gender is confounded by cancer type with regard
to breast and prostate, but there probably isn’t enough power to explore this just
within CRC cases.

**Response:** We did look at unadjusted response rates among men and women with colon
cancer. For men, the response rate was 47.1%, compared to 48.9% for women. This was
not a significant difference in the sample of 400 colon cancer patients.

6. This manuscript requires some serious editing as far as grammar,
terminology-use, etc. Perhaps, most glaring is the use of “mail surveys” instead
of “mailed surveys.” Even more disturbing is the depersonalized way of referring
to the participants on this study. First of all, it is hoped that this survey did in fact
yield data that were useful in terms of a greater overarching project (and if so, it
would be helpful to mention this). Secondly, the language that is used should
indicate that these surveys were completed by individuals.

**Response:** AAPOR calls them “mail surveys” and they have been referred to that way in
many papers in the literature (Childers& Ferrel, 1979; Dillman; Sinclair & Clark, 1993;
Fox, Crask & Kim, 1998; Smith, Stein Mehta. Et al. 2006). However, since the reviewer
prefers the term “mailed” we have changed the reference throughout the paper.
Throughout the paper, we refer to the respondents as “cancer patients”. However, we did
use the passive voice, so we have made an effort to change the voice to a more active one
throughout (e.g., “cancer patients returned…” instead of “the surveys were returned”).
We have also added language to the discussion to explain how the data from this survey
was used in the overarching project (see page 13).

Minor concerns:
Title: The term “representative,” is used…representative of what? Few pilot studies
involve 1200 participants – the terminology here needs to be
reconsidered.

**Response:** These cancer patients are representative of patients with these three cancers in
the state of Pennsylvania. We have no alternate wording for “pilot study”, as “pre-test”
does not seem better. While the sample size is large, this was indeed a pilot of a much
larger survey, (with a final n=2072), with the purpose of determining whether the
methods were practical and testing what questionnaire length and incentive structure
would be most sensible, and determining the need to oversample particular
subpopulations.

**Reviewer 2:**
**Reviewer's report**
**Title:** Response rates to a mailed survey of representative cancer patients:
incentive and length effects.
**Reviewer:** Elaine McColl

**Major Compulsory Revisions**
1. Was a power calculation carried out, or was the choice of 400 patients in each
of the samples essentially arbitrary? If a power calculation was carried out, please report
it. If not, please justify why not. In factorial designs, the power for the interaction terms is
of course less than for the main effects analyses.
Response: The purpose of this pilot study was to determine whether a mail ed survey was feasible for this population (no other study that we knew about had surveyed this population and achieved a high response rate) overall and secondarily whether a long questionnaire and a small incentive would stand in the way of completion. With a sample of 1200 divided into four conditions (or for any cancer divided into two conditions), we knew we could detect differences of more than 5% (alpha=.05, power=80%; expected difference 60 vs. 64%) From the perspective of a decision to go forward that was sufficient for our practical purpose; if the observed differences were less than 5% between conditions we would go ahead with the smaller incentive and longer questionnaire (additional comment added on page 6.)

Please justify the use of prostate cancer patients as the reference group in your dummy coding of type of cancer. Why not colon cancer, since this included both men and women?
Response: We agree with the reviewer that it is more appropriate to use colon cancer as the reference category. The regression has been re-run and table 2 updated with colon cancer as the reference group.

**Minor Essential Revisions**

Page 5. Presumably just as the breast cancer sample comprised only women, the prostate cancer sample comprised only men. This should be made explicit, along with the gender split in the colon cancer sample.
Response: Language to this effect has been added (see page 5).

Page 5. To avoid confusion with a cross-over trial design, I would recommend substituting 'two by two fully crossed experiment' with two by two factorial experiment'.
Response: This change has been made (see page 5).

Page 6. Reference to 'interviews' in the presentation of calculation of response rates is confusing, given that this was a postal survey. Please use a different term ('questionnaires' would be fine).
Response: This change has been made. See page 6.

