Author's response to reviews

Title: The association between survey timing and patient-reported experiences with hospitals: results of a national postal survey

Authors:
  Oyvind A Bjertnaes (oan@kunnskapssenteret.no)

Version: 2 Date: 1 July 2011

Author's response to reviews: see over
Reviewer: 1

Discretionary Revisions:

1. General comments: The article addresses an important topic, as surveys remain an important tool in patient satisfaction research, but also face significant methodological challenges. The questions posed by the author are suitable and the methods utilized in the analyses are both appropriate and sufficiently defined. The review of the literature is up-to-date, but may require some expansion. Finally, the conclusions drawn from the analyses are appropriate, with limitations stated. Abstract: I would recommend using alternative terminology regarding existing research on the effect of survey timing, as “insecure” fails to fully describe the issues involved. The author may choose to simply indicate that the existing research is “contradictory” and then be more specific in terms of where studies differ. Terminology should be changed throughout the manuscript for consistency.

The term “contradictory” is now used throughout the manuscript.

2. Background: Second sentence – it is recommended that the author expand/be more specific in terms of descriptions of the methodological challenges faced in patient satisfaction research.

This is now described more specifically in the Background (first paragraph):

“Surveys are frequently used to measure patient experiences and satisfaction with health care (3, 4), but their value is subject to several methodological challenges. A review of the literature showed that many reports of studies on patient satisfaction present little documentation of the reliability and validity of the instruments used (5), raising major concern about the quality of these studies. Another methodological challenge relates to the relatively high rates of non-response, especially for postal surveys (6), which can impair their generalizability. A third methodological challenge relates to the use of patient-reported experiences or patient satisfaction in benchmarking between different units, where fair comparisons require adjustment for differences in the patients’ characteristics (7, 8).”

3. Throughout the manuscript the author alternatively discusses “scales for patient experiences” and “patient satisfaction.” I would recommend that the author use the latter terminology throughout, as my understanding is that the scales were designed to measure satisfaction with care.

Unfortunately, there is no consensus in the literature about these terms. The term “patient reported experiences” is now primarily used to describe our study, while “patient evaluation” is used as a broader term covering both patient reported experiences and patient satisfaction. Sometimes both patient reported experiences and patient satisfaction are used because references are made simultaneously to different studies using divergent terms.

Minor Essential Revisions
4. **Background:** Revision is needed in terms of the conclusions/logic of the first paragraph. The author states that results of previous studies are contradictory, but then states that timing of surveys might have substantial effects on the estimated level of patient satisfaction. In the face of contradictory evidence, it is recommended that the author present evidence from other fields/subject areas where timing is important to support this contention. This would also expand the literature review beyond the single review paper cited.

This has been dealt with by formulating a new paragraph in the Introduction (second paragraph):

“Many studies have focused on these methodological issues, but few rigorous studies have been conducted to assess the effect of survey timing on patient evaluation. This is problematic because both patient recall and clinical and health-related quality-of-life outcomes might produce different patient evaluations depending on when the survey is applied to patients (3). Following a longitudinal study, two different theoretical models for patient satisfaction were suggested: (i) an immediate post-visit satisfaction model that includes demographics, patient expectations, patient functioning and patient–doctor interaction; and (ii) a model for 2-week/3-month satisfaction that includes demographics, expectations, patient functioning and symptom improvement (9). This suggests that the results of evaluations of patients would vary with the survey time point.”

5. **Background (second paragraph):** It is recommended that the author incorporate support/citations for the possible interpretations regarding survey timing (see note above).

See point 4 and 11.

6. **Statistical Analyses:** Recommend that the author provide justification for the recode of survey time into the categories chosen.

This is now briefly described in the Methods/Statistical analysis (included in the second paragraph):

“Survey time is a continuous variable and was included as such in the regression analysis described below. In bivariate analysis, the survey-time variable was grouped by week, which gave large groups for statistical comparisons. These bivariate analyses are secondary and no other tests were conducted to assess the appropriateness of this grouping.”

7. **Statistical Analyses:** Incorporate last sentence concerning SPSS use into the last paragraph in this section.

Done.

8. **Results:** Not sure what “lower patient experience scores” means. Recommend that the author operationalize in terms of “patient satisfaction.”

