

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Patient-reported experiences with general practitioners: a randomized study of mail and web-based approaches following a national survey
AUTHORS	Iversen, Hilde Hestad; Holmboe, Olaf; Bjertnaes, Oyvind

VERSION 1 – REVIEW

REVIEWER	DR WILLIAM E HOGG L'Institut du Savoir Montfort, University of Ottawa, Canada
REVIEW RETURNED	19-Jan-2020

GENERAL COMMENTS	Well conducted study. It might help if you provide a sample size justification and explain why you oversampled Group A. Also you might explain why you included all GPs from the smaller practices and not from the larger practices.
-------------------------	---

REVIEWER	Manuella Lech Cantuaria University of Southern Denmark, Denmark
REVIEW RETURNED	06-May-2020

GENERAL COMMENTS	<p>This article is aimed at comparing response rates, background characteristics, data quality and main study results for a survey of patient experiences with general practitioners (GPs) administered by mail and web-based approaches.</p> <p>In general, the article is well-structured and concise. However, it needs to be significantly improved prior to publication, especially in terms of the overall objective definition (and background behind the objective), the methodology description and the discussion of the results obtained. Besides, more details need to be provided in terms of outcome definition and study limitations. A non-response analysis needs to also be included. My specific comments are the following:</p> <p>Abstract:</p> <ol style="list-style-type: none">1) Page 2, line 8 (i.e. objectives): What the authors describe here are the specific objectives. What is the overall reason for the authors to do these comparisons? Is it to evaluate the potential use of Web-surveys in Norway? This should be clarified in one or two sentences.2) Page 2, line 23 (i.e. Settings): More details need to be provided here.3) Page 2, line 29 (i.e. Participants): According to the journal's instructions, more information should also be provided (e.g. number of participants).4) Page 3, line 5 (i.e. Conclusions): The general conclusion should also be reformulated. Here, it reads more as a repetition of the results. What is the final message from the study? <p>Introduction:</p>
-------------------------	--

	<p>1) Page 5, line 10: the abbreviation GP need to be assigned here. Besides, how does the system work in case of immigrants and individuals born after 2001? The first sentence should be rewritten, so this information is clear for the readers.</p> <p>2) Page 5, line 37-42: This sentence (“the results from two previous randomized studies and also studies of survey-mode preferences in different patient populations indicate that there is a rather modestly developed web mode preference overall.”) is rather confusing. Please re-write it, so that is clear why this is a motivation for the study.</p> <p>3) Overall, the introduction needs to be better developed, in order for us to understand the overall objective of the study.</p> <p>Methods:</p> <p>1) Page 6 (line 32- 46): The text of the first paragraph of the Methods section is a bit confusing. I suggest the text to be rewritten and the inclusion of a figure illustrating the sampling plan.</p> <p>2) Page 6 (line 55 – 58): More details are needed in terms of the design and response options offered to participants. For example: For group A, a printed version of the questionnaire is sent together with the invitation? How do they return the printed questionnaire after filled up? Does it imply any cost to them? And for the electronic response, was an URL provided for them to access the questionnaire? More info here is necessary as those can explain the differences in response rate, why most people in group A answered the paper version, etc. This should also be better discussed in the discussion section and included as a limitation. Besides, for the participants of group B, do you have any registration of how many e-mail addresses were valid? And how many residents have their e-mail registered in the national system?</p> <p>3) Page 7 (line 3-6): Similar to the previous comment: More information regarding the reminders need to be provided. For example, for group A, is the printed version of the survey sent again together with the reminder? What about the URL? Besides, how much time did it take for reminders to be sent? All of those points are very important to understand the results found.</p> <p>4) Page 7 (line 22-29): Information on the approvals should be given together in the section Approval in page 8. Some information is duplicated.</p> <p>5) Page 7 (line 26): Where is the nonresponse analysis? This is indeed important for the study, but I cannot see it here.</p> <p>6) Page 7 (Measurements section): More information is needed to explain how the responses were converted to a scale from 0-100.</p> <p>7) Page 8 (Statistical analysis): Methods for calculating response rates should be specified. See AAPOR: https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx</p> <p>8) Page 9 (Patient and public involvement section): This section should be moved to the beginning of the Methods, where the survey is explained. This should also be more elaborated. Besides, the authors say” Patients were included in the development process of the instrument”, What exactly does this mean?</p> <p>Results</p> <p>1) Page 9 (line 53): Why was 0.7 the criterion chosen for cronbach's alpha. References are also needed. This can be included in the Methods section.</p> <p>2) Page 10 (line 17 – to 22): The sentence “The proportion of patients aged 30–49 years was higher in Group B than in Group A, and while Group A contained a higher proportion of patients who were aged >= 67 years” is confused and should be rewritten.</p> <p>3) Page 10 (line 22): You say “web respondents were more likely”.</p>
--	---

