PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form ([see an example](#)) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Who is More Likely to Use Doctor-Rating Websites, and Why? A cross sectional study in London.
AUTHORS	Miraldo, Marisa; Galizzi, Matteo; Stavropoulou, Charitini; Desai, Mihir; Jayatunga, Jeevana; Joshi, Mitesh; Parikh, Sunny

VERSION 1 - REVIEW

REVIEWER	Simone Ghislandi Assistant Professor Bocconi University, Milan, Italy I do not have competing interests with the study.
REVIEW RETURNED	10-Jul-2012

THE STUDY	Patients are not representative of the general population because the convenience sampling is from two areas of London. This limit however is correctly recognized by the authors. Statistical methods are appropriate but not adequately well described.
RESULTS & CONCLUSIONS	As specified in the attached revision file, results should be presented in a way more in line with a good medical journal.
GENERAL COMMENTS	**General comments:** The paper focuses on the role of internet as a source of health-related information among British citizens. In particular, the main objective of the work is to understand if and how the general British population uses doctor-rating websites. A total of 200 individuals are randomly selected from the borough of Hammersmith and Fulham in London. Age and race seem to represent the most important predictors of the use and awareness of these services However, (dis)satisfaction about own GPs is also an important factor, as it is the GP-patient gender concordance. The paper is important for two reasons. First of all, it represents one of the first attempt to investigate in detail how the information about health treatments and facilities flows to patients. In the general framework which dominates the policy agenda, the patient is typically assumed to be able to use efficiently the available information in order to always make the "best" choice. It becomes then very important to understand how this choice is really made and how information is acquired and processed for this purpose. Second, this paper provides a first profile of the standard user of doctor-rating websites, showing that it is not possible to consider these websites as used by the "general population". This particular finding is also important for the understanding of the statistics related to the healthcare management and marketing literature, since it proves that samples of websites users over-represent the youngsters and the non whites.

The paper is good but improvable. In particular, authors should clarify better the motivation and the pros of the study and should make the paper more readable to a more general population. Writing and tables must be simplified in order to meet the requirements of a good medical journal.

Specific comments:

INTRODUCTION

• It would be useful to insert a paragraph on the link between the objective of this study and the more general topic of the free choice by patients within a health system. This would facilitate the reader in understanding better the motivation of the analysis.

• Since there is not a section for the review of the literature, are the papers cited here in the Introduction all the relevant existing ones on the topic?

SURVEY DESIGN AND DATA COLLECTION

• What are the main differences between the convenience sampling used here and the benchmark of the random sampling? Could you please motivate better on the reasons for using convenience sampling?

• Within the areas, how were the subjects chosen? Was there some stratified sampling?

• Who conducted the survey? When and at what time of the day?

DATA ANALYSIS

• Table 1. This Table should be extended, reorganized and used as a reference for the paper. Besides the description, it should provide the basic statistics about the respondents (e.g. how many respondents for each items, mean age and sex).

• The part describing the variables in page 7 should be cut and substituted by a reference to the description column in Table 1. This should facilitate the reader and give a better organization to the paper.

• There is no need of reporting the formula of a probit, which is a well-known non-linear regression method.

• Why did authors decide to use a probit instead of a more common logit? In medical journals, readers are more familiar with odds ratio rather than coefficients. On the other hand, could authors explain better what is an ordered probit and why it is introduced here?

• Again, table 2 should include OR and standard errors. Also, what is the need of reporting 4 models?

• How trustable are the answers to the questions about the likelihood to use online rating websites?

• Table 3 is difficult to read. It could be a good idea to create a more synthetic set of variables, either taking a stepwise regression and showing only the final results or by selecting what authors consider the variables that are more likely to be the relevant ones.

• Also, I am not sure there is the need to use seventeen different specifications. Could author provide a more synthetic set of results?

REVIEWER	Tara Lagu, MD MPH Research Scientist, Center for Quality of Care Research Baystate Medical Center Assistant Professor, Tufts University School of Medicine 280 Chestnut St. 3rd fl

REVIEW RETURNED	17-Jul-2012

| **THE STUDY** | This is a study that addresses an important gap in the literature: what are the characteristics of patients who use doctor rating sites? The authors surveyed a convenience sample of 200 people to determine if they were aware of the sites, which is novel and important, but the study is plagued with significant limitations. I could not determine the content of the survey, which is important, because awareness of such sites is one thing, actual use of the sites is something else entirely and is not addressed in this manuscript, as far as I can tell. Also, this a 200 person survey (relatively small) and most medical readers would benefit from simple descriptive statistics and associations. The authors leave this out and, I feel, make the analysis unnecessarily complicated. Leaving out a lot of what the medical reader would consider to be important significantly weakens the manuscript. Finally, the format of this manuscript is atypical for a medical journal-usually it's intro, methods, results, discussion. This one is intro, survey design, data analysis, results and discussion combined. It would not be difficult to alter the manuscript for a medical audience, but I would definitely recommend it before publishing this manuscript in BMJ.

The introduction is quite long and includes information that should probably be moved into the discussion section (e.g., paragraphs 2, 3).

Methods:

I suggest that the authors include a copy of the questionnaire as an appendix (and not just the variable). The content of the survey is unclear. What exactly were they trying to determine? Just awareness of the sites? Or the degree to which they use the sites? These are definitely two different things but are not distinguished by the authors.

The convenience sample is a limitation of the study. It is unclear who, exactly, the respondents represent, but it does seem that they most represent the neighborhood in which data collection occurred, which skews younger and unemployed-seems like they went into public places in the middle of the day, when the proportion of unemployed people is higher (because everyone who is employed is at work).

In the data analysis section, the authors begin presenting results, which I find confusing.

The number of responses is never stated as far as I could find, but I think is 200, 29 of whom were aware of doctor-rating websites. The authors then construct an elaborate model to examine predictors of doctor-rating sites. This, to me, skips an important step: examining the frequencies of age groups, ethnicities, education, income, etc.-things that a general reader would want to know (and could readily understand). They also do not examine bivariate relationships which would help determine which variables belong in the model. |
| **RESULTS & CONCLUSIONS** | Table 3 is not appropriate or necessary for the general medical reader, except perhaps as an appendix.

The discussion section does not cite the prior literature that has reported on many of these phenomena in the past (e.g., use of the |

	internet by age, gender, etc, use of health care data by consumers (pew, Kaiser family foundation) and does not cite literature on the content and scope of physician rating websites in England and the US (authors: Wachter, Lagu, Lopez, Jha) (at least one of these includes NHS choices). These should be included as background at least.
	Furthermore, the discussion does not seem to center on the most important question for clinicians: what is the significance of the lack of patient awareness? Are the 29 patients who used the data actually using it? Or are they just aware of it? Do these sites contribute towards improving care quality or do they just "add to the noise?"

VERSION 1 – AUTHOR RESPONSE

REVIWER 1- Dr. Simone Ghislandi

R1.1
In particular, authors should clarify better the motivation and the pros of the study and should make the paper more readable to a more general population.

Reply:

We thank the referee for this suggestion. The general motivation and main contributions of the study are now explicitly stated and clarified from the very beginning of the article. Please see the new Introduction section in the main text.

R1.2
Writing and tables must be simplified in order to meet the requirements of a good medical journal.