See point 3 above.
Discussion: Second sentence is not clear as written. Would recommend that this be restated in terms of time intervals between discharge and survey.

This has been clarified as follows in the Discussion (first paragraph):

“Half of the studies identified in a review found that patient evaluations were poorer when measured longer after the health-care visit (3). This is in line with the results of several other studies not documented in the review (15-16, 18, 21–23).”

Discussion: Clarify discussion of mode and timing (second paragraph). Additionally, the conclusion drawn from these two studies may be stated in a more cautionary fashion, as it is still difficult to be definitive based on only two empirical works.

This has been changed to clarify mode and timing effects, in addition to changes following the revised regression analysis (see point 12 for information about revised regression analysis). Revised paragraphs 2-4 of the Discussion are as follows (the fourth paragraph point to a need for research on possible reasons for the decline over time, which is a main finding following the new multilevel analysis):

“Surveys of patient-reported experiences and patient satisfaction might be conducted in many different ways. The choice of survey mode is one important decision where a range of different possibilities exist, such as between postal or electronic questionnaires, telephone interviewing or face-to-face interviews. The choice of survey mode is important for several reasons, but particularly in the context of the present study since it can affect patient evaluation. For example, telephone interviews have been shown to generate more extremely positive answers than other data collection modes, possibly because of the presence of an interviewer (10). This mode effect was also identified in a patient-experience study comparing telephone interviews and the mail mode (11). Other mode effects have also been identified in research into patient experience and patient satisfaction. Survey modes including on-site data collection and interview methods have resulted in lower expressed dissatisfaction (3, 12), probably due to the degree of anonymity and less pressure to give socially acceptable answers (3).

While mode effects are related to how the data are collected, survey-time effects are related to the time it takes from the health-care encounter to the patient receives the questionnaire. If the distribution time affects patient evaluation, a survey-time effect has occurred. The presence of mode effects in this kind of research means that studies assessing the effect of survey time need to standardize the data-collection mode in order to avoid confounding time and mode effects. Almost all studies showing a worsening in patient evaluation over time changed the data-collection mode between the different measurements (13-16, 18, 21–23). Therefore, the timing effects might be related to the mode change rather than being actual timing effects. The current study standardized the data-collection mode and found a significant association between survey time and patient-reported experiences for three of the six scales. This is in line with the aforementioned studies, but contradicts another study from Switzerland in which the data-collection mode was also standardized (19). However, the Swiss study only included one hospital, a specific patient group and a relatively small sample. Consideration of all of the available data suggests that there is a negative association between survey timing and patient-reported experiences and satisfaction. The lack of studies with a rigorous design point to a
need for more high-quality research on survey-time effects, including studies with other population groups, other data-collection modes and a longer time span.

“Research on possible reasons for the decline over time is also needed. It is well known that patient evaluations tend to be highly skewed against positive evaluations (28). Several studies have indicated that psychosocial factors prompt positive evaluations (3, 29-30), such as gratitude, loyalty, luck and equity (30). Consequently, one possible explanation for the decline over time is a weaker effect of these psychosocial factors, meaning that patients on average feel more comfortable being critical as time passes. However, other explanations are possible, such as patient recall or changes in clinical and health-related quality-of-life outcomes (3). A fruitful research approach to assess these causes could combine a longitudinal approach and qualitative data.”

Major Compulsory Revisions:
11. Background: Clarify distinction between survey mode and timing of the survey. As currently written, there may be some confusion.

A new revised third paragraph has been included in the Introduction, following reviewers’ comments in point 11 and 5:

“Effects related to survey mode and survey time are easily confounded in methodological patient experience and patient satisfaction research. Survey mode effects are related to how the data are collected. Examples of survey modes are electronic questionnaires, telephone interviewing and face-to-face interviews. Telephone interviews have been shown to generate more extremely positive answers than other data collection modes (10), also in patient experience and patient satisfaction research (3, 11), while on-site data collection results in a more favourable patient evaluation (3, 12). These effects should be distinguished from survey-time effects, which are related to the time it takes from the health-care encounter to the patient receives the questionnaire. Questionnaires might be distributed immediately after a healthcare encounter, a short time after the encounter or a long time after the encounter. If the distribution time affects patient evaluation, a survey-time effect has occurred. Several studies have investigated the association between survey timing and patient evaluation (13–23), and most of them found that patient evaluation is poorer when measured at a longer time after the encounter (13-16, 18, 21–23). However, a closer investigation of these studies shows that in all except one (18), the data collection mode changed between the different measurements; the aforementioned timing effects might therefore have been due to changes in data collection mode. In fact, the best-designed study concerning the association between survey timing and patient satisfaction found little association between survey timing and patient satisfaction measures (19). That study grouped patients into three different mailing intervals: 1, 5 and 9 weeks after discharge. All groups received and completed a postal survey at home, and so the data-collection mode was standardized between the groups.”

