PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The routine collection of patient-reported outcome measures (PROMs) for long-term conditions in primary care: a cohort survey
AUTHORS	Peters, Michele; Crocker, Helen; Jenkinson, Crispin; Doll, Helen; Fitzpatrick, Ray

VERSION 1 - REVIEW

REVIEWER	Chris Salisbury and Mairead Murphy Centre for Academic Primary Care University of Bristol
REVIEW RETURNED	11-Oct-2013

GENERAL COMMENTS

High priority revisions which we recommend should be made are listed below. A number of other revisions, which may also be desirable, are detailed in the attached word document.

1. Clarify the research question or questions and underlying hypothesis (see detailed comments)
2. Clarify the outcome measures (see detailed comments)
3. Present full details of the logistic regression to take account of confounding
4. Add totals to the bottom of all tables
5. Correct numerical inaccuracies (see detailed comments)
6. Include a flow diagram, commencing with the 5596 patient originally extracted and ending with the 1136 final responders, which reconciles to the table totals throughout.
7. Review the abstract, discussion and conclusions, and ensure key messages in the abstract and conclusions drawn are supported by the data.
8. Review the additional desirable recommendations shown in the commentary, and consider each in turn.

Note to authors: the review template allows only for “Yes” or “no” responses to the review checklist. In many cases, we would have preferred to indicate “to some extent” or “to a large extent”. However, as this was not allowable, we indicated “no” on questions where we believe revisions to the manuscript are recommended as a priority.

1. Is the research question or study objective clearly defined?

No

The research question is clearly stated on the first line of the abstract, under Objective: “To evaluate the feasibility of using PROMs for LTCs in primary care.” In fact, based on the outcome measures and analysis presented, it seems that the authors are addressing 2 research questions. If this is the case, we would suggest that these should be stated as two separate but related questions, i.e.: 1) to assess the feasibility of collecting longitudinal PROM data for LTCs in primary care and 2) to assess the appropriateness of the EQ-5D for this purpose.

We also found ourselves wondering what the implied hypothesis was for this study. Did the authors expect the EQ5D to show any change? We would not anticipate much change in this measure in patients with long term conditions who are not receiving any specific intervention. So in a sense this study is effectively a study of the test-retest reliability of the EQ5D, although that doesn't appear to be the intention.

Abstract

2. Is the abstract accurate, balanced and complete?

No

The abstract would benefit from greater clarity in the research question (specifically whether there is 1 only, or a primary and a secondary) and in the outcome measures. The conclusions described in the abstract aren't completely representative of the results section; the latter focussing on response rates, and the former only touching on this. This is described below under "Discussion and Conclusions"

Study Design

3. Is the study design appropriate to answer the research question?

Yes

Description of Methods

4. Are the methods described sufficiently to allow the study to be completed?

Yes

Yes. However it would have been useful to have more information about the guidance that the researchers the practices about the grounds to exclude patients, because in our experience the level of exclusions relates strongly to what the practices are asked to do. Also, in the method, it wasn't clear why different numbers of practices provided patients for each LTC. (Presumably this was to get 175 for each LTC, taking into account the varied prevalence of these conditions, but it is not clear.)

Research Ethics

5. Are research ethics addressed appropriately?

Yes.

Outcome Measure(s):

6. Are the outcome measures clearly defined?

No

The outcome measures are specified in the abstract, but could be more clearly defined, in particular the second one. They would be clearer if split out into 2 separate lines (to reflect the two research questions). The first outcome measure is response rates (more specifically 2 rates: baseline response and follow-up response). The second outcome measure currently reads “and the EQ-5D, a generic patient reported outcome measure.” The following should be clarified:

1. What is the outcome? Is the outcome the absolute value? Or change in EQ-5D index score / VAS score?
2. What hypothesis, or research question, does this outcome measure test? Is there an implied hypothesis that the EQ-5D should show some change (why should that be so?) , and if it fails to, that this would demonstrate its lack of usefulness as a PROM for LTCs? (If so, this should be explained and justified.)
3. The study uses the EQ-5D-3L, as opposed to the EQ-5D-5L. Recent studies of EQ-5D done on LTCs show 5L has more discriminatory power. 5L does not, however, have a directly elicited value set available, only a mapping, so it is justifiable to use 3L on this basis. The outcome measure section might benefit from brief explanation on the choice of 3L over 5L.