**Discretionary Revisions**

Page 3: The cited literature on previous research into the impact of various factors on response rates to postal surveys is predominantly from the US. I recommend inclusion also of citations of reviews carried out in the UK (but of international literature). Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, Kwan I. Increasing response rates to postal questionnaires: systematic review. British Medical Journal 2002;324;1183-1192. http://bmj.bmjjournals.com/cgi/reprint/324/7347/1183

Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, Kwan I and Cooper R. Methods to influence response to postal questionnaires (Cochrane


All of these provide supporting evidence for the issues discussed in this paper. More specifically, reference could also be made to Jacoby A. Possible factors affecting response to postal questionnaires: findings from a study of general practitioner services. Journal of Public Health Medicine. 12(2):131-5, 1990. This paper compared longer (50 questions over 16 pages) and shorter (30 questions over 8 pages) questionnaires in a health survey context and like the paper under review found no effect of length on response rates.

**Response:** This is a good point. We have added these citations to the literature review section (Beginning on page 4).

Page 13. I agree with your recommendation that a more systematic study would examine the effects of different dollar (or whatever currency is appropriate to the country of research) amounts of incentive. I would also suggest adding that the focus should not be solely on the effectiveness in terms of response rates, but also on cost-effectiveness, and in particular on the marginal cost per extra questionnaire returned (rather than the average cost). Few of the studies identified by Edwards et al and McColl et al in the reviews above considered cost issues.

**Response:** Language as been added to the discussion section (see page 14).

**Reviewer 3**
**Reviewer:** Timothy J Beebe
**Reviewer comment:**
First, the introduction and discussion sections ought to cover the relative merits and pitfalls of the various data collection modes in greater depth. As it stands, the advocacy for the mailed form appears one-sided as there is ample evidence that of the three main modes (mail, telephone, face-to-face), mailed surveys garner the lowest levels of participation. Also, much of the literature cited deals with populations and/or content unrelated to the study focus. For example, the coverage of the questionnaire length literature cites studies focused on adolescent populations, physician surveys or web-only modes. Few focus on general populations, patients (let alone cancer patients), or health surveys. More digging seems warranted.

**Response:** We have added more detail on relative merits and weaknesses of each mode. We have also added references that focus on general populations, health surveys or
patients and deleted some that were less relevant. See new text, beginning in the introduction.

Second, there are variables that appear in methods section or results that receive no coverage in the introduction. For example, is there no literature covering the potential impact of incentive amount or questionnaire length on willingness to participate in follow-up? How about for cancer stage? If information is provided in the methods or results sections, it ought to receive some attention in the introduction, otherwise it appears out of the blue.

**Response:** The section on willingness to participate in the follow-up has been omitted in accordance with recommendation of other reviewers.

On a related note, the most interesting findings are the generally null results relating to the manipulated variables of incentive amount and questionnaire length. All of the other information seems extraneous, especially given that most of the results are statistically non-significant. The paper and the tables in particular would be more straightforward if the focus was only on the manipulations.

**Response:** We have omitted the other variables to ensure that the focus is on the incentive and length manipulations.

Third, is there any way to pull some information off the cancer registry sampling frame that would support the assessment of nonresponse bias? As the authors indicate, nascent evidence suggests a very weak relationship between response rates and nonresponse bias. Most registries have quite a bit of information on its members that could support the comparison of respondents and nonrespondents. Also, did patient self-reported information vary by the conditions? Addition of such information to the manuscript would heighten its relevance as few studies attend to that information, especially in health surveys. Without it, the manuscript strikes one as warranting publication as a short note rather than a full research article.

**Response:** 1) we do provide some analyses of substantive results by condition and did not find any significant differences. These are reported on page 12.