Reply:

We fully agree with the referee's suggestion. We have now greatly simplified and clarified both the tables and the text all throughout the article in order to make it more readable to the broadest audience of medical journals. Please refer to the new tables in the Appendix section of the paper.

R1.3
It would be useful to insert a paragraph on the link between the objective of this study and the more general topic of the free choice by patients within a health system. This would facilitate the reader in understanding better the motivation of the analysis.

Reply:

We thank the referee for these very helpful comments. We have now immediately clarified the motivation and main contributions of the paper and added a paragraph explicitly linking the objective of the study to the more general topic of patients' choice. Please see the new Introduction section on pages 4-5.

R1.4

Since there is not a section for the review of the literature, are the papers cited here in the Introduction all the relevant existing ones on the topic?

Reply:

Although we agree with the referee that explicitly including a section on the literature's findings would be of potential interest, conducting a systematic review of the literature was outside the direct scope of the article. In the original version of the article, therefore, we only selected the studies that were relevant to the UK context.

In the revised version, however, we have now updated the introduction and discussion with more literature from the US as also suggested by the second referee. Please see new text in the Introduction and Discussion sections on pages 4-5 and 11-15 respectively.

R1.5

What are the main differences between the convenience sampling used here and the benchmark of the random sampling? Could you please motivate better on the reasons for using convenience sampling?

Reply:

The objective to involve respondents from the general population informed the choice of using subjects directly sampled from the field. An online survey would have exclusively reached the segment of already active internet users, thus failing to address the main goal of the study, whether the users of doctor-rating websites are fairly representative of the general public. The choice of convenience sampling was mainly dictated by practical and financial issues. Given the variety of respondents' observable characteristics between different London boroughs, a random sample of London inhabitants would have required a complex stratification sampling that was practically unfeasible to implement given the envisaged time frame and the available resource capacity.

We fully agree with the referee, however, that the convenience sampling employed in the analysis indeed presents limitations, and, in the article, in fact, we openly acknowledge those limitations of the analysis. We have also added a new section after the discussion of the results which explicitly addresses the limitations of the study.

The main differences between the convenience sampling used in this study and the benchmark of a random sampling as, for instance, obtained by the ONS census survey, is thoroughly discussed in the result section.

Please see the new Methods section on pages 5-7 and the new Discussion section on pages 11-15.

R1.6

Within the areas, how were the subjects chosen? Was there some stratified sampling? Who conducted the survey? When and at what time of the day?

Reply:

Following the referee's suggestion, we have now clarified the sampling process, timings and location of the data collection in the text. We have also added sentence on the text clarifying who and how

exactly conducted the survey. Please see the revised text on pages 6-7.

R1.7
Table 1. This Table should be extended, reorganized and used as a reference for the paper. Besides the description, it should provide the basic statistics about the respondents (e.g. how many respondents for each items, mean age and sex).

Reply:

As suggested by the referee, Table 1 has been thoroughly extended and reorganized to include descriptive statistics. Please refer to new Table 1.

R1.8
The part describing the variables in page 7 should be cut and substituted by a reference to the description column in Table 1. This should facilitate the reader and give a better organization to the paper.

Reply:

Following referee's suggestion, this part has now been removed and the new Table 1 is now referenced in the text.

R1.9
There is no need of reporting the formula of a probit, which is a well-known non-linear regression method. Why did authors decide to use a probit instead of a more common logit? In medical journals, readers are more familiar with odds ratio rather than coefficients. On the other hand, could authors explain better what is an ordered probit and why it is introduced here?

Reply:

We originally employed a binary probit and an ordered probit model to fit Awareness and the IntentionToUse discrete variables, respectively, to ensure a reasonable comparability between the empirical results obtained for the two set of regressions. The models, in fact, only differ in the number of values that the dependent variables can take, while the underlying structure of the error terms is the same. The probit model was originally selected because the error terms are assumed to follow a standardized Gaussian distribution, and the convergence of the maximum likelihood optimization routines is particularly quick for both the binary and the ordered probit models.

We agree with the referee, however, that readers of medical journal are more familiar with the alternative logistic specification. The logistic specification, in fact, is particularly appealing with respect to the probit counterpart because its results can be readily expressed in terms of odds ratio, which can be interpreted more intuitively than estimate coefficients. Following the referee's suggestion, in the revised version of the paper, we have therefore re-run all the sets of estimations using a binary and an ordered logit model to fit the Awareness and the IntentionToUse discrete variables, respectively, instead of the previously employed probit specifications. Also, following the referee's suggestion, estimate results are now presented in terms of the odds ratio, together, again, with the standard error, and levels of significance. We thank the referee for this helpful suggestion.

Results are now presented in the new Tables 3 and 4 (before labeled as Table 2 and 3), to be possibly included in the Appendix. As it can be seen, the new set of logistic regressions presents estimates that are qualitatively fully in line with the results previously presented in the probit models. The description of the methods, together with the discussion of the results, have been accordingly amended to mirror the new specification employed in the estimates. Please see new Results section on pages 8-11.

R1.10
Again, table 2 should include OR and standard errors. Also, what is the need of reporting 4 models?

Reply:

We have now removed the probit analysis and have carried an ordered logit as suggested by the referee. We indeed agree with the referee that readers of medical journal are more familiar with the alternative logistic specification. The logistic specification, in fact, is particularly appealing with respect to the probit counterpart because its results can be readily expressed in terms of odds ratio, which can be interpreted more intuitively than estimate coefficients. Following the referee's suggestion, in the revised version of the article, we have therefore re-run all the set of estimations using a binary logit model to fit the Awareness. Also, following the referee's suggestion, estimate results are now presented in terms of the odds ratio, together, again, with the standard error, and levels of significance.

Moreover, as suggested by the second referee, we have also included, among the regressors, two further variables that the pair-wise correlation analysis suggests to be positively correlated with Awareness, namely: the ordered discrete variables capturing the fact that the respondents consider as important factors in making their decisions i) the financial performance of the hospital (HC_FinPerform); and ii) the online doctor rating websites (SI_DoctorRating).

We have nevertheless decided not to include among the regressors other variables which also showed significant correlations, such as the binary variables for being aware of the existence of the Choose and Book system (CBAwareness), or for having internet access (Web_Access). We were in fact concerned with the possibility that those variables were likely to raise endogeneity or multi-collinearity issues, respectively, being the first essentially co-determined with the dependent variable, and showing the second a nearly perfect association with the awareness status.

In the new Table 3 (in the first version of the paper labelled as Table 2) we thus present the results for the estimates of four logit empirical models: model 1 directly corresponds to the first probit model presented in the previous draft, while models 2-4 extends the corresponding probit models 2-4 to the inclusion of the two above new variables. Following the suggestion by the referee, the results are again presented in a step-wise format.