	<p>What are the web-respondents? Is it only people from group B or it also includes web respondents from group A? The table only compares group A with group B, so the sentence needs to be rewritten. Besides that, the sentence does not reflect necessarily the results that are shown in Table 2.</p> <p>4) Page 10 (line 38): According to the table, the question where the difference between responses is 3.5 is the waiting time for NOT urgent appointments. Results should then be revised.</p> <p>5) A nonresponse analysis is really needed in this paper. It is hard to draw conclusions from the differences between the respondents if we do not know whether the samples are representative. It may be, for example, that only younger residents have their e-mail addresses in the national register) – this should be very well addressed in the discussion.</p> <p>Discussion</p> <p>1) Page 11 (line 40-45): You talk here about a CAHPS survey. First, what CAHPS means? Second, it is necessary that more details on the study design are given, so the reader can understand the comparison between this and their study.</p> <p>2) Page 12 (line 36 – 40): Again, a non-response analysis needs to be included here, to know whether these results reflect the actual situation. Most importantly, we need to know whether the Group A and Group B samples were representative. If, for example, only younger residents have their e-mail address registered in the national system, it is expected that Group B respondents are younger. This information should be thoroughly discussed in the article.</p> <p>3) In general, the discussion needs to be improved, so that it includes a comparison between the results found in this study with previous studies.</p> <p>4) Discussion repeats results and lack information on different important aspects: a) effect of reminder letters in the different groups; b) limitations of the study in terms of the design; c) Differences in e.g. missing data.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Well conducted study.

It might help if you provide a sample size justification and explain why you oversampled Group A.

We agree with your comment, the reason for the oversampling of Group A was not explained thoroughly enough in the manuscript. The aim of the current study was to compare the standard mail survey mode of data collection with web-based data collection in Norway. However, the survey was commissioned by The Ministry of Health and Care Services, and the main aim was to evaluate the GP scheme. Part of this evaluation comprised the national patient experience survey, and the sampling plan aimed to give a nationally representative sample and results.

The standard data collection procedure in national patient experience surveys in Norway is post-discharge postal surveys, with pen-and-paper questionnaire included and an option to answer electronically. Previous surveys and research we have conducted in this area has shown that web mode surveys have lower response rates than other modes. In the national survey among patients visiting GPs in Norway in 2014, 18% of the respondents chose the electronic response option. However, comprehensive Web-based surveys is a new, potentially cost-effective and quick data collection method that must be developed and tested. The current study of patient experiences with GPs is the first to explore a purely electronic protocol in the national program for monitoring and

reporting on health-care quality using patient experience surveys in Norway. Also, we had not previously explored the quality of the email addresses collected from the national register for contact information.

Previous research has shown only marginal differences in patient experiences between patients in web-based surveys and other modes, giving some support to the lack of association between low response rate for web-based surveys and the amount of non-response bias. However, low response rates threaten the legitimacy of surveys at the clinic and in the public domain, as well as the ability of the surveys to identify important differences in patient-reported experiences between providers and over time.