7. Statistics:

We were confused by the power calculation, which relates to our earlier comments about what hypothesis the study was designed to test. If the primary aim of the study is (as stated) to test the feasibility of using PROMS for LTCs, is it necessary to power on effect size? Why would a 0.3 effect size be expected in a general population of people with LTCs receiving standard primary care interventions over this time?

8. References:

Are the references up to date and appropriate?

Yes

One exception may be literature on the non-responsiveness of the EQ-5D for people with LTCs, as this is not a new finding, although it is slightly presented as such. Some studies, including the Health Survey for England have already shown little change in the EQ-5D for people with LTCs, so this should be referred to.
(HSE data referred to in DoH Policy Paper: **Setting Levels of Ambition for the NHS Outcomes Framework,** *A technical annex to support Delivering our NHS care objectives: a consultation on the draft mandate to the NHS Commissioning Board Chapter 4: Enhancing quality of life for people with long-term conditions, published July 2012)*

9. ***Do the results presented address the research question or objective?***

Not very clearly

<u>Results Presented (First Outcome Measure)</u>

The results presented on the first outcome measure (response rates for baseline and follow-up surveys) do address the primary research question. Given that the primary research question is to assess the feasibility of using PROMs for LTCs in primary care it is this outcome measure which is of most interest. There are some areas which could do with more detail:

1. **Baseline Response Rate 38.4%**: This is where the highest proportion of responders are lost, but it has less detail than the loss to follow-up (e.g. breakdown by age). This may be because this information was only captured at the stage of questionnaire completion, and is therefore not available. If the authors did not collect data on the age and sex of all those asked to complete the questionnaire, that would be a shame and they should state this as a limitation. It would have been quite possible to do that. However, if they did collect these data it would be worthwhile presenting it so that we can understand the response bias. Even if they do not have age data, they might still discuss that, for example, that the low response rate in asthma may be confounded by age. The logistic regression results would also be useful to see.
2. **Initial Cohort extracted (n = 5596)**: Are there any possible reasons why the total patients extracted might be different apart from asthma and stroke, which have different criteria? Epilepsy and COPD show quite substantial differences – with at least one practice identifying more than 3 times as many patients in the epilepsy cohort. It would be useful to highlight potential reasons for such a difference.
3. **Patients excluded by practices (n = 1111/1112)**: The

variation among practices in number of exclusions is interesting and surprising. Is there any more information available on why there is so much variation? Was the list of exclusions used given to practices in advance? (i.e. death, blindness, co-morbidities etc). Did they capture exclusion by reason?

Results Presented (Second Outcome Measure)

The results presented for the second outcome measure are all in Table 8. The conclusions which should be drawn from this are not fully clear (see comments on clarity of second outcome measure). Additionally, the results of the logistic regression are summarised, and somewhat lost in the text. It would be useful to present these results.

10. Are the results presented clearly

No

Most of the results are presented in terms of ANOVAs, simple comparisons without adjustment for other confounding variables. The more appropriate comparison would be to show the unadjusted and adjusted relationships between the explanatory variables and the outcome variables using logistic regression, to take account of confounding. Although the authors state in the methods that they did do logistic regression, they only present the results as a summary sentence. But we would suggest that these are the most important results, and should be presented in detail.

There are a number of numerical inaccuracies throughout. Some of these are listed below. Adding totals to the bottom of all tables, and providing a flow diagram would make the correction and the checking of all figures easier for both the authors and the readers.

1. Table 2 shows an incorrect sign under % difference for epilepsy: should be +32.4%
2. Table 3 would be easier to read with totals at the bottom, and the absolute total of patients included as well as the %, so that it can be tied back to the 4484 mentioned in the abstract. Also, a quick calculation shows it to be 4485 when these are added, not 4484. What happened to the missing patient?
3. Table 4 would be easier to read with total at the bottom. Also as with Table 3, the totals do not match those given in the text: 4485 again for the total patients written to, at 1716 responders, as opposed to 1721 described in text. Some of the percentages are also not quite right: e.g. top line: 395/1334 =29.6%, not 30.0%.
4. Table 5 N = 97 as stated in text and as calculated by total. But if N = no. practices for each LTC covered, it should be 98 (10 +16 +10 +23 +20 +19)

5. Table 6 total patients contact is 1589 (as opposed to abstract which is 1590)

Other tables may also need to be checked. A flow chart would be very helpful to readers in following the analysis, starting with the 5596 patients extracted (a total which is not presented anywhere), and illustrating exclusions and non-responders through to the final 1136.