12. Limitations: Author should discuss potential skewness based on distribution of the survey time variable. May also be important to discuss in previous sections concerning methodology and analyses.

Survey time should have been standardized across hospitals following the survey protocol, with individual variation ranging from 1-15 days. A further inspection of the distribution of the survey time variable across hospitals showed that survey time vary significantly between hospitals, with hospital means ranging from 8.6 to 20 days. This means that some hospitals
were more compliant than others regarding the survey protocol, with skewness on the survey time variable across hospitals as a result. Furthermore, additional analysis showed that patient reported experiences also vary significantly between hospitals for all scales, indicating a need to adjust for the hospital level in regression analysis. To account for potential confounding effects related to the hospital level, it was decided to change from ordinary regression to multilevel regression in table 3. Multilevel regression divides the variance between different levels, and gives more accurate significance tests at the patient level in situations where the cluster effect is significant. This change in statistical analysis resulted in a number of changes in the manuscript, which is described in detail below:

Revised Methods/Statistical analysis:

“Multilevel linear regression analysis was used to assess the association between survey timing and the six patient experience scales, controlling for gender, age, self-perceived health, education, admission type, number of admissions in the previous 2 years, response time and hospital. Patient clustering within hospitals might inflate t values in ordinary linear regression models so as to produce a type I error, which was the reason for using multilevel regression. The multilevel model divides the total variance in patient-reported experiences into variance at the hospital (macro) level versus the patient (micro) level. The hospitals were included as random intercepts, and all variables from the ordinary regression as fixed effects at the patient level. Standardized variables at level 1 were used in the regression; consequently, standardized regression coefficients were computed.”

Revised Results (regression):

“Multilevel regression analysis revealed that several background variables were significantly associated with patient-reported experiences (Table 3). Survey time was significantly related to three of the six patient-reported experience scales: doctor services (Beta=–0.424, p<0.05), information about examinations (Beta=–0.566, p<0.05) and organization (Beta=–0.528, p<0.05). All associations were negative, indicating that the patient-reported experience scores declined with increasing survey time. The most important predictor of all patient-reported experience scales was patients’ self-perceived health, with Beta values ranging from –2.66 (for hospital and equipment) to –4.70 (for information about examinations). Other significant predictors were response time (five scales), gender (two scales), age (six scales), education (five scales), admission type (six scales) and number of admissions in the previous 2 years (five scales).”

A new paragraph was also inserted in the Discussion about possible limitations related to skewness based on distribution of the survey time variable:

“A third possible limitation of the study concerns its observational research design, causing uncertainty related to potential confounding variables. The gold standard for effect research is randomized trials, in which the aim is for only random variations to exist between study groups and for there to be a direct link between intervention and effect. However, a multicentre randomized trial on this topic would present large practical and methodological challenges, both regarding which time frames to use (intervention) and how to apply sample frames and randomization across hospitals. Another suitable design could have been a longitudinal approach, but that was not possible in this study. The present study adjusted for the most important sociodemographic predictors of patient experiences and patient satisfaction, reducing the probability of confounding effects related to variables not included. The study was based on data from all hospitals in Norway, and the survey-time variable was
registered and analyzed as a continuous variable at the individual level. The former feature increased the external relevance of the study, and the latter gave the opportunity to use survey time in days in the analysis, providing more detailed information than only groups based on, say, weeks or months.”

The first paragraph of the Discussion was changed because of different results following multilevel regression:

“Research on the effect of survey timing on patients’ evaluations of health services has produced contradictory results (3). Half of the studies identified in a review found that patient evaluations were poorer when measured longer after the health-care visit (3). This is in line with the results of several other studies not documented in the review (15-16, 18, 21–23). However, the most rigorously designed study on this topic found little association between survey timing and patient satisfaction (19). Our main finding supports the former studies, showing that survey time was significantly and negatively related to three of six scales for patient-reported experiences with hospitals.”