Accordingly, considering the commission and the lack of experiences from a purely electronic protocol, we evaluated the risk of randomizing the total sample in two groups as too high, and chose to include fewer patients in the subsample.

We have added the following text in the Methods section, page 8, lines 18-24:

“The current study was the first to explore a purely electronic protocol in the national program of patient experience surveys in Norway. Also, we have not previously explored the quality of the email addresses collected from the national register for contact information. Considering the commission of achieving national representative results and the uncertainty regarding the responses from a purely electronic protocol, we evaluated the risk of randomizing the total sample in two groups as too high and chose to include fewer patients in the subsample.”

Also you might explain why you included all GPs from the smaller practices and not from the larger practices.

Thank you for the relevant comment. The sampling plan aimed to give a nationally representative sample.

GP practices were randomly selected after stratification by the number of GPs at the practices and the municipality types. The preconditions for the sampling frame were to report the results on a national level and to be able to estimate satisfactory intraclass correlation coefficients on the GP practice level. We did not aim to benchmark at the GP level. With the patient sample size chosen, we explored how the ICC varied dependent on the number of GPs at the practice level and found that at least four GPs were needed per GP practice to reach an acceptable ICC, and that not much were gained by including more GPs per practice.

We have made the following adjustments in the Method section, added to highlight the sampling plan and stratification (page 8, lines 4-9):

“The preconditions for the sampling frame were to report the results on a national level and to be able to estimate intraclass correlation coefficients on the GP practice level. We did not aim to benchmark at the GP level. With the patient sample size chosen, we explored how the ICC varied dependent on the number of GPs at the practice level and found that at least four GPs were needed per GP practice to reach an acceptable ICC, and that not much were gained by including more GPs per practice.”

Reviewer 2

This article is aimed at comparing response rates, background characteristics, data quality and main study results for a survey of patient experiences with general practitioners (GPs) administered by mail and web-based approaches.

In general, the article is well-structured and concise. However, it needs to be significantly improved prior to publication, especially in terms of the overall objective definition (and background behind the objective), the methodology description and the discussion of the results obtained. Besides, more details need to be provided in terms of outcome definition and study limitations. A non-response analysis needs to also be included. My specific comments are the following:

We agree with your comments, and the issues referred to above is now addressed, see our comments below.

Abstract

1) Page 2, line 8 (i.e. objectives): What the authors describe here are the specific objectives. What is the overall reason for the authors to do these comparisons? Is it to evaluate the potential use of Web-surveys in Norway? This should be clarified in one or two sentences.

Thank you for these useful suggestions. We have added the following under “objectives” in the Abstract:

“The standard data-collection procedure in the national patient experience survey programme in Norway is post-discharge mail surveys, which include a pen-and-paper questionnaire and an option to answer electronically. A purely electronic protocol has not previously been explored, and the aim of this study was to compare response rates, background characteristics, data quality and main study results for a survey of patient experiences with general practitioners (GPs) administered by the standard mail data-collection procedure and a web-based approach.”

2) Page 2, line 23 (i.e. Settings): More details need to be provided here.

Thanks. We have changed the text to the following:

“In Norway every inhabitant is assigned to an individual GP, and the setting for the current survey is patients of GPs. Regular GP practices were randomly selected after stratification by the number of GPs at the practices and the municipality types.”

3) Page 2, line 29 (i.e. Participants): According to the journal's instructions, more information should also be provided (e.g. number of participants).

The text is changed to the following:

“The sample consisted of 6,999 patients aged 16 years and older registered with a GP in November 2018.”

4) Page 3, line 5 (i.e. Conclusions): The general conclusion should also be reformulated. Here, it reads more as a repetition of the results. What is the final message from the study?

“Conclusions” has been adjusted as follows:

“Web-based surveys are faster and less expensive than standard mail surveys, but their low response rates and coverage problems threaten their usefulness and legitimacy. Initiatives to increase response rates for web-based data collection, more non-response research and strategies for tailoring data collection to different groups should be key elements in future research.”