As it can be seen, overall, the results from the logit regression analysis fully confirm the main findings presented in the previous draft, while show significance for few more variables. In particular, as it was the case with the probit specification, the regression analysis confirms that:
i) older individuals are less likely to be aware of the rating websites;
ii) in most specifications, white British and white non-British respondents appear less likely to be aware of the websites;
iii) among the other socio-demographic factors, only income is sometimes (marginally) significant, pointing to the fact that respondents with higher reported levels of income tend to be less aware of the websites, while neither education or gender turn out to be significant predictors of awareness
iv) among the characteristics of the providers that respondents consider important when making their

decisions on where to receive healthcare, in one specification the reputation of the doctor has a strong positive effect, while both clinical and financial performance rates of the providers show negative significant effects: thus, the respondents who consider the reputation of the doctor important in deciding where to receive care are more likely to be aware of the rating websites, while this is less often the case for respondents putting a higher weight on financial or clinical performance ratings;

v) concerning the sources of information, in one specification respondents who consider the hospital statistics important in deciding where to receive care were more likely to be aware of the rating websites, with a quite strong and significant effect;

vi) among the variables on doctor-patient relationship, whenever included in the regressors, the gender match between the GP and the patient predicts higher awareness of the website ratings, with a noticeable effect;

vii) finally, although in one specification the respondents who feel that their GPs spend a sufficient time in their consultation are less likely to be aware of the internet rating websites, both the statistical significance and the estimated odds ratio do not appear robust across specifications.

The text in the article has been amended accordingly. Please see the new Results and Discussion sections on pages 8-11 and 11-15.

R1.11
How trustable are the answers to the questions about the likelihood to use online rating websites?

Reply:

We reckon that the answers on the likelihood to use doctor rating websites in the future should not be considered a necessarily perfect predictor of actual website usage behaviour. An analysis of the determinants of the actual usage of the websites, however, would be affected by the more serious limitation of a very small variation in the dependent outcomes, with a very little proportion of the respondents reporting to actually use the websites: from the 29 respondents that were aware of the existence of doctor-rating websites only six reported to have used these websites. The limitation of such a little variation in the dependent outcomes would seriously reduce the predictive power and reliability of any regression analysis based on actual usage, while not eliminating the issues related to the fact that these are, in any case, self-reported data on usage behaviors.

R1.12-13
Table 3 is difficult to read. It could be a good idea to create a more synthetic set of variables, either taking a stepwise regression and showing only the final results or by selecting what authors consider the variables that are more likely to be the relevant ones.
Also, I am not sure there is the need to use seventeen different specifications. Could author provide a more synthetic set of results?

Reply:

Following the referee's suggestions, we have now greatly simplified the presentation of the estimation results. As suggested, we have now selected only six specifications of the empirical model that are likely to be the most relevant ones, and presented them in a step-wise format. The new Table 4 (previously labelled as Table 3) in the Appendix reports estimation results in terms of the odds ratio, together with the standard error, and levels of significance. Please see new Results section on pages 8-11, new Discussion section on pages 11-15 and the new Conclusions section on pages 15-16.

REVIEWER 2- Professor Tara Lagu

R2.1
I could not determine the content of the survey, which is important, because awareness of such sites is one thing, actual use of the sites is something else entirely and is not addressed in this manuscript, as far as I can tell.

Reply:

We thank the referee for this helpful comment. We fully agree with the referee that considering awareness is entirely a different question than looking at actual usage of sites, and that the article should clarify and explicitly address this important aspect.

On a related issue, we also agree with the referee that the answers on the likelihood to use doctor rating websites in the future should not be considered a necessarily perfect predictor of actual website usage behaviour. An analysis of the determinants of the actual usage of the websites, however, would be affected by the more serious limitation of a very small variation in the dependent outcomes, with a very little proportion of the respondents reporting to actually use the websites: from the 29 respondents that were aware of the existence of doctor-rating websites only six reported to have used these websites. The limitation of such a little variation in the dependent outcomes would seriously reduce the predictive power and reliability of any regression analysis based on actual usage, while not eliminating the issues related to the fact that these are, in any case, self-reported data on usage behaviours.

Following the referee's suggestions, we have now explicitly stated from the very beginning of the article the objectives of the analysis and the main content of the survey. We have also clarified and explicitly addressed the research questions and the dependent variables for the empirical analysis. Furthermore, in the revised draft of the article, we have added a paragraph in the text that thoroughly describes the content of the questionnaire in the survey. As required by the referee, we have also added the whole questionnaire in the Appendix. Please see the Introduction section on pages 4-5 and the new Methods section on pages 5-7.

R2.2
Also, this a 200 person survey (relatively small) and most medical readers would benefit from simple descriptive statistics and associations. The authors leave this out and, I feel, make the analysis unnecessarily complicated. Leaving out a lot of what the medical reader would consider to be important significantly weakens the manuscript.

Reply:

We are grateful to the referee for having raised this important point. Following the referee's suggestion, Table 1 has been thoroughly extended and reorganized to include descriptive statistics. Please refer to new Table 1. Moreover, we have now expanded the text to explicitly describe the main characteristics of our sample (please see page 8).

R2.3
Finally, the format of this manuscript is atypical for a medical journal-usually it's intro, methods, results, discussion. This one is intro, survey design, data analysis, results and discussion combined. It would not be difficult to alter the manuscript for a medical audience, but I would definitely recommend it before publishing this manuscript in BMJ.

Reply:

We thank the referee for this very helpful suggestion. We have now extensively re-organized the article to fit the typical format of a medical journal. As suggested by the referee, the manuscript is now divided in an introduction, three distinct sections on methods, results, and discussion, respectively, and a brief conclusion. Please refer to the new structure of the manuscript. In particular, the description of the results, therefore, is now separated from the discussion, which we hope will contribute to a higher clarity and readability of the article by the broad audience of a medical journal.

R2.4
The introduction is quite long and includes information that should probably be moved into the discussion section (e.g., paragraphs 2, 3).

Reply:

Following the referee's suggestion, we have now removed those paragraphs from the introduction and instead added them in the new section that has been added for discussing the results, as also required by the referee. We hope the introduction now reads better.
However, given the suggestions of the other referee, who required to introduce more background on patient choice and to discuss its links with our study, we have not been able to further cut the introduction. Please see the new Introduction section on pages 4-5.

R2.5
I suggest that the authors include a copy of the questionnaire as an appendix (and not just the variable).

Reply:

We thank the referee for this suggestion. We have now included a whole copy of the questionnaire in the Appendix.

R2.6
The content of the survey is unclear. What exactly were they trying to determine? Just awareness of the sites? Or the degree to which they use the sites? These are definitely two different things but are not distinguished by the authors.

Reply:

We thank the referee for this helpful comment. We totally agree with the referee that considering awareness is entirely a different question than looking at actual usage of sites, and that the article should clearly state what the content of the survey is. Following the referee's suggestions, we have now added a paragraph in the text that explicitly describes the content of the survey in detail. Please see page 6.

R2.7

The convenience sample is a limitation of the study. It is unclear who, exactly, the respondents represent, but it does seem that they most represent the neighbourhood in which data collection occurred, which skews younger and unemployed-seems like they went into public places in the middle of the day, when the proportion of unemployed people is higher (because everyone who is employed is at work).

Reply:

The choice of convenience sampling was mainly dictated by practical and financial issues. Given the variety of respondents' observable characteristics between different London boroughs, a random sample of London inhabitants would have required a complex stratification sampling that was practically unfeasible to implement given the envisaged time frame and the available resource capacity.

We fully agree with the referee, however, that the convenience sampling employed in the analysis indeed presents limitations, and, in the article, in fact, we openly acknowledge those limitations of the analysis. We have also added a new section after the discussion of the results which explicitly addresses the limitations of the study.

We also acknowledge that the study should better clarify which, and how, respondents were approached in the field, and more explicitly discuss any characteristics of the respondents that were likely to be over- or under-represented in the resulting sample. The referee is right when suspecting that the convenience sampling tends to over-represent respondents who are currently not employed, such as unemployed, retired and students.