11. Are the discussion and conclusion justified by the results?

No

The conclusion and discussion is only partly justified by the results presented, and the section would benefit from some redrafting to link the two more clearly:

1. The research question was about the feasibility of using PROMs for LTCs in primary care. It is difficult to unpick from the discussion what the answer to this research question is. This will also be clearer once the research question and outcome measures are more clearly specified. It would be useful to have the key messages, or contributions to knowledge highlighted at the start of the discussion, and tied back to the research question(s).
2. It states for the first time in the discussion “response rates were the main variable of interest in this study”. This should be reflected in the abstract, results and conclusion, and in the Key Messages section at the front which should focus on response rates, the conclusions that can be drawn from them, and feasibility of collecting PROM data. Instead this focuses on the appropriateness of EQ-5D and the (speculative) point that response rates could be improved by making PROMs more engaging for patients.
3. The last paragraph of the discussion describes 3 ways in which baseline response rates could be improved. These should be either based on the research in this paper or referenced to other papers, as they currently read as somewhat speculative.

12. Are the study limitations discussed adequately

Yes

Limitations are discussed adequately. However, the limitation described as “the logistics of remotely identifying eligible patients from GP databases” is worthy of a few sentences explanation. GP systems should readily allow such identification, since these conditions are clearly specified in the QOF and data about them is already collected remotely in order to calculate QOF payments for each practice. So it would be helpful to clarify whether the authors experienced technical problems, specific to this study, or whether there are any fundamental problems likely to impede routine collection of PROMs.

13. Supplementary reporting Complete?

yes

14. To the best of your knowledge, free from concerns over publication ethics?

Yes

15. Is the standard of written English acceptable for publication

The standard of written English is good, but needs a review for accuracy of language used throughout – e.g. (pg. 8 “There were some disease-specific differences in consent, including gender….”)

REVIEWER	Joanne Greenhalgh School of Sociology and Social Policy University of Leeds UK No competing interests - though I do know two of the authors of this manuscript
REVIEW RETURNED	04-Dec-2013

GENERAL COMMENTS

This is an important paper and addresses a highly topical issue.

My main issue lies in the mismatch between the stated objectives of the study and the data presented. The stated objectives of the study are to evaluate the feasibility of using PROMs to monitor the quality of life of individuals with LTC in primary care. I think the study evaluates some important aspects of the feasibility of collecting PROMs data in primary care but not does not fully evaluate the feasibility of using it. The main focus of the paper lies in describing the logistical difficulties in identifying patients with LTC, the reasons why practices excluded patients, the response rates obtained and variation in response rates by patient characteristics and place. These are extremely important and the paper merits publication on this basis alone - it is important to know about the feasibility of collecting the data. I would therefore recommend that the stated objectives are modified to reflect the focus on feasibility of collection, rather than use.

Furthermore, the focus on using PROMs to monitor the quality of life of individuals suggests that the data are intended to be used at the individual patient level to inform the planning of their care by individual GPs, whereas the previous sentences in the last para of the introduction highlight the challenges of using PROMs as an