2) There are two types of non-response studies that might be undertaken: The important non-response bias papers (e.g., Keeter et al.) have all focused not on differences in characteristics of those who did and did not respond, but on differences on variables measured in the questionnaires between those who responded to sampling procedures with low response rates and those procedures with high response rates. Obviously this type of non-response study could not be undertaken with the current data set. The closest parallel we do offer is the analysis of outcome responses by experimental groups within the sample – as noted above, given the lack of response rate differences between experimental groups it is no surprise that there were also no differences in questionnaire responses by experimental conditions. They did not produce any response bias, per se.

We have compared respondents and non-respondents on characteristics found in the PCR (see Table 5). We are able to examine whether response rates differed by variables
available in the PCR and showed that there were some differences (by race and stage and marital status, particularly). However, this does not provide estimates of non-response bias on the outcomes of interest (communication behaviors of patients.)

We were concerned about the effects of non-response on our results. Thus we did make use of information relevant to this concern for the full sample in later analyses. We have developed weights for our analyses of the full sample (n=2072) to adjust our sample for non-response bias vis a vis the full PCR population on six characteristics that were available in the PCR data base (age, gender, stage, date of diagnosis, race-ethnicity, marital status, each within the three cancers.) Comparing results with these weights applied versus not applied (thus correcting for non-response bias) has affected our substantive analyses minimally, across a number of papers now in press or published.

**Some specific comments**

**Introduction:** More references are needed to support some of the assertions in the first paragraph (e.g., “…average response rates [for web surveys] are not high compared to those for phone surveys.”). Also, some mention of how measurement error related to data collection mode should be offered (see Link et al. regarding web surveys vs. mail and telephone).

**Response:** See new text, page 1.

Finally, casting the hypotheses in table form is atypical and distracting. These should be woven into the text. Also, they appear one-tailed but the statistical tests appear two-tailed.

**Response:** These hypotheses have been woven into the text and the table deleted. They are rephrased as research questions so that they are two-tailed, in accordance with the statistical tests (see introduction and background section).

**Methods:** In the description of the mailing procedure, it isn’t clear what “…return of the mail survey served as evidence of consent…” given that they had not received a questionnaire in the first mailing (unless I’m missing something).

**Response:** This paragraph was unclear. We thank the reviewer for pointing that out. The language has been changed to clarify that if the questionnaire sent as part of the second mailing was returned, it served as evidence of consent.

Also, a lot of real estate in the paper is devoted to describing the differences between AAPOR RR2 and RR4. This text should be cut down. The authors might also consider testing to see if the results varied by use of the two rates. If things don’t vary, then results (including tables) and discussion should focus on only one for parsimony and clarity.

**Response:** The two rates permit us to do different forms of analysis. The RR4 rate is the best estimate for the overall response rate and is the rate that captures how well the mailing procedures functioned. Ideally that rate would have been used for all analyses. However it could not be confidently calculated within experimental conditions so we used the RR2 rate for those analyses. While the rates are not sharply different (RR2=62% versus RR4=67%) if we had to use only one rate we would have been required to choose
the lower rate since it could be used for the experimental analyses, and we are reluctant to see future citations which should focus on the overall response rate, ending up referring to the 62% rate rather than to the 67% rate, which is the appropriate estimate. This is particularly notable for colorectal cancer patients where the RR2 rate was 47% and the RR4 rate 57%, due to the relatively high expected mortality rate from colorectal cancer. To respond to the concern about space taken up by this discussion it has been cut back sharply, formulas have been eliminated and only substantive comments relating to the specifics of this study have been retained.

Results: The footnoted point at the bottom of p. 10 should be mentioned in the methods section and woven into the results. Formally including item nonresponse as a measure of data quality would increase the potential impact of the paper (see earlier point in general comments). For the tables, make sure that the reference category for each of the variables is clearly defined and noted for easy interpretation.

**Response:** We have described the item non response analysis in the methods and added it to the results section. We have clarified the reference category in the tables.