Following referee's suggestion, on page 8 we have now expanded the discussion of the descriptive statistics, and explained in greater detail in the article who exactly the respondents were, as well as the sampling process, timings and location of the data collection. We have also clarified who, and how, conducted the survey. Moreover, the main differences between the convenience sampling used in this study and the benchmark of a random sampling as, for instance, obtained by the ONS census survey, is thoroughly discussed in the result section in pages 8-9.

R2.8

In the data analysis section, the authors begin presenting results, which I find confusing.

Reply:

Motivated by the referee's comment, we have now moved the descriptive statistics to the beginning of the results section. Please see the new results section on pages 8-9.

R2.9

The number of responses is never stated as far as I could find, but I think is 200, 29 of whom were aware of doctor-rating websites.

Reply:

We are grateful to the referee for having pointed out this source of ambiguity in the text. In order to meet the referee's comment, we have now explicitly reported, both in the main text and in a summary table containing the description of the variables and their descriptive statistics (Table 1 in the Appendix) the total number of respondents (200) as well as of all non-missing responses for each variable collected in the survey. Please refer to the new Table 1 for details. Furthermore, we have now explicitly stated also the number of respondents who were aware of the websites (29) and who were actually using them (6). Please see new text on the Methods and Results sections on pages 5-7 and 8-11.

R2.10
The authors then construct an elaborate model to examine predictors of doctor-rating sites. This, to me, skips an important step: examining the frequencies of age groups, ethnicities, education, income, etc.-things that a general reader would want to know (and could readily understand).

Reply:

We thank the referee for having addressed this important point. Following referee's suggestion, Table 1 has been thoroughly extended and reorganized to include the number of non-missing observations, frequencies, and descriptive statistics (e.g. mean and standard deviation). Please refer to new Table 1 for details. Moreover, we have now expanded the text to explicitly describe the main characteristics of our sample. Please see sub section "Sample" on pages 6-7.

R2.11
They also do not examine bivariate relationships which would help determine which variables belong in the model.

Reply:

As suggested by the referee, we have actually looked at bivariate correlations to inform our analysis as submitted in the previous draft of the article. We now explicitly report the pairwise correlations in Table 2 in the Appendix and refer to this table in the text on the Statistical Analysis section (see page 7).

We have opted, however, to base the decision of which variables to include in the empirical models on other considerations together with the significance levels reported in the correlation analysis. Pairwise correlations, in fact, although highly useful to get preliminary insights in the structure of the data, needed to be accompanied by considerations about potential issues of endogeneity, multi-collinearity, and omitted variables.

Endogeneity and multi-collinearity issues, for instance, pointed against the inclusion in the regression analysis of variables which could be endogenously determined with the dependent variables. That was the case, for instance, of the variables for being aware of the existence of the Choose and Book system (CBAwareness), or for having internet access (Web_Access) in our first set of estimates, to study the determinants of being aware or unaware of the doctor rating websites (Awareness).

Although these variables showed significant correlations with the Awareness binary dependent variable, these variables were essentially co-determined, or nearly perfectly collinear with the latter , thus failing to show sufficient variability in the data: for instance, all subjects who were aware of the websites (Awareness=1), turned out to indeed have internet access (Web_Access=1); while 26 out of 29 of those that were unaware of the rating websites (Awareness=0) also did not know the Choose and Book system (CBAwareness=0). For these reasons we opted to not include these variables as regressors in the probit, and then in the logit, estimates for the awareness.

On the other hand, in an attempt to prevent omitted variables problems, we decided to also include in the regressors variables which did not show significantly correlation with the awareness binary variable.

Following the suggestion by the first referee, we have now re-run the regression analysis for the binary Awareness, outcome using a logit specification instead of a probit. In particular, as requested by the first referee, we have now included both odds ratios and standard errors in Table 3 (previously labelled as Table 2), in place of the estimated coefficients only.

Moreover, as suggested by the second referee, we have now included, among the regressors, in some specifications of the new logit analysis of the determinants of Awareness two further variables that the pairwise correlation analysis suggests to be positively correlated with Awareness, namely: the ordered discrete variables capturing the fact that the respondents consider as important factors in making their decisions i) the financial performance of the hospital (HC_FinPerform); and ii) the online doctor rating websites (SI_DoctorRating).

In the new Table 3 (previously labelled as Table 2), we thus present the results for the estimates of four empirical models: model 1 directly corresponds to the first probit model presented in the previous draft, while models 2-4 extends the corresponding probit models 2-4 to the inclusion of the two above new variables. Following the suggestion by the first referee, the results are again presented in a step-wise format.

As it can be seen, overall, the results from the logit regression analysis fully confirm the main findings presented in the previous draft, while show significance for few more variables. In particular, as it was the case with the probit specification, the regression analysis confirms that:
i) older individuals are less likely to be aware of the rating websites;
ii) in most specifications, white British and white non-British respondents appear less likely to be aware of the websites;
iii) among the other socio-demographic factors, only income is sometimes (marginally) significant, pointing to the fact that respondents with higher reported levels of income tend to be less aware of the websites, while neither education or gender turn out to be significant predictors of awareness;
iv) among the characteristics of the providers that respondents consider important when making their decisions on where to receive healthcare, in one specification the reputation of the doctor has a strong positive effect, while both clinical and financial performance rates of the providers show negative significant effects: thus, the respondents who consider the reputation of the doctor important in deciding where to receive care are more likely to be aware of the rating websites, while this is less often the case for respondents putting a higher weight on financial or clinical performance ratings;
v) concerning the sources of information, in one specification respondents who consider the hospital statistics important in deciding where to receive care were more likely to be aware of the rating websites, with a quite strong and significant effect;
vi) among the variables on doctor-patient relationship, whenever included in the regressors, the gender match between the GP and the patient predicts higher awareness of the website ratings, with a noticeable effect;
vii) finally, although in one specification the respondents who feel that their GPs spend a sufficient

time in their consultation are less likely to be aware of the internet rating websites, both the statistical significance and the estimated odds ratio do not appear robust across specifications.

The text in the article has been amended accordingly. Please see new Results and Discussion sections on pages 8-11 and 11-15.

R2.12
Table 3 is not appropriate or necessary for the general medical reader, except perhaps as an appendix.

Reply:

Following the referee's suggestion, we have now greatly simplified the presentation of the estimation results in Table 4 (previously labelled as Table 3), and included the new Table in the Appendix. As suggested also by the first referee, we have now selected only six specifications of the empirical model that are likely to be the most relevant ones, and presented them in a step-wise format. The new Table 4 (previously labelled as Table 3), in the Appendix reports estimation results in terms of the odds ratio, together with the standard error, and levels of significance. The presentation of the results, as well as their discussion, in the corresponding sections, have also been modified accordingly. Please see new Results and Discussion sections on pages 8-11 and 11-15.

R2.13
The discussion section does not cite the prior literature that has reported on many of these phenomena in the past (e.g., use of the internet by age, gender, etc, use of health care data by consumers (pew, Kaiser family foundation) and does not cite literature on the content and scope of physician rating websites in England and the US (authors: Wachter, Lagu, Lopez, Jha) (at least one of these includes NHS choices). These should be included as background at least.

Reply:

We are very grateful to the referee for this helpful comment and for having pointed out to us this relevant stream of the literature. We have now enriched the new section with the discussion of the results by including evidence from this interesting literature.