	indicator of the effectiveness of primary care - ie as an aggregate indicator of the quality of care. It would be helpful for the authors to make explicit the different intended uses of PROMs data and the lack of clarity around this in primary care (and indeed the tensions that surround it), as this has important implications for the discussion of their findings. As a reader I really wanted to know more about the qualitative aspects of feasibility, such as more details about why practices excluded individuals, how the whole enterprise was perceived by practices and by patients and what value they thought that data might have. It would be useful to know if the team collected this data and if so whether it will be published elsewhere. It would be useful to know if they collected any data to shed more light on what lay behind the regional variations in response rates. The paper also reports the change in the EQ5D over time and notes the lack of change over time. Here the paper does begin to consider some aspects of the validity of using PROMs data as an indicator of the quality of primary care. The paper would benefit from a deeper critical and perhaps more speculative discussion of the implications of the low response rates and the lack of change in EQ5D scores for PROMs as an indicator of the quality of primary care. For example, is it going to be used to as a way of benchmarking the quality of care provided by GP practices? We know from the use of PROMs in elective surgery that low response rates increase the risk that providers may be mislassified as an outlier (see work of Andrew Street and Matt Sutton) so this may be even more of an issue in primary care given that response rates are much lower. Furthermore, the lack of change in EQ5D scores over time raises serious questions about whether change in scores (or even absolute single point in time scores) would be useful in disriminating between practices. While the paper does not set out to evaluate whether the EQ5D scores could discriminate between practices, but it might be useful to raise this in the discussion - or at least raise the issue that the lack of clarity about how PROMs data will be used hinders our understanding of how the low response rate and low change over time impact on the validity of the indicators. There are also broader questions to consider about what PROMs data is intended to reflect the quality of - since people with LTC are not only cared for in primary care but also by a range of other services spanning secondary care, social care, the voluntary sector and their own families. The paper does touch on these both in the introduction and in the discussion, but again, a deeper discussion would strengthen the paper. In particular, in the introduction it might be helpful to expand on exactly how PROMs data is expected to drive change in the organisation and delivery of health services.

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Chris Salisbury and Mairead Murphy

Institution and Country Centre for Academic Primary Care

University of Bristol

High priority revisions which we recommend should be made are listed below. A number of other revisions, which may also be desirable, are detailed in the attached word document.

Reviewers' comments	Authors' response
1. Clarify the research question or questions and underlying hypothesis (see detailed comments)	This has been clarified in the abstracts and methodology section.
2. Clarify the outcome measures (see detailed comments)	Clarified in abstract and methods.
3. Present full details of the logistic regression to take account of confounding	Table with results from logistic regression and associated text have been inserted into the results section.
4. Add totals to the bottom of all tables	Totals have been added to all tables.
5. Correct numerical inaccuracies (see detailed comments)	These have all been corrected.
6. Include a flow diagram, commencing with the 5596 patient originally extracted and ending with the 1136 final responders, which reconciles to the table totals throughout.	Flow diagram has been inserted.
7. Review the abstract, discussion and conclusions, and ensure key messages in the abstract and conclusions drawn are supported by the data.	We believe that the messages in the abstract are supported by the data. The messages should be clearer now that the study outcomes have been clarified. This reviewer ask for less speculation, but the second reviewer asked for more speculation. As authors, we would like to keep it as it is.
8. Review the additional desirable recommendations shown in the commentary, and consider each in turn.	See below

Other comments

1. Is the research question or study objective clearly defined?	
The research question is clearly stated on the first line of the abstract, under Objective: “To evaluate the feasibility of using PROMs for LTCs in primary care.” In fact, based on the outcome measures and analysis presented, it seems that the authors are addressing 2 research questions. If this is the case, we would suggest that these should be stated as two separate but related questions, i.e.: 1) to assess the feasibility of collecting longitudinal PROM data for LTCs in primary care and 2) to assess the appropriateness of the EQ-5D for this purpose.	This has been clarified.
We also found ourselves wondering what the implied hypothesis was for this study. Did the authors expect the EQ5D to show any change? We would not anticipate much change in this measure in patients with long term conditions who are not receiving any specific intervention. So in a sense this study is effectively a study of the test-retest reliability of the EQ5D, although that doesn’t appear to be the intention.	The rationale for assessing change using the EQ-5D has been included (start of methods section)
Abstract	
2. Is the abstract accurate, balanced and complete?	
The abstract would benefit from greater clarity in the research question (specifically whether there is 1 only, or a primary and a secondary) and in the outcome measures. The conclusions described in the abstract aren’t completely representative of the results section; the latter focussing on response rates, and the former only touching on this. This is described below under “Discussion and Conclusions”	The study outcomes have been clarified. Both the discussion and abstracts focus on both response rates and change in EQ-5D. The abstract has been modified to clarify where the EQ-5D data is referred to.
Description of Methods	
4. Are the methods described sufficiently to allow the study to be completed?	
Yes	
Yes. However it would have been useful to have more information about the guidance that the researchers the practices about the grounds to	Further information on patient exclusion has been inserted in the methods. The rationale for different number of practices