Introduction

1) Page 5, line 10: the abbreviation GP need to be assigned here. Besides, how does the system work in case of immigrants and individuals born after 2001? The first sentence should be rewritten, so this information is clear for the readers.

The abbreviation GP is written out in full in the first sentence, and we have changed the information on the GP scheme to the following (page 5, lines 2-5):

“Norway introduced the regular general practitioner (GP) scheme in 2001. All inhabitants who are registered in the National Registry as living in Norway have the right to a GP/family doctor. Migrants eligible to stay in Norway for more than six months are entitled to enrol in the scheme.”

2) Page 5, line 37-42: This sentence (“the results from two previous randomized studies and also studies of survey-mode preferences in different patient populations indicate that there is a rather modestly developed web mode preference overall.”) is rather confusing. Please re-write it, so that is clear why this is a motivation for the study.

Thank you for these useful suggestions. We have tried to explain more explicitly why this is a motivation for the study and have rewritten the current section.

Page 5, lines 19-22; page 6, lines 1-6:

“The results from previous studies of survey-mode preferences in different patient populations both in Norway and other countries indicate that there is a rather modestly developed web mode preference.^{2–11} In the national patient experience survey among patients visiting general practitioners in 2014 in Norway, only 18% of respondents answered electronically.

A main limitation of previous studies has been the lack of e-mail addresses in the sample frame, and the implication that even the electronic group had to be invited by a postal invitation, adding to costs, and precluding the possibility of testing a comprehensive electronic data collection option. A purely electronic protocol has not previously been explored in the national program for monitoring and reporting on health-care quality using patient experience surveys in Norway.”

Page 7, lines 12-13:

“We considered these potential advantages and possibilities as important arguments for performing further research into web-based surveys.”

3) Overall, the introduction needs to be better developed, in order for us to understand the overall objective of the study.

See point 2 above, we have rewritten the Introduction section.

Methods

1) Page 6 (line 32- 46): The text of the first paragraph of the Methods section is a bit confusing. I suggest the text to be rewritten and the inclusion of a figure illustrating the sampling plan.

Thank you for the relevant comment. We have rewritten the paragraph to the following (page 8, lines 3-14):

“The sample consisted of patients aged 16 years and older registered with a GP in November 2018. The preconditions for the sampling frame were to report the results on a national level and to be able to estimate intraclass correlation coefficients on the GP practice level. We did not aim to benchmark at the GP level. With the patient sample size chosen, we explored how the ICC varied dependent on the number of GPs at the practice level and found that at least four GPs were needed per GP practice to reach an acceptable ICC, and that not much were gained by including more GPs per practice. The sampling plan had a three-stage design. First, regular GP practices were randomly selected after stratification by the number of GPs at the practices and the municipality types. Second, all the GPs were included in the selected practices that had up to four GPs, while four of them were randomly selected in the practices that had five or more GPs. Third, we randomly selected 14 adult patients from the list of patients of each GP.”

2) Page 6 (line 55 – 58): More details are needed in terms of the design and response options offered to participants. For example: For group A, a printed version of the questionnaire is sent together with the invitation? How do they return the printed questionnaire after filled up? Does it imply any cost to them? And for the electronic response, was an URL provided for them to access the questionnaire? More info here is necessary as those can explain the differences in response rate, why most people in group A answered the paper version, etc. This should also be better discussed in the discussion section and included as a limitation. Besides, for the participants of group B, do you have any registration of how many e-mail addresses were valid? And how many residents have their e-mail registered in the national system?

We have included additional information about the the national register for contact information and design and response options in the following sections:

Introduction (page 6, line 17-21):

“The use of Internet in the population is growing. In 2018, 90% of all Norwegian citizens used the Internet at a daily basis.¹⁶ In all age groups under 60 years, between 90-99 percent reported to use the Internet daily, but corresponding results for those between 60-69 years was 81% and for those aged 70 years or more 67%. Seventeen percent of the citizens aged 70 years or more reported that they never used the Internet.”