More specifically, in the introduction and the discussion sections of the revised version of the article, we have compared the slow uptake of online ratings in our study with evidence from the US, which substantially confirms the relatively limited uptake of these websites even in a more market-oriented system. This evidence mainly comes from the studies by Lagu, Hannon, Rothberg et al. (2010); Gao et al. (2012) mentioned by the referee. In the results section, we have also related our finding of a relatively limited awareness of the rating websites among respondents in our sample to the comparable figures documented by the study by the Kaiser family Foundation, also suggested by the referee.

In the revised manuscript, moreover, we have expanded the discussion on the interaction between Internet information and the doctor-patient relationship, which is one of the main focuses of the study. We refer our findings to the relevant evidence on both the UK and the US provided by the works by Greaves et al. (2012a,b); and by Malone et al. (2005), McCartney (2009), Lagu and Lindenauer (2010), Lagu, Hannon, Rothberg et al. (2010); Gao et al. (2012), López et al., 2012; Wachter (2012), respectively. We feel that this important stream of literature, in particular, has allowed a much better

understanding and a more thorough interpretation of some of our findings.

In addition, in the results section of the revised document we provide further explanations on the impact of socio-economic variables on the likelihood of using online ratings, in light of the findings by Brodie et al. (2000) and Ybarra and Suman (2006). Finally, the final discussion and conclusions have also being expanded to show how the contribution of our study relates to the broader evidence on the topic, and in particular to the existing literature on the interaction between Internet information resources and doctor-patient relationship (e.g. (Nwosu and Cox, 2000; Broom, 2005; Malone et al, 2005; Gorrindo, 2008; Lagu and Lindenauer, 2010; López et al., 2012).

Please see the new Introduction (pages 4-5), Results (pages 8-11), Discussion (pages 11-15) and Conclusion (pages 15-16) sections.

R2.14
Furthermore, the discussion does not seem to center on the most important question for clinicians: what is the significance of the lack of patient awareness? Are the 29 patients who used the data actually using it? Or are they just aware of it? Do these sites contribute towards improving care quality or do they just "add to the noise?"

Reply:

We gratefully thank the referee for this very important comment. Following the referee's suggestion, we have now attempted to explicitly discuss the significance of our findings on the lack of awareness, in the final part of the new section containing the discussion of all the results. Please see pages 14-15.

VERSION 2 – REVIEW

REVIEWER	Simone Ghislandi Assistant Professor of Health Economics Department of Policy Analysis Bocconi University, 20136 Milan, Italy I have no conflict of interest with the paper reviewed.
REVIEW RETURNED	29-Aug-2012

GENERAL COMMENTS	The paper is well-written and ready for publication. Just a few questions more that I believe should be addressed: • You do not find any effect of education age and gender on the likelihood of their intention to use. This is strange. How would you comment on this? Is it a sign that "intention" does not imply actual usage or it is a finding that actually we should expect this? • Answer to your questionnaire can depend on the subjects having experienced the HC machine sometime in their life. Experience can even be indirect (through family). You do have a question in your questionnaire about this. And, I believe, this variable is the HC_PastExp in Table 3. However, OR are close to 1 and not significant. I would not expect something like this. Could you please comment on the role of experience in this type of studies and in your results in particular? • Websites uptake seems to be low. Do authors think this is due to the early stage of the "choice" model? Or they believe there is only a limited potential use of these websites also for the future? I think this issue should come out more clearly from the paper.

REVIEWER	Tara Lagu
	Baystate Medical Center
	USA
REVIEW RETURNED	27-Aug-2012

THE STUDY	This is a revised manuscript that describes a primary data collection that aims to determine who uses online healthcare reviews.
	The survey clearly represents quite a bit of work and appears to be well-constructed, but the format of the paper is still somewhat atypical for a medical manuscript. Most of my concerns are not methodological. They are more about writing style and choices about manuscript organization, although I am somewhat concerned about the complicated analysis the authors conducted given the extremely small sample of patients who reported they were aware of the sites.
	Introduction:
	The introduction is still long. It lists the study aims before providing sufficient background to the reader and also introduces some topics (e.g., Choose and book) which may not be directly relevant to the manuscript. Including information from the paragraphs below (but repackaging in two succinct and well-written paragaphs) is all that is really necessary:
	"Both the NHS Plan1 and the NHS Improvement Plan2, in fact, set out the changes required for the NHS to become more patient-focussed. Greater patient involvement in the running of the NHS has gone hand in hand with the policymakers' drive to improve the quality of public healthcare services. The 'bottom-up' approach to shape a more patient-centred NHS has typically focused on three main areas: i) giving users more choice and personalisation; ii) making funding respond to users' choices; and iii) engaging users through greater involvement. This was with the aim of creating a patient-led service promptly responding and supporting patients' health needs.3
	Lord Darzi's 2008 report "High Quality Care For All - The Next Stage Review"4 acknowledged that improvements to the NHS should focus on improving the quality of services, and that the best way of achieving this would be to ensure that services are locally responsive to the needs of the community. This would involve empowering providers and patients as decentralised decision-makers in order to foster a culture of continuous quality improvement and innovation.
	In practice, however, relative little evidence is available on whether, and to what extent, doctor-rating websites are actually known and actively used in the UK. A study by the Kings Fund6 explored the information sources used by patients in making decisions about where to receive care. Only 4% of the patients used the NHS Choices website, with the majority instead drawing information from their own experiences (41%), advice from GP (36%), advice from friends and family (18%), and other websites (1%). Similarly, a national survey on patients' choice by the Department of Health found that the NHS Choices website was only used by 5% of

respondents.7 These figures are consistent with the evidence from the US where usage of doctor rating websites is relatively more diffused and established.8,9

Moreover, very little is known about the profile of individuals who are more likely to make active use of these sites. Appleby and Alvarez10 found that women in England desire patient choice more than men (69% to 56%), suggesting that women may also be more likely to use patient sources of information such as rating sites. This is in line with findings from the US where women and younger adults are more active 'online health information seekers'.11"

These paragraphs, repackaged, folllowed by the aims is probably enough information for the background section. All the rest is probably not necessary.

Methods:

This paragraph should go into the discussion section:

The field survey was considered the most appropriate administration mode to involve a sample of respondents from the general population. An online survey, in fact, by exclusively reaching the segment of active internet users, would have failed to address the main goal of the study, whether the users of doctor-rating websites are fairly representative of the general public.

The authors do not use standard methodologic language when describing survey design and pilot testing. I would recommend that they examine other studies that have done something similar and mimic their style.

Where did the authors get their sample size of 200?

| **RESULTS & CONCLUSIONS** | Results:
The authors have include a lot of commentary in the results section which is not typical for a medical manuscript. Generally, medical journals do not include any commentary in the results section. For example,

This paragraph should go into the discussion section:
As common in field surveys of this type, the convenience sampling tended to over-represent respondents who were currently not working, or were at home, and thus had time to fill out the questionnaire: the proportion of subjects who were not currently working, as given by the sum of the respondents who reported to be unemployed, retired, or students, indeed amounts to 29% of the sample.