exclude patients, because in our experience the level of exclusions relates strongly to what the practices are asked to do. Also, in the method, it wasn’t clear why different numbers of practices provided patients for each LTC. (Presumably this was to get 175 for each LTC, taking into account the varied prevalence of these conditions, but it is not clear.)	has been inserted in the methods.
Outcome Measure(s):	
6. Are the outcome measures clearly defined?	
No	
The outcome measures are specified in the abstract, but could be more clearly defined, in particular the second one. They would be clearer if split out into 2 separate lines (to reflect the two research questions). The first outcome measure is response rates (more specifically 2 rates: baseline response and follow-up response). The second outcome measure currently reads “and the EQ-5D, a generic patient reported outcome measure.” The following should be clarified: 1. What is the outcome? Is the outcome the absolute value? Or change in EQ-5D index score / VAS score? 2. What hypothesis, or research question, does this outcome measure test? Is there an implied hypothesis that the EQ-5D should show some change (why should that be so?), and if it fails to, that this would demonstrate its lack of usefulness as a PROM for LTCs? (If so, this should be explained and justified.)	The outcomes have been clarified in the abstract and methods. Further clarification on the rationale for assessing change on the EQ-5D has been added.
3. The study uses the EQ-5D-3L, as opposed to the EQ-5D-5L. Recent studies of EQ-5D done on LTCs show 5L has more discriminatory power. 5L does not, however, have a directly elicited value set available, only a mapping, so it is justifiable to use 3L on this basis. The outcome measure section might benefit from brief explanation on the choice of 3L over 5L.	The study started before the EQ-5D-5L was available. The baseline data collection started in September 2010, and the article on the development and preliminary testing of the EQ-5D-5L was not published until 2011.
7. Statistics:	
We were confused by the power calculation,	This is explained by the rationale

which relates to our earlier comments about what hypothesis the study was designed to test. If the primary aim of the study is (as stated) to test the feasibility of using PROMS for LTCs, is it necessary to power on effect size? Why would a 0.3 effect size be expected in a general population of people with LTCs receiving standard primary care interventions over this time?	(beginning of methods section) on evaluating change by means of the EQ-5D.
8. References:	
Are the references up to date and appropriate?	
Yes One exception may be literature on the non-responsiveness of the EQ-5D for people with LTCs, as this is not a new finding, although it is slightly presented as such. Some studies, including the Health Survey for England have already shown little change in the EQ-5D for people with LTCs, so this should be referred to. (HSE data referred to in DoH Policy Paper: Setting Levels of Ambition for the NHS Outcomes Framework, A technical annex to support Delivering our NHS care objectives: a consultation on the draft mandate to the NHS Commissioning Board Chapter 4: Enhancing quality of life for people with long-term conditions, published July 2012)	This reference has been inserted into the discussion.
9. Do the results presented address the research question or objective?	
Not very clearly	

Results Presented (First Outcome Measure) The results presented on the first outcome measure (response rates for baseline and follow-up surveys) do address the primary research question. Given that the primary research question is to assess the feasibility of using PROMs for LTCs in primary care it is this outcome measure which is of most interest. There are some areas which could do with more detail: 1. Baseline Response Rate 38.4%: This is where the highest proportion of responders are lost, but it has less detail than the loss to follow-up (e.g. breakdown by age). This may be because this information was only captured at the stage of questionnaire completion, and is therefore not available. If the authors did not collect data on the age and sex of all those asked to complete the questionnaire, that would be a shame and they should state this as a limitation. It would have been quite possible to do that. However, if they did collect these data it would be worthwhile presenting it so that we can understand the response bias. Even if they do not have age data, they might still discuss that, for example, that the low response rate in asthma may be confounded by age. The logistic regression results would also be useful to see.	No data was collected on non-responders as to minimize burden on practice staff. This is now discussed in the limitations of the study. The full logistic regression results have been included in the revised version.
2. Initial Cohort extracted (n = 5596): Are there any possible reasons why the total patients extracted might be different apart from asthma and stroke, which have different criteria? Epilepsy and COPD show quite substantial differences – with at least one practice identifying more than 3 times as many patients in the epilepsy cohort. It would be useful to highlight potential reasons for such a difference.	It was participants with diabetes and stroke (not asthma) where the search criteria differed from QOF. The epilepsy search was problematic due to a mistake in the initial searches. The difference is COPD was likely to be related to problems searching some of the GP clinical systems. This has been explained in more details in the limitations.
3. Patients excluded by practices (n = 1111/1112): The variation among practices in number of exclusions is interesting and surprising.Is there any more information available on why there is so much variation? Was the list of exclusions used given to practices in advance? (i.e. death, blindness, co-morbidities etc). Did they capture exclusion by reason?	More explanation is given in the methods and limitations of the study.
Results Presented (Second Outcome Measure)	
The results presented for the second outcome measure are all in Table 8. The conclusions	The full logistic regression data has been