Introduction (page 6, line 23-24):

“A total of 88% of the population was registered in the national register for contact information in November 2018. 17”

Methods (page 9, line 1-12):

“Patients in Group A were mailed an invitation with both pen-and-paper and electronic response options. The invitation included a cover letter describing the purpose of the study, a paper

questionnaire, a prepaid envelope and information and a login code to be able to respond electronically. The patients in Group B received an email invitation with an electronic response option only. The email invitation included information about the purpose of the study, a link to the online survey and a login code. Two reminders were sent to non-respondents in both samples using the same contact mode as for the first invitation. The first reminder was sent to both groups around three weeks after the first contact. The second reminder was sent around six weeks after the first contact. All reminders to Group A were sent by mail and included a new invitation, the paper questionnaire, the postage-paid envelope and the login code to enable electronic responses. Group B were sent a new email invitation with a link to the survey and a login code in both reminders.”

Results (page 12, line 11-13):

“15% of the patients in the electronic arm lacked a valid email address in the national register, and 5% of the patients in the standard mail survey mode lacked a valid mailing address (Fig. 1).”

Discussion (page 15, line 8-10):

“However, as many as 15% of the patients in the electronic arm lacked a valid email address in the national register, the corresponding number we could not reach in the standard mail data-collection was 5%.”

Discussion (page 16, line 6-10):

“The surveys were designed to be as similar as possible, including the invitation letter, the content, layout and structure of the questionnaire and the timing of the first contact and reminders. The invitations to the patients in Group A and Group B were sent the same week and non-respondents in both groups received two reminders”.

3) Page 7 (line 3-6): Similar to the previous comment: More information regarding the reminders need to be provided. For example, for group A, is the printed version of the survey sent again together with the reminder? What about the URL? Besides, how much time did it take for reminders to be sent? All of those points are very important to understand the results found.

See our response to 2) above.

4) Page 7 (line 22-29): Information on the approvals should be given together in the section Approval in page 8. Some information is duplicated.

As suggested, we have removed the text on page 7, and changed the approval section on page 12 to the following:

“The Data Protection Officer at the NIPH recommended that the study be approved, and it was formally approved by the research director of the division for health services at the NIPH. The Norwegian Directorate of Health approved the use of data about non-respondents in the nonresponse analysis, except those of patients who withdrew themselves from the study.

Return of the questionnaire represented patient consent in the study, which is the standard procedure in all patient experience surveys conducted by the Norwegian Institute of Public Health.”

5) Page 7 (line 26): Where is the nonresponse analysis? This is indeed important for the study, but I cannot see it here.

Thank you for raising an important question. We have now conducted non-response analysis for both samples, with all the independent variables available for non-respondents. The results are shown in table 4. The following changes have been made to the Introduction, Methods, Results and Discussion sections:

Introduction (page 6, line 8-15):

“The literature on the effects of background characteristics on the responses to different data collection methods are inconsistent.^{2–10} Non-response bias has been studied in four patient populations in Norway through follow-up telephone interviews with non-respondents, ^{12–15} including non-respondents in a survey on patient experiences with GP. ¹⁵ The results have shown minor differences between the postal respondents from the national surveys and the postal non-respondents who have provided answers through follow-up interviews. In general, the impact of non-response bias in the large-scale surveys have been considered relatively small.”

Introduction (page 7, line 2-7):

“Potential variations in the population coverage between paper- and web-based questionnaires and the risk of selection bias from using the Internet for questionnaire surveys are reduced, but a major concern with protocols that use only digital responses is leaving out people without available digital contact information. When comparing the standard mail survey mode of data collection with web-based data collection the characteristics of non-respondents and respondents in both groups should be explored.”

Methods (page 11, line 15-20):

“Differences in respondent characteristics between respondents and non-respondents in Group A and respondents and non-respondents in Group B were also tested using Pearson chi-square tests for categorical variables and independent-samples t-tests for continuous variables. Variables available on non-respondents were gender, age, time on the list of the GP, number of consultations during the past 24 months and number of diagnosis during the past 24 months.”