This paragraph should go into the discussion section

Comparing the sample with the Census data for the borough the mean age of our sample was slightly older than that for the borough (39.57 years compared to 35.2 years).15 Our sample however was closer to the national mean age of 38.5 years. The range of ages seems to show a positive skew, with a greater frequency of people aged 40 years and under. This is consistent with the 2001 census data for Hammersmith and Fulham which showed the borough contained a larger proportion of young people aged 20-29 (23.8%) than the rest of England (12.66%) (ONS, 2001).15 Age is an |

important demographic to consider when analysing our results as age has been shown to be important in internet usage.11

This should go in the discussion section
This corresponds to less than 15% of our sample, indicating that the awareness and, consequently, usage of these online sources is still quite limited in the UK, although significantly higher than what the previous studies have shown.6

As should this

A slow uptake of online ratings has also been reported in the US, a more market-oriented health system. It is indicative that only 6% of Americans were aware of Hospital Compare, the quality reporting website maintained by the Centres for Medicare and Medicaid Services (CMS).17

And this

Looking at the characteristics of the providers that respondents consider important in making their decisions on where to receive healthcare, in one specification the reputation of the doctor has a strong positive effect, while both clinical and financial performance rates of the providers show negative significant effects. Thus, the respondents who consider the reputation of the doctor important in deciding where to receive care are more likely to be aware of the rating websites, while this is less often the case for respondents putting a higher weight on financial or clinical performance ratings, perhaps signalling that those respondents may be more familiar with alternative sources of information.

"Concerning the sources of information, in one specification respondents who consider the hospital statistics important in deciding where to receive care, turn out to be more likely of being aware of the rating websites, with an effect which is particularly significant and quite remarkable in terms of odds ratio. This may signal the possible existence of 'complementary' effects between the two sources of information, according to which individuals who give importance to hospital statistics are also more likely to actively seek for doctor rating websites."

Again, I am not convinced the intensive statistical analysis adds a lot to the manuscript. This is a simple study-why the complicated analysis? What does it add?

Discussion:
Given my feedback above, I again suggest the authors rethink the discussion section. Is the primary goal of the paper to show that people are mostly unaware of these sites?

I know the relationship between use of the sites and characteristics of users is an additional point the authors want to convey, but given the convenience sample/sampling bias and low reported use, do we believe these results? I'm not sure what to make of them-and think that the authors are drawing a lot of conclusions from the 29 people who said they knew of the rating sites they were questioned about.

I would recommend that the authors again rethink the paper, simplify the analysis, and chose a couple of important conclusions and put

	those into the discussion section rather than focusing on so many different areas and not really covering any of them well.

VERSION 2 – AUTHOR RESPONSE

REVIEWER TARA LAGU

COMMENT 1: Introduction: The introduction is still long. It lists the study aims before providing sufficient background to the reader and also introduces some topics (e.g., Choose and book) which may not be directly relevant to the manuscript. Including information from the paragraphs below (but repackaging in two succinct and well-written paragraphs) is all that is really necessary:
"Both the NHS Plan1 and the NHS Improvement Plan2, in fact, set out the changes required for the NHS to become more patient-focussed. Greater patient involvement in the running of the NHS has gone hand in hand with the policymakers' drive to improve the quality of public healthcare services. The 'bottom-up' approach to shape a more patient-centred NHS has typically focused on three main areas: i) giving users more choice and personalisation; ii) making funding respond to users' choices; and iii) engaging users through greater involvement. This was with the aim of creating a patient-led service promptly responding and supporting patients' health needs.3
Lord Darzi's 2008 report "High Quality Care For All - The Next Stage Review"4 acknowledged that improvements to the NHS should focus on improving the quality of services, and that the best way of achieving this would be to ensure that services are locally responsive to the needs of the community. This would involve empowering providers and patients as decentralised decision-makers in order to foster a culture of continuous quality improvement and innovation.
In practice, however, relative little evidence is available on whether, and to what extent, doctor-rating websites are actually known and actively used in the UK. A study by the Kings Fund6 explored the information sources used by patients in making decisions about where to receive care. Only 4% of the patients used the NHS Choices website, with the majority instead drawing information from their own experiences (41%), advice from GP (36%), advice from friends and family (18%), and other websites (1%). Similarly, a national survey on patients' choice by the Department of Health found that the NHS Choices website was only used by 5% of respondents.7 These figures are consistent with the evidence from the US where usage of doctor rating websites is relatively more diffused and established.8,9
Moreover, very little is known about the profile of individuals who are more likely to make active use of these sites. Appleby and Alvarez10 found that women in England desire patient choice more than men (69% to 56%), suggesting that women may also be more likely to use patient sources of information such as rating sites. This is in line with findings from the US where women and younger adults are more active 'online health information seekers'.11"
These paragraphs, repackaged, followed by the aims is probably enough information for the background section. All the rest is probably not necessary.

REPLY 1
We thank the reviewer for this very helpful comment. We have now substantially reduced and reshaped the introduction as suggested by the reviewer. We make no longer reference to Choose and Book and we focus only on NHS Choices and Dr Foster Intelligence which are the key doctor-rating websites we asked our sample about. We feel that the introduction reads now better. Please see pages 4-5 of the revised document.

COMMENT 2: Methods: This paragraph should go into the discussion section: The field survey was considered the most appropriate administration mode to involve a sample of respondents from the general population. An online survey, in fact, by exclusively reaching the segment of active internet users, would have failed to address the main goal of the study, whether the users of doctor-rating websites are fairly representative of the general public.

REPLY 2
Following the suggestion by the reviewer, the paragraph has now been moved into the discussion section. In particular it can be found in the "Limitations of the study" subsection on pages 15-16 of the revised document.

COMMENT 3: The authors do not use standard methodologic language when describing survey

design and pilot testing. I would recommend that they examine other studies that have done something similar and mimic their style.

REPLY 3
Following the reviewer's comment, we looked carefully at the guidelines for articles to be submitted to the BMJ (http://www.bmj.com/about-bmj/resources-authors/article-submission/article-requirements), and, unfortunately, we were not able to find specific information on how to report on the survey design and pilot testing. We have also examined several other studies published on the BMJ, as suggested by the reviewer.
For instance the following two papers have been closely read: 1) Justin Berk, and Achyuta Adhvaryu. The impact of a novel franchise clinic network on access to medicines and vaccinations in Kenya: a cross-sectional study. BMJ Open 2012 2:e000589; doi:10.1136/bmjopen-2011-000589; 2) Carstensen, J et al. How does comorbidity influence healthcare costs? A population-based cross-sectional study of depression, back pain and osteoarthritis. BMJ Open 2012 2:e000809; doi:10.1136/bmjopen-2011-000809.
These two studies have similar characteristics to ours, and also seem to use a very similar language to describe the methods. In light of these studies that we have been able to find and examine, we are unclear on which direction to take regarding this specific comment, and therefore we have not been able to significantly alter the text to specifically address this point.

COMMENT 4: Where did the authors get their sample size of 200?
REPLY 4
We thank the reviewer for this important question that allows us to clarify the sample size calculation. The sample size was originally calculated having in mind the envisaged analysis to be conducted using the collected survey data. The analysis was intended to develop both bivariate correlation analysis and statistical regression analysis. Since the interpretation of the estimated coefficients in the statistical regression analysis can be readily re-conducted to the corresponding correlation coefficients, we conducted initial sample size calculations in terms of the former.