The Conclusion was changed because of different results following multilevel regression:

”Survey time was significantly and negatively related to three of the six scales for patient-reported experiences with hospitals. The effect was small compared to the most important sociodemographic variables, self-perceived health and age, indicating that protocols for local and national inpatient experience surveys with hospitals have some flexibility in relation to the time between discharge and posting the survey. However, large-scale hospital comparisons of patient-reported experiences should consider survey time as an adjustment factor if it is not standardized across hospitals. The generalizability of the survey-time effect to other topics and other modes is uncertain, but a negative association has been found in most of the other studies referenced, including patient populations in primary care and hospital in- and outpatient care. However, more high-quality research on this topic is needed, including studies with other population groups, other data collection modes and a longer time span.”

Reviewer: 2

Minor essential revisions:

13. This is a well-written paper about a well-designed experiment. The paper is well situated in the literature and provides a valuable contribution to those interested in measuring how patients evaluate health care. Specific comments follow. Define what you mean by “survey timing” earlier in the abstract and manuscript.

Done. See also point 11, reviewer 1.

14. Methods: Does timing impact any other outcomes such as response rate, time to response or respondent representativeness? These are all important considerations that could be addressed with your data. In the last paragraph of the discussion you suggest that response rate could differ by survey time – you could empirically test this. In this same place you state that non-response bias is not a concern, however, there could be different levels of nonresponse bias by survey timing.
Reviewer is right, these are all relevant outcomes. However, the data set is from 2006 and data about non-respondents has been made anonymous due to juridical and data security issues. Furthermore, the main focus in the study is the association between survey timing and patient-reported experiences. Two studies that included response rate as an outcome measure are referenced in the Discussion.

15. Methods, data collection: Provide more clarity with respect to the division into 14-day periods. This part of the methods section is unclear. It seems like survey time should be randomly distributed across patient characteristics, but this was not found in your results. Is there something systematically different between those with different survey times?

Yes, a systematic difference was found between hospitals regarding survey time. This has resulted in a change from ordinary regression to multilevel regression analysis, and a number of changes in the manuscript (see response to point 12 above).

The division into 14-day periods has been clarified as follows in the Methods/data collection (included in the first paragraph):

“The inclusion period from September 1 to November 23 was divided into six 14-day periods. The purpose of this division was to secure relatively recent experiences at the time of patient evaluation, which was based on previous practice and not scientific evaluation. According to the protocol, each hospital transferred the contact information about discharged patients to the Knowledge Centre immediately after each 14-day period. The Knowledge Centre mailed questionnaires to patients following each transfer.”

16. Methods, statistical analysis: How were the 4 groups for the survey time variable chosen. If this was done by some manner other than being theory driven, were any sensitivity tests conducted?

See response to point 6 above (reviewer 1) regarding survey time groups.

17. Results: Include response rate here.

Done.

Discretionary revisions:


This has been changed to “contradictory” (see response to point 1 above, reviewer 1).

19. Abstract, results: include the response rate.

Done.

20. Methods: Discussion: Would be useful to discuss the potential generalizability of these results to other topics, other populations and other modes.
See response to point 12 above, where the revised conclusion includes reflection on the generalizability issue.

21. Minor issues not for publication. Discussion: You state that the response rate is “average” do you mean “in line with”?

This has been changed to “in line with”.

Reviewer 3

Major compulsory revisions:

22. This study finds that delays in sending the survey is associated with worse reported experiences, but that this seems to be explained away by controlling for patient health status at time of interview. The authors conclude that “response time” (how defined in this article) is “unrelated” to patient experiences and that based on these analysis, “Protocols for local and national inpatient experience surveys with hospitals have flexibility in relation to time from discharge to posting of survey, at least up to five weeks. My key concerns with the article relates to the interpretation of prior literature and the conclusions/implications drawn from this study. These issues can be addressed. My major concerns: Can you the authors please clarify the key predictor variable: “Response time was computed as the difference between date for response registration and date for first postal mailing.” Does response registration refer to compiling list of all discharges in the past 14 days (in which case response time is really “first mailing time”) or is it days between discharge and survey completion? The answer has important implications for my comments. Elsewhere, the authors describe “survey timing” which implies that they are measuring days until survey is initially sent, which is entirely under the control of the survey firm/hospital.