which should be drawn from this are not fully clear (see comments on clarity of second outcome measure). Additionally, the results of the logistic regression are summarised, and somewhat lost in the text. It would be useful to present these results.	inserted.
10. Are the results presented clearly	
No Most of the results are presented in terms of ANOVAs, simple comparisons without adjustment for other confounding variables. The more appropriate comparison would be to show the unadjusted and adjusted relationships between the explanatory variables and the outcome variables using logistic regression, to take account of confounding. Although the authors state in the methods that they did do logistic regression, they only present the results as a summary sentence. But we would suggest that these are the most important results, and should be presented in detail.	The logistic regression results are now fully presented.

There are a number of numerical inaccuracies throughout. Some of these are listed below. Adding totals to the bottom of all tables, and providing a flow diagram would make the correction and the checking of all figures easier for both the authors and the readers. 1. Table 2 shows an incorrect sign under % difference for epilepsy: should be +32.4% 2. Table 3 would be easier to read with totals at the bottom, and the absolute total of patients included as well as the %, so that it can be tied back to the 4484 mentioned in the abstract. Also, a quick calculation shows it to be 4485 when these are added, not 4484. What happened to the missing patient? 3. Table 4 would be easier to read with total at the bottom. Also as with Table 3, the totals do not match those given in the text: 4485 again for the total patients written to, at 1716 responders, as opposed to 1721 described in text. Some of the percentages are also not quite right: e.g. top line: 395/1334 =29.6%, not 30.0%. 4. Table 5 N = 97 as stated in text and as calculated by total. But if N = no. practices for each LTC covered, it should be 98 (10 +16 +10 +23 +20 +19) 5. Table 6 total patients contact is 1589 (as opposed to abstract which is 1590) Other tables may also need to be checked. A flow chart would be very helpful to readers in following the analysis, starting with the 5596 patients extracted (a total which is not presented anywhere), and illustrating exclusions and non-responders through to the final 1136.	These have all been corrected and a flow chart has been added.
11. Are the discussion and conclusion justified by the results?	

The conclusion and discussion is only partly justified by the results presented, and the section would benefit from some redrafting to link the two more clearly: 1. The research question was about the feasibility of using PROMs for LTCs in primary care. It is difficult to unpick from the discussion what the answer to this research question is. This will also be clearer once the research question and outcome measures are more clearly specified. It would be useful to have the key messages, or contributions to knowledge highlighted at the start of the discussion, and tied back to the research question(s). 2. It states for the first time in the discussion “response rates were the main variable of interest in this study”. This should be reflected in the abstract, results and conclusion, and in the Key Messages section at the front which should focus on response rates, the conclusions that can be drawn from them, and feasibility of collecting PROM data. Instead this focuses on the appropriateness of EQ-5D and the (speculative) point that response rates could be improved by making PROMs more engaging for patients. 3. The last paragraph of the discussion describes 3 ways in which baseline response rates could be improved. These should be either based on the research in this paper or referenced to other papers, as they currently read as somewhat speculative.	The research objectives have been clarified. This reviewer would prefer less speculation, but the second reviewer asked for more speculation. Hence, we have opted to keep it as it is. The 3 ways to improve baseline response rates (last paragraph of discussion) are the authors’ opinion and we believe it is reasonable to speculate how to address limitations identified by this study.
12. Are the study limitations discussed adequately	
Yes	
Limitations are discussed adequately. However, the limitation described as “the logistics of remotely identifying eligible patients from GP databases” is worthy of a few sentences explanation. GP systems should readily allow such identification, since these conditions are clearly specified in the QOF and data about them is already collected remotely in order to calculate QOF payments for each practice. So it would be helpful to clarify whether the authors experienced technical problems, specific to this study, or whether there are any fundamental problems	The limitations arising from the logistics of the search have been described in more detail in the limitations section.

likely to impede routine collection of PROMs.	
15. Is the standard of written English acceptable for publication	
The standard of written English is good, but needs a review for accuracy of language used throughout – e.g. (pg. 8 “There were some disease-specific differences in consent, including gender….”)	This example has been corrected.