One of the goals of the study was to explore the likely determinants of the willingness to use doctor-rating websites. This would typically imply to look at the correlation coefficients between the likelihood to use and a broad range of behavioural and socio-demographic variables. In our survey the likelihood to use was modeled as a discrete ordered variable taking values 1, 2, and 3 for subjects reporting to be 'not likely', 'quite likely', and 'likely' to use the websites in the future, respectively. As explained in detail in Table 1, most variables in the survey were modeled as discrete variables, taking values between 1 and 5. Preliminary analysis of the data from the pilot survey suggested that most responses to the survey questions were distributed according to normal distributions around mean values in the range of 2.8-3.9 and standard deviations of about 1.2. Similar distributions were ex-post confirmed by the figures obtained by the data collected in the main survey, as summarised in Table 1.

Our data analysis was expected to look at the correlation coefficient between the likelihood of using the websites on the one hand, and a typical survey response, on the other, a correlation coefficient that here we can indicate as rws. The minimum sample size to test the null hypothesis of no significant correlation between the two variables (that is rws=0) can be calculated given an a priori belief on its effect size in the population. Published correlations between likelihood to use and behavioural responses, however, were unfortunately unavailable, and we therefore had to rely on hints from the pilot and on our most conservative estimates.

Assuming, conservatively, that the population correlation coefficient is rws=0.2 (a "low" effect size, since the variance of s would account for just 4% of the variance of w), and under the assumptions that both w and s are normally distributed a bi-directional test (both positive and negative correlation were expected) with 95% significance level reaches a standard 80% power level at a minimum sample of n=200 subjects (Cohen, 1969, page 90). We thus targeted a sample size of 200 respondents.

Following the reviewer's suggestion we have now briefly reported the above sample calculation at the end of the 'Sample' section on page 6.

COMMENT 5 Results: The authors have included a lot of commentary in the results section which is

not typical for a medical manuscript. Generally, medical journals do not include any commentary in the results section. For example,

This paragraph should go into the discussion section:

As common in field surveys of this type, the convenience sampling tended to over-represent respondents who were currently not working, or were at home, and thus had time to fill out the questionnaire: the proportion of subjects who were not currently working, as given by the sum of the respondents who reported to be unemployed, retired, or students, indeed amounts to 29% of the sample.

This paragraph should go into the discussion section: Comparing the sample with the Census data for the borough the mean age of our sample was slightly older than that for the borough (39.57 years compared to 35.2 years).15 Our sample however was closer to the national mean age of 38.5 years. The range of ages seems to show a positive skew, with a greater frequency of people aged 40 years and under. This is consistent with the 2001 census data for Hammersmith and Fulham which showed the borough contained a larger proportion of young people aged 20-29 (23.8%) than the rest of England (12.66%) (ONS, 2001).15 Age is an important demographic to consider when analysing our results as age has been shown to be important in internet usage.11

REPLY 5
We thank the reviewer for these useful suggestions. These paragraphs have now been merged with the new discussion section. In particular we have created a new section in the discussion where we discuss the representativeness of our sample by comparing its characteristics to the statistics from the Borough and the National Census. Please see pages 11-12 of the revised document.

COMMENT 6 This should go in the discussion section: This corresponds to less than 15% of our sample, indicating that the awareness and, consequently, usage of these online sources is still quite limited in the UK, although significantly higher than what the previous studies have shown.6
As should this: A slow uptake of online ratings has also been reported in the US, a more market-oriented health system. It is indicative that only 6% of Americans were aware of Hospital Compare, the quality reporting website maintained by the Centres for Medicare and Medicaid Services (CMS).17

REPLY 6: Following the reviewer's suggestion, these two paragraphs have been merged with the Awareness subsection of our Discussion. Please see page 12.

COMMENT 7: This should go in the discussion section: Looking at the characteristics of the providers that respondents consider important in making their decisions on where to receive healthcare, in one specification the reputation of the doctor has a strong positive effect, while both clinical and financial performance rates of the providers show negative significant effects. Thus, the respondents who consider the reputation of the doctor important in deciding where to receive care are more likely to be aware of the rating websites, while this is less often the case for respondents putting a higher weight on financial or clinical performance ratings, perhaps signalling that those respondents may be more familiar with alternative sources of information.
This should go in the discussion section: "Concerning the sources of information, in one specification respondents who consider the hospital statistics important in deciding where to receive care, turn out to be more likely of being aware of the rating websites, with an effect which is particularly significant and quite remarkable in terms of odds ratio. This may signal the possible existence of 'complementary' effects between the two sources of information, according to which individuals who give importance to hospital statistics are also more likely to actively seek for doctor rating websites."

REPLY 7: For the most, these paragraphs describe the results rather than commenting them. We have thus left in the Results section only the parts that refer to the results description. Following the reviewer's suggestion, the commentary sentences have been merged with the Discussion text. Please see pages 9 and 12-13.

COMMENT 8: Again, I am not convinced the intensive statistical analysis adds a lot to the manuscript. This is a simple study-why the complicated analysis? What does it add?

REPLY 8: We thank the reviewer for this helpful suggestion. We confirm that we have paid careful attention to select and run the simplest possible statistical models for the multiple regression analysis, choosing the binary and discrete variables models frequently employed in health economics research

but also in medical journals. Moreover, following the reviewers' suggestions, we have re-run the entire set of estimations using the logistic, rather than the probit, specification of the models, and changed the presentation of the results in terms of odds ratios (together with standard errors), rather than estimated coefficients, precisely in order to further facilitate the immediate interpretation of the results by the broadest possible audience. The employment of multiple regression analysis adds highly informative insights to the bivariate correlations analysis, as it allows to check the findings and to test the robustness of the statistical associations in a multivariate environment. As such, it allows to rigorously account for several issues that may occur in the underlying structure of the data and that are not otherwise detectable when just using bivariate correlations analysis. These include potential issues of omitted variables, multi-collinearity, and endogeneity, and we thus feel that retaining it in the article may be of some interest to the reader.

Yet, we see the reviewer's point for simplifying the text on the statistical analysis. In the section on 'Results' of the included revised draft, we have now added a paragraph at the beginning of each of the two sub-sections on the 'Results on awareness' and on 'Results on the likelihood to use online rating websites'. These two new paragraphs immediately highlight the main findings that directly emerge from the bivariate correlations analysis reported in Table 2, for the variables 'Awareness' and 'Intention to use', respectively. An overview of the results from the correlation analysis is thus presented to the reader before the presentation of the results from the more structured statistical analysis reported in Table 3 and Table 4, respectively. Please see pages 8 and 10 of the revised document.

COMMENT 9: Discussion: Given my feedback above, I again suggest the authors rethink the discussion section. Is the primary goal of the paper to show that people are mostly unaware of these sites? I know the relationship between use of the sites and characteristics of users is an additional point the authors want to convey, but given the convenience sample/sampling bias and low reported use, do we believe these results? I'm not sure what to make of them-and think that the authors are drawing a lot of conclusions from the 29 people who said they knew of the rating sites they were questioned about.
I would recommend that the authors again rethink the paper, simplify the analysis, and chose a couple of important conclusions and put those into the discussion section rather than focusing on so many different areas and not really covering any of them well.