Discussion: Reference 40 addressing the issue of average response rates for mailed surveys (55.6%) seems less germane than one done by Asch et al. focusing specifically on health surveys (see Asch, D.A., M.K. Jedrziewski, and N.A. Christakis. 1997. “Response rates to mail surveys published in medical journals.” Journal of Clinical Epidemiology 50: 1129-1136).

**Response:** We thank the reviewer for bringing this study to our attention and have referenced it. However, the point the authors were making was that in the last decade at the time the study was published, mail surveys have not achieved response rates that high. This study was published in 1997.

Also, did the ACS study mentioned in that same paragraph really look at a mixed mode mail and telephone survey (the one where they got a response rate of 34%)? If so, that point should be clearer as it really doesn’t represent an apples-to-apples comparison with the authors’ mail-only investigation.

**Response:** The reference to the ACS study has been edited to clarify that it was a mixed mode study.

Finally, the cancer-related findings are interesting. But, one wonders whether the observed differences in response across breast, prostate, and colon cancers hold after formally controlling for demographic (non-gender-related) differences in case-mix across those groups.

**Response:** Education was not available as a variable, However, we did control for age, race and marital status and found that the differences by cancer do still hold. The same is true after adding stage of cancer. We have added a note regarding this to the table with regression results and the methods section.
Reviewer 4
Reviewer: Phil Edwards

MAJOR COMPULSORY REVISION

Page 4: The authors provide a good summary of some of the literature on the effects of incentives and questionnaire length on response, but miss two published meta-analyses of these effects [Reviewer included references]

Response: These references have been added to the literature review section (beginning on page 1).

Page 5: The method of randomisation of participants should be described: how was the randomisation sequence generated, who conducted the randomisation, any measures used to conceal the allocation sequence, etc.

Response: Randomization was conducted using a random number generator in SPSS. Only the research director had access to the master list. Each subject was assigned a code corresponding to one of four conditions (e.g., short survey, $3/long survey, $5, etc. These codes were placed on the inside back cover of the survey for identification upon their return). This language has been added to the manuscript (see page 7).

It would be more informative if authors could provide an estimate of the main effects of the incentive and length on the percentage responding (with 95% confidence intervals). Table 3 includes odds ratios and standard errors (for logistic regression coefficients) which are good summary statistics that estimate the intervention effects, but I cannot find any reference to these results in the manuscript. Throughout the results section it would be useful to report the odds ratio for response with 95% CIs, or to include the p-values. Also, please give p-values rather than “p=NS”.

Response: Main effects are described on page 9 of the results section and also shown in table 2 (in the total columns). Odds ratios and p-values have been added.

MINOR ESSENTIAL REVISIONS

Page 5: The description of the materials used is very helpful, but more details about the incentive would be useful (i.e. was the $5 incentive made up of five one dollar bills or one five dollar bill?). Also, what time period was specified at which response rates were calculated? Who packaged and sent the questionnaires?

Response: A $5 bill was used for the incentive. Response rates were calculated 4 months after the initial mailing was sent. Questionnaires were packaged and sent by members of the grant team and other graduate research assistants.

Language to this effect has been added to the methods section (see page 6)

Discussion of the results could include statistical power. For example, what effect sizes was the study powered to detect? Also, rather than reporting that there was a “lack of interactions” it would be more accurate to say “lack of evidence for interactions” as the study was not powered to detect these.

Response: (see language above and in the text as to the differences that the study was designed to find.) The language on page 11 has been changed to reflect a “lack of evidence”.


DISCRETIONARY REVISIONS

Pages 6–8: whilst it is useful for researchers to know about the differences between RR2 and RR4, there is rather a lot of text devoted to the explanation which rather detracts from the main messages of the paper. The only difference between the two formulae is the proportion of cases of unknown eligibility that are eligible. Either this material can be appended, or else simply provide a reference to the AAPOR source.

Response: We have kept both rates as noted above but sharply cut back the real estate devoted to their explication.

References:
AAPOR Standard definitions.
http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions&Template=/CM/ContentDisplay.cfm&ContentID=1273