Results (page 13, line 22-25; page 14, line 1-4):

“Significant differences were found between Group A and Group B within respondents and non-respondents with respect to gender and age (Table 4). Non-respondents tended to be more likely to be men and to be younger than respondents in both groups. Significant differences were also found for time on the list of the GP, number of consultations during the past 24 months and the two variables about number of diagnosis the last two years for Group A. Respondents tended to have been longer on the GPs list, and to have a higher number of consultations and diagnosis during the last two years. We found no additional significant differences between respondents and non-respondents in Group B.”

Discussion (page 17, line 1-14):

“There are several methods for assessing non-response bias, including comparison of respondents and non-respondents on background variables.²⁵ When we compared respondents with non-respondents, we found that men and younger patients were underrepresented as respondents in both groups. These differences are normally handled by non-response weighting, but such weights are only able to compensate for variables available in the sampling frame. We did not conduct further analysis of non-respondents, but previous follow-up studies of non-respondents in Norway indicate small additional bias¹²⁻¹⁵ However, none of these have included a purely digital protocol, which warrant future non-response research for digital protocols. The coverage challenges for the digital sampling frame should be part of this research, as 12% of the population was not registered in the register, and 15% of the registered persons lacked a valid email address. This coverage challenge is an additional weakness of purely digital approaches and should be compensated with other response options for those excluded.”

6) Page 7 (Measurements section): More information is needed to explain how the responses were converted to a scale from 0-100.

New sentence under Measures (page 10, line 6-7):

“Single item and index scores were transformed linearly from the 1 to 5 scale to a scale of 0–100.”

7) Page 8 (Statistical analysis): Methods for calculating response rates should be specified. See AAPOR: <https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx>

We have included additional information about how the response rate was calculated in the suggested paragraph (page 10, line 22-24):

“The survey response rate by group was calculated as the proportion of eligible patients (ie, not those who had moved to a new house, died, or were otherwise ineligible) and who returned a completed survey (AAPOR response rate 4.0).¹⁸”

8) Page 9 (Patient and public involvement section): This section should be moved to the beginning of the Methods, where the survey is explained. This should also be more elaborated. Besides, the authors say “Patients were included in the development process of the instrument”, What exactly does this mean?

Thanks. We have included the Patient and Public Involvement section after Measures, and changed the text according to the comment above (page 10, line 12-19):

“Patients were included in the development process of the instrument, to secure that the questionnaire included the most important topics for patients. To identify important topics, we assessed reviews of the literature and consulted a reference group comprising GPs, researchers and representatives from health authorities and patient organisations throughout the process of questionnaire development. The questionnaire was tested through cognitive interviews with patients. First, eight face-to-face interviews and nine telephone interviews were conducted. After an extensive revision, we conducted another 11 face-to-face interviews with patients. The revised version was tested in a pilot study.”

Results

1) Page 9 (line 53): Why was 0.7 the criterion chosen for cronbach’s alpha. References are also needed. This can be included in the Methods section.

Thanks for the useful suggestions. We have included both information on the criterion and references in the relevant paragraph at page 11 (line 3-6):

“The item-total correlation coefficient quantifies the strength of an association between an item and the remainder of its indicator, with a coefficient of 0.4 considered acceptable.¹⁹ Cronbach’s alpha assesses the overall correlation between items within an indicator, and an alpha value of 0.7 is considered satisfactory.^{19,20}”

2) Page 10 (line 17 – to 22): The sentence “The proportion of patients aged 30–49 years was higher in Group B than in Group A, and while Group A contained a higher proportion of patients who were aged \geq 67 years” is confused and should be rewritten.

Thanks, we have changed to (page 13, line 13-16):

“The proportion of patients aged 30–49 years was higher in Group B than in Group A (37.4% compared to 23.8%). In group A, 31.8% of the patients were aged \geq 67 years, a much higher proportion than in group B where the corresponding proportion was 19.4%.”