REPLY 9
We are very grateful to the reviewer for these helpful suggestions. The relatively limited uptake and awareness of these websites, although not the primary goal of the paper, is an important point we would like to make, as it adds to the discussion on whether patients are well informed about the options they have, as also requested by the other reviewer. We fully agree with the reviewer, however, that the discussion in the article should better focus on a couple of main conclusions and develop them with greater emphasis. Following the reviewer's suggestion, we have now sharpened up the overall discussion and the text throughout the paper by focusing on what we see as the main message of the paper. In particular, our current discussion focus is on emphasising the importance of the doctor-patient relationship on the intention to use doctor-rating websites, and on discussing the substitute and complementary roles of the Internet and the physician. This is a new set of findings that the previous literature has not fully discussed. The results on awareness are only lightly cross referenced in the conclusion and discussion sections of the paper on pages 11-18.
The abstract has been changed to reflect these changes.

REVIEWER: Simone Ghislandi

COMMENT 1: You do not find any effect of education age and gender on the likelihood of their intention to use. This is strange. How would you comment on this? Is it a sign that "intention" does not imply actual usage or it is a finding that actually we should expect this?

REPLY 1
We agree with the reviewer that this is a striking finding, a priori not expected, and we are grateful to have addressed our attention to it. We believe that this may be due to the difference between the determinants of actual use and the intentions to use. In particular our findings, together with evidence from the literature, seem to suggest that even though people may declare an intention to use there

may exist barriers to access that are related to socio-economic and demographic characteristics that prevent them from doing so.

We have now introduced several comments related to this in our discussion section. See pages 13 and 16 of the revised document. We have also updated the text with new references that we have added in these paragraphs.

COMMENT 2: Answer to your questionnaire can depend on the subjects having experienced the HC machine sometime in their life. Experience can even be indirect (through family). You do have a question in your questionnaire about this. And, I believe, this variable is the HC_PastExp in Table 3. However, OR are close to 1 and not significant. I would not expect something like this. Could you please comment on the role of experience in this type of studies and in your results in particular?

REPLY 2
This is indeed an interesting comment and we thank the reviewer for pointing it to our attention. Let us first clarify that the HC_PastExp variable refers to the participant's personal past experience with the provider. So, it does not refer to the indirect experience, i.e. through the family. Yet, the feedback from family and friends is captured by another variable, SI_family, which was, only marginally, significantly correlated with the participant's intention to use the websites (see Table 2).

Going back to the personal past experience of the participant we have looked at all the relevant papers we quote in our study and we found that, interestingly enough, there is no reference to past experience. However, and even though the particular variable has not been significant, the relationship of the patient with the provider, i.e. the doctor, has received a lot of attention in our paper and indeed has been shown to play an important role in determining the patient's intention to use these websites. This aspect is acknowledged on pages 12-16 of our discussion section.

COMMENT 3: Websites uptake seems to be low. Do authors think this is due to the early stage of the "choice" model? Or they believe there is only a limited potential use of these websites also for the future? I think this issue should come out more clearly from the paper.

REPLY 3
We thank the reviewer for this important observation. As discussed in the paper, the limited evidence that exists on the topic in the UK showed a low uptake of these websites in the past (Dixon et al 2010). Our results show that although still low, the uptake of these websites has increased significantly. This may be the result of great choice, which makes people look for more information. What is interesting though, is that even in the US, a more market-oriented health system, the use of similar sites is not much higher. This makes us believe that the slow uptake in the UK cannot be attributed only to the early stage of the "choice" model but perhaps, it is due to the important role that more traditional information channels, such as the relation with the physician, still play. This is the main reason why we have emphasised the behavioural aspects of the doctor-patient relation, and have extensively discussed the 'substitute' and 'complementary' effect between the Internet and the physician as information providers. Following the reviewer's suggestions, we have now included an explicit discussion of this important point in the 'General discussion' sub-section.

VERSION 3 - REVIEW

REVIEWER	Tara Lagu Baystate medical center USA
REVIEW RETURNED	12-Sep-2012

THE STUDY	The manuscript is improved although still not in the classic format of a medical journal. The authors have removed a lot of the analysis from the results section, which is good, but some still remains, e.g., the paragraph that begins"It is interesting that....."

	I'm still not convinced their modeling strategy is necessary given their extremely small sample size. some justification for this would be helpful.
RESULTS & CONCLUSIONS	The discussion section is disjointed. It should be a fluent and coherent explanation of the results, starting with the main findings. What are the most important results? Summarize them in a few sentences, a single paragraph. Then discuss the implications in subsequent paragraphs. The discussion section is still too long. Why is it divided into "discussion" general discussion" and "conclusions?" This needs to be reformatted for a more medical style. Because of the length, the disjointedness and the organization of the discussion section, I am left unsure of what the author's key findings are. I am also unsure, looking at the "key findings" section on the first page if the authors actually sampled "the general public" or can use this phrase.

VERSION 3 – AUTHOR RESPONSE

Reply to the referee

Comment 1: "The manuscript is improved although still not in the classic format of a medical journal. The authors have removed a lot of the analysis from the results section, which is good, but some still remains, e.g., the paragraph that begins" It is interesting that....."

Reply 1: Following the reviewer's suggestion, we have now edited the sentences in the discussion, completely removing the comments and leaving only the results part.

Comment 2: I'm still not convinced their modelling strategy is necessary given their extremely small sample size. Some justification for this would be helpful.

Reply 2: We confirm that we have used the simplest possible modelling strategy as also stated in our previous replies. We have tried to address this issue in our previous revisions as suggested by both referees. In particular, we have introduced the correlation analysis as suggested by the reviewer. We have chosen the simplest possible multivariate models for discrete dependent variables to complement the bivariate correlation analysis. As also suggested by the other referee, we have already changed the modelling strategy and used standard logistic models, presenting their results in terms of odds ratios, as commonly done in similar studies. We have also paid extra attention in explaining and justifying all these modelling choices in the text and in simplifying all over the jargon in order to render the text easily readable to all audiences as suggested by the referees. At this stage, therefore, we are unsure what more can be possibly done in order to further justify the analysis.

Comment 3: The discussion section is disjointed. It should be a fluent and coherent explanation of the results, starting with the main findings.

Reply 3: We thank the reviewer for this suggestion. We agree that the discussion section should start with the main findings, as suggested by the referee.
Therefore, we have : i) moved the sample description to the results section, merging it with the descriptive statistics section to avoid duplication of text; ii) added the comments regarding the

convenience sampling into the 'Limitations' section.

Comment 4: What are the most important results? Summarize them in a few sentences, a single paragraph. Then discuss the implications in subsequent paragraphs. The discussion section is still too long. Why is it divided into "discussion" general discussion" and "conclusions?" This needs to be reformatted for a more medical style. Because of the length, the disjointedness and the organization of the discussion section, I am left unsure of what the author's key findings are.

Reply 4: We are very grateful to the reviewer for these helpful suggestions. We have now introduced a paragraph at the beginning of the discussion section summarizing the main results as suggested. The discussion after that paragraph now expands each of those results and relates the findings to the literature.
Moreover, as suggested, we have substantially reduced the length of the discussion. We have also incorporated the text in the previous "General Discussion" sub-section into the "Intention to use" and "Awareness and usage" sub-sections.
The section "Limitations of the study" has also been polished. Finally, we have reduced the conclusions to avoid text repetition with the discussion section.

Comment 5: I am also unsure, looking at the "key findings" section on the first page if the authors actually sampled "the general public" or can use this phrase.

Reply 5: We have now edited the key findings as well as the whole manuscript, by removing the term "general public", as suggested.