Reviewer Name Joanne Greenhalgh

Institution and Country School of Sociology and Social Policy

University of Leeds

UK

This is an important paper and addresses a highly topical issue.

My main issue lies in the mismatch between the stated objectives of the study and the data presented. The stated objectives of the study are to evaluate the feasibility of using PROMs to monitor the quality of life of individuals with LTC in primary care. I think the study evaluates some important aspects of the feasibility of collecting PROMs data in primary care but not does not fully evaluate the feasibility of using it. The main focus of the paper lies in describing the logistical difficulties in identifying patients with LTC, the reasons why practices excluded patients, the response rates obtained and variation in response rates by patient characteristics and place. These are extremely important and the paper merits publication on this basis alone - it is important to know about the feasibility of collecting the data. I would therefore recommend that the stated objectives are modified to reflect the focus on feasibility of collection, rather than use.	The objectives have been clarified in the abstract and methods. Further details on the logistical difficulties have been included (in limitations).
Furthermore, the focus on using PROMs to monitor the quality of life of individuals suggests that the data are intended to be used at the individual patient level to inform the planning of their care by individual GPs, whereas the previous sentences in the last para of the introduction highlight the challenges of using PROMs as an indicator of the effectiveness of primary care - ie as an aggregate indicator of the quality of care. It would be helpful for the authors to make explicit the different intended uses of PROMs data and the lack of	This study looked at the use of aggregate PROMs data, we have clarified this in the introduction.

clarity around this in primary care (and indeed the tensions that surround it), as this has important implications for the discussion of their findings.	
As a reader I really wanted to know more about the qualitative aspects of feasibility, such as more details about why practices excluded individuals, how the whole enterprise was perceived by practices and by patients and what value they thought that data might have. It would be useful to know if the team collected this data and if so whether it will be published elsewhere. It would be useful to know if they collected any data to shed more light on what lay behind the regional variations in response rates.	Some further detail has been included. Qualitative interviews were also conducted after completion of the surveys, but presenting these data is beyond the scope of this manuscript.
The paper also reports the change in the EQ5D over time and notes the lack of change over time. Here the paper does begin to consider some aspects of the validity of using PROMs data as an indicator of the quality of primary care. The paper would benefit from a deeper critical and perhaps more speculative discussion of the implications of the low response rates and the lack of change in EQ5D scores for PROMs as an indicator of the quality of primary care. For example, is it going to be used to as a way of benchmarking the quality of care provided by GP practices? We know from the use of PROMs in elective surgery that low response rates increase the risk that providers may be mislassified as an outlier (see work of Andrew Street and Matt Sutton) so this may be even more of an issue in primary care given that response rates are much lower. Furthermore, the lack of change in EQ5D scores over time raises serious questions about whether change in scores (or even absolute single point in time scores) would be useful in disriminating between practices. While the paper does not set out to evaluate whether the EQ5D scores could discriminate between practices, but it might be useful to raise this in the discussion - or at least raise the issue that the lack of clarity about how PROMs data will be used hinders our understanding of how the low response rate and low change over time impact on the validity of the indicators.	This reviewer asks for more speculation. As the first reviewer asks for less speculation, we have chosen to leave this part unchanged. We have introduced further evidence on response bias.
There are also broader questions to consider about what PROMs data is intended to reflect the quality of - since people with LTC are not only cared for in primary care but also by a range of other services spanning secondary care, social care, the voluntary sector and their own families. The paper does touch on these both in the introduction and in the discussion, but again, a deeper discussion would strengthen the paper. In particular, in the introduction it might be	Whilst these comments are very valuable, they are beyond the scope of this paper.

helpful to expand on exactly how PROMs data is expected to drive change in the organisation and delivery of health services.	

VERSION 2 – REVIEW

REVIEWER	Joanne Greenhalgh University of Leeds, UK
REVIEW RETURNED	19-Jan-2014

GENERAL COMMENTS	The authors have adequately responded to my comments