3) Page 10 (line 22): You say “web respondents were more likely”. What are the web-respondents? Is it only people from group B or it also includes web respondents from group A? The table only compares group A with group B, so the sentence needs to be rewritten. Besides that, the sentence does not reflect necessarily the results that are shown in Table 2.

The current result refers to Table 3, not Table 2. We have changed the sentence, see below (page 13, line 16-17).

“The respondents in Group B were more likely to report that they had been in contact with their GP during the previous month than respondents in Group A.”

4) Page 10 (line 38): According to the table, the question where the difference between responses is 3.5 is the waiting time for NOT urgent appointments. Results should then be revised.

Thanks. We have corrected the text according to the results in the table (page 14, line 6-8).

5) A nonresponse analysis is really needed in this paper. It is hard to draw conclusions from the differences between the respondents if we do not know whether the samples are representative. It may be, for example, that only younger residents have their e-mail addresses in the national register – this should be very well addressed in the discussion.

See our responses to point 5 under Methods. The Discussion section is adjusted (page 15, line 5-11).

“The vast majority (88%) of the Norwegian population is included in the national register and uses the Internet on a daily basis, reducing potential variations in the population coverage between paper- and web-based questionnaires and the risk of selection bias from using the Internet for questionnaire surveys. However, as many as 15% of the patients in the electronic arm lacked a valid email address in the national register, the corresponding result for patients in the standard mail data-collection was 5%. Furthermore, only 18% of the contacted sample in the web-based approach responded.”

Conclusions (page 18, line 19-20):

“Men and younger patients were underrepresented as respondents in both groups.”

Discussion

1) Page 11 (line 40-45): You talk here about a CAHPS survey. First, what CAHPS means? Second, it is necessary that more details on the study design are given, so the reader can understand the comparison between this and their study.

Thanks. We have now included the following explanations in the last part of the relevant paragraph (page 15, line 14-17):

“A recent Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey produced corresponding results when comparing protocols based on web responses via an email invitation and mail.⁷ The mail protocol yielded more than twice the response rate of the web approach.”

2) Page 12 (line 36 – 40): Again, a non-response analysis needs to be included here, to know whether these results reflect the actual situation. Most importantly, we need to know whether the Group A and Group B samples were representative. If, for example, only younger residents have their e-mail address registered in the national system, it is expected that Group B respondents are younger. This information should be thoroughly discussed in the article.

Thanks. See our previous comments regarding non-response, especially point 5 above.

3) In general, the discussion needs to be improved, so that it includes a comparison between the results found in this study with previous studies.

Thank you. We have rewritten the Discussion section according to your comment. See also our responses to previous points above.

4) Discussion repeats results and lack information on different important aspects: a) effect of reminder letters in the different groups; b) limitations of the study in terms of the design; c) Differences in e.g. missing data.

You are right, this should be specified. We have rewritten the discussion.

VERSION 2 – REVIEW

REVIEWER	Manuella Lech Cantuaria University of Southern Denmark, Denmark
REVIEW RETURNED	05-Jul-2020

GENERAL COMMENTS	<p>The authors have addressed most of my previous comments. There are still some minor comments, but in general, I believe the article has improved considerably and can be considered for publication.</p> <p>Abstract:</p> <p>1) Page 2, line 13 (i.e. Settings): What is meant by “stratification by number of GPs”? Also, is it correct to say “the setting for the current survey is patients of GPs”? Consider rewriting the entire “Settings” section.</p> <p>2) Page 3, line 10 (i.e. Conclusions): Rewrite the term “non-response research” or consider removing it.</p> <p>Introduction:</p> <p>1) Page 5, line 3: Consider replacing “as living in Norway” by “Norwegian residents”.</p> <p>2) Page 5, line 20: Replace the term “rather modestly developed” by</p>
-------------------------	--