Response time and survey time are two separate variables, this should have been presented more clearly. Response time depends on how quick each individual respond to the questionnaire: the time it took from the Knowledge centre sent out the first questionnaire to the patient to the date the Knowledge centre received and registered the questionnaire response from the patient. Response time is of secondary interest in the article, included as a covariate in the regression models. As the reviewer write, survey time is under the control of the hospitals and the Knowledge Centre. Survey time was computed as the difference between date for first postal mailing and discharge date. Survey time is the key independent variable in this study: is there an association between how long it takes from discharge to the patient are mailed the questionnaire and patient-reported experiences? According to the survey protocol the number of survey time days should vary from approximately 1-15 days. However, a number of hospital transfers to the Knowledge centre were delayed which resulted in a survey time variation from 1-41 days (mean 12 days). This survey time variable was included as a main predictor in the regressions. Based on comments from the reviewers and further evaluation, the main statistical analysis was changed to multilevel regression to adjust for potential confounding effects related to the hospital level (see point 12 above).

A clarification about response time and survey time has been included in the Methods-Statistical analysis (third paragraph):
“Response time has been shown to be associated with patient-reported experiences (27). Response time was computed as the difference between the dates of response registration and first postal mailing. This variable is influenced by how rapidly each individual responded to the questionnaire, and was included as a covariate in the regression models described below. The main focus of the regression analysis was the association between survey timing and the level of patient-reported experiences, adjusted for all the other predictors including response time.”

23. Literature review is incomplete. The literature review contains one randomized trial of time between discharge and when survey sent based on a small sample (<2000) and several other observational studies that involved mode change. The authors say that the other observational studies which tend to find worse experiences for patients with greater delays when interviewed may be an artefact of change in survey mode and that the exception (based on the randomized study) which finds worse outcomes only for the longest duration group for two of six measures. The literature review can be strengthened by referring to Elliott et al. (2005; Health Services Research) CAHPS analysis. This is a very large observational dataset. They find that time between discharge and survey completion (which is related to survey protocol and patient characteristics, since patients with worse health outcomes are more difficult to track) is negatively related to patient experiences and this study did not allow mode to change over time.

The study by Elliott et al. is now referenced in the Methods/statistical analysis (see also point 22 above about survey time and response time). The present study also identified a negative association between response time and patient-reported experiences.

24. Description of the randomized trail can be more detailed. In particular, this study had a compelling design, but a very small sample size. The authors should note this limitation and that, with a larger sample size such as that used in the current study or CAHPS, it is likely that the lower total problem score reported by patients in the group with the greatest time since discharge (>5 weeks) group would have reached statistical significance. As it is, two of six measures that go into the total score did reach statistical significance.

See point 10 above where limitations of this study are included in the revised paragraph.

25. The recommendations are misleading. The authors conclude that time since discharge (up to 5 weeks) is unrelated to patient experiences and therefore survey firms have flexibility and when they administer. But the results show that as in other studies (including CAHPS), patients in the group with more days since discharge have worse outcomes and we know from other work that patients with more health problems tend to be more critical of their health experiences. Hospitals that wait up to 5 weeks will disadvantage the health profile of their respondents and their ratings. Some flexibility is probably ok (and probably necessary from an implementation point of view), but adjustments should be made to account for both time since discharge and patient health status to ensure a level playing field for hospitals being evaluated. These results are consistent with such adjustments.
After accounting for the hospital cluster effect, the present study found a significant negative association between survey time and patient-reported experiences for three of six scales. Consequently, recommendations have been changed (see point 12 above).

26. The article would also benefit from a close read by the BMC copywriter to clarify some of the English usage, but is generally well-written.

A professional editing service has been used to improve the style of written English.

27. If “response time” refers to time between discharge and when patient completes the survey, then the authors should clarify this and should also note that response time is a function of time it takes to track patient which is confounded with patients’ characteristics (e.g., institutional transfers, readmissions, etc.) and survey protocol.

See point 22 above.