	<p>something else, in order to improve sentence clarity.</p> <p>3) Page 6: Consider changing the paragraphs orders. The first paragraph (line 1-6) could be closer from the final paragraph in page 7, as it directs the reader to the study's objective. Besides, the 3 new paragraphs (page 6, line 8 to 25 and page 7, lines 2-7) should be summarized, in order to improve clarity of ideas.</p> <p>Methods:</p> <p>1) Page 8 (line 5-6): What do you mean by "We did not aim to benchmark at the GP level? Consider rewriting.</p> <p>2) Page 8 (line 7): State what ICC refers for.</p> <p>Discussion</p> <p>1) Page 16 (line 5): When it is said that "these results are in line with other findings" –which specific aspects are you actually referring to?</p> <p>2) Page 18: Please make clear here other limitations that your study has, e.g. limitations in the design and number of participants.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

Reviewer 2

The authors have addressed most of my previous comments. There are still some minor comments, but in general, I believe the article has improved considerably and can be considered for publication.

We want to thank the reviewer for positive and constructive feedback. We have tried to answer and incorporate the points in the resubmission and hope that the manuscript is acceptable for publication.

Abstract:

1. Page 2, line 13 (i.e. Settings): What is meant by "stratification by number of GPs"? Also, is it correct to say "the setting for the current survey is patients of GPs"? Consider rewriting the entire "Settings" section.

Thanks. We have rewritten the section to the following:

"General practitioner offices in Norway."

2. Page 3, line 10 (i.e. Conclusions): Rewrite the term "non-response research" or consider removing it.

We agree and have removed the term "non-response research".

Introduction:

1. Page 5, line 3: Consider replacing "as living in Norway" by "Norwegian residents".

Done.

2. Page 5, line 20: Replace the term "rather modestly developed" by something else, in order to improve sentence clarity.

We have changed the sentence to the following:

“The results from previous studies of survey-mode preferences in different patient populations both in Norway and other countries indicate that web mode surveys have lower response rates than other modes.2–11”

3. Page 6: Consider changing the paragraphs orders. The first paragraph (line 1-6) could be closer from the final paragraph in page 7, as it directs the reader to the study's objective. Besides, the 3 new paragraphs (page 6, line 8 to 25 and page 7, lines 2-7) should be summarized, in order to improve clarity of ideas.

The paragraph orders have been changed accordingly, and the new paragraphs on page 6 and 7 have been rewritten.

Methods:

1. Page 8 (line 5-6): What do you mean by “We did not aim to benchmark at the GP level? Consider rewriting.

Thanks, we deleted this confusing and redundant sentence.

2. Page 8 (line 7): State what ICC refers for.

We have rewritten the current sentence to the following:

“With the patient sample size chosen, we explored how the intraclass correlation coefficient (ICC) varied dependent on the number of GPs at the practice level and found that at least four GPs were needed per GP practice to reach an acceptable ICC, and that not much was gained by including more GPs per practice.”

Discussion

1. Page 16 (line 5): When it is said that “these results are in line with other findings” –which specific aspects are you actually referring to?

Thanks for a good suggestion. We have clarified by including the additional text in the sentence:

“These results are in line with other findings, that have shown only marginal differences in patient experiences and satisfaction between patients in web-based and other modes.2,3,7,8,9”

2. Page 18: Please make clear here other limitations that your study has, e.g. limitations in the design and number of participants.

Thanks. We have included the following about other limitations in the Discussion section:

“The survey was not linked to a specific contact with the GP or GP office, or actual use e.g. the last six months, which might have resulted in lower response rates and implies that we were unable to make any assumptions about specific contacts. Differences in respondent characteristics between respondents and non-respondents in both groups were tested, but not differences in patient reported

experiences since we lacked a follow-up study of non-respondents. However, the impact of non-response bias in previous large-scale surveys have been relatively small.12–15”

VERSION 3 – REVIEW

REVIEWER	Manuella Lech Cantuaria University of Southern Denmark, Denmark
REVIEW RETURNED	12-Sep-2020
GENERAL COMMENTS	The authors have addressed my previous comments from revision 1 and 2. In general, I believe the article has improved considerably and can be considered for publication.