PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Explaining variation in cancer survival between eleven jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey
AUTHORS	Rose, Peter; Rubin, Greg; Perera, Rafael; Almberg, Sigrun; Barisic, Andriana; Dawes, Martin; Grunfeld, Eva; Hart, Nigel; Neal, Richard; Pirotta, Marie; Sisler, Jeff; Konrad, Gerald; Toftegaard, Berit; Thulesius, Hans; Vedsted, Peter; Young, Jane; Hamilton, Willie; ., ICBP Module 3 Working Group

VERSION 1 - REVIEW

REVIEWER	Peter Murchie Centre of Academic Primary Care, Division of Applied Health Sciences, University of Aberdeen, UK
REVIEW RETURNED	19-Jan-2015

GENERAL COMMENTS	This paper reports on a significant and heroic research effort to discern differences in referral thresholds between primary care practitioners in six different countries. The methods used are innovative and, at the current time, probably represent as good an effort at could be made at investigating this important question. The authors do, perhaps, gloss over the main limitation of their work - namely are they showing cause or effect? The clear implication as argued is that GPs referral thresholds differ in differ countries, and that GPs in countries with poorer survival are too choosy in whom to refer. By extent lowering those thresholds (i.e. weakening the gate-keeper function) in countries with poorer survival will improve early detection. The reality, however, is likely that GP referral behaviour is dictated by the resources that they have available to them, such that the most effective measures are likely to be targeted at health service organisation rather than GP behaviour. I think this point could be reflected in the conclusions of the paper. A second, and obvious point, is that there are likely to be subtle confounders that are not discussed here - for example poorer outcomes may just reflect better IT systems, and one suspects that availability of private routes to referral and diagnosis are likely to have had an impact on the reported results. Both these points could be acknowledged more explicitly. That said, this is a significant and important piece of work and deserves to be published. I have the following specific comments: Abstract: The objectives should be more specific to the aims of the study rather than the somewhat vague - "the study investigates factors that may contribute to cancer suvival differences....."

	Introduction: The presentation of references supporting the notion that longer delays mean poorer outcomes is a little one-sided. The introduction should reflect the notion that it is far from clear that there is a strong relationship between moderate delays and outcome. "Public awareness of signs and symptoms and beliefs about cancer appear to be quite similar...." this is a striking comment to not be supported by a reference or further justification which it would be good to see. As per my comment above, it would be good to the introduction acknowledge that the most powerful predictor of GPs readiness to investigate or refer symptoms possibly indicating cancer are likely to be health-service related and not to be intrinsic characteristics. It seems unlikely that GPs in different countries would vary in judgements that a patients's symptoms might warrant cancer. Methods: It would be good to have a little bit more detail about how the online survey was administered. Was a standard approach used? Could there be differences in the way this was done between jurisdictions which might have affected the response/engagement with the project. A few sentences on how the survey was validated would be useful, even though a reference is provided. It would reassure the reader that this wasn't just simple face validity. It seems a pity that respondents were aware that this study was linked to cancer. This might have reduced the differences seen in a blinded study and be more meaningful. Perhaps the authors could speculate on this? Linked to above, the differing sampling and approach method could be a major source of bias. More detail needs to be provided about the different approaches uses so that we can be reassured on this point. The wide variation in response rates (5% - 45%) underpins this point! Why were out-of-hours practitioners included. OOH practitioners would not usually refer non-emergency symptomatic patients in the UK. Is this different elsewhere? Results: Was it legitimate to include the Northern Irish data even though that jurisdiction fell beneath the recruitment target? A brief comment is needed. Discussion: I would like to see some reflection on my points above in the limitations section: what was really being measured here - GP behaviour or health service provision - and what can we learn from this study about the link between them. The paragraph beginning "Those responding were not wholly representative..." should be rewritten I think to pick up on the risk that the recruitment methods have introduced confounding factors that the analysis will not control for e.g. prevalence of private insurance, diagnostic centres; or differential use of incentives to take

	part in the study. Within the existing paragraph the statement that "it is unlikely that selection bias will operate in different directions in each country" is a brave one. It very well might, given the non-uniform recruitment methods so this needs a closer justification and explanation. Furthermore, why should this study have picked up the very best and most interested GPs? I can't see there is any justification for believing that. Furthermore, why should the best and most interested GPs investigate or refer earlier? The opposite might be true. A clearer statement of "where next" should be provided. The current one is a little vague and all-encompassing. What should be the first targets? What is the role for routine health service data in future efforts? Should this be used to validate the current findings as a first step to ensure the validity of the methods? Overall: This is a large and complex study addressing a fascinating question in an innovative way. Despite, this, there are obvious limitations, some of which are not given due prominence in the current report. I believe, however, with appropriate attention to these the manuscript is an important addition to the literature, and raises some key issues for further investigation.

REVIEWER	María-Dolores Chirlaque Murcia Cancer Registry. Regional Health Council, Murcia, Spain
REVIEW RETURNED	24-Feb-2015

GENERAL COMMENTS	The present manuscript on primary care practitioners practice and specific cancer survival like outcome is an interesting and current issue to study. The draft, in general, is complete, well analysed, and the discussion addresses the main points. The design of the study is ecological because analyze correlation between readiness of PCPs to investigate symptoms of cancer and cancer survival. This aspect is not deal with along the draft and ecological bias should be describe, analyzed and take into account like possible factor influencing in the results. More update results on survival have been published by Allemani C et al (Global surveillance of cancer survival 1995-2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2)) in Lancet (Nov 2014) including results up to 2009; You mention, like a weaknesses in the discussion, that the latest comparative survival data for the three cancers in participating jurisdictions is from 2007. A suggestion is to review this draft and to test whether it provides novel results. Aspects related to the readiness of PCPs to investigate or refer for suspected cancer have been widely argued in the present draft. However, aspect related to 1 year survival and 5 year survival conditional on surviving at least 1 year (conditional 5 year survival) would need some clarifications. 5-year relative survival conditional on 1-year survival is considered as helpful for patients and clinicians. The difference between 1 and 5 years relative survival is that 1-year directly provides information on early survival, while the conditional

	5-year survival goes five years into the future and four years back again. Thus, 1-year RS is indirectly related to the proportion of late stage tumours. A considerable proportion of patients with advanced disease die during the first year, and those who survive have a more uniform stage distribution. Differences between conditional and unconditional relative survival are lower for cancers with relatively good prognoses, like colorectal, and higher for cancer with unfavourable prognosis like lung. Other point is to compare with relative survival instead of observed survival. What type of survival have you used in the present draft like outcome? This issue has not been explained in methods. Factor on process of care like primary care structure or PCPs characteristics have been correctly addressed in the present draft. However, factors of the own patients or tumours (demographics, comorbidities, stage/grade, histology, etc.) are important determinant of overall survival and could be commented in the discussion. I miss a specific paragraph mentioning that an ethical committee had approved the study or that participants gave consent to participate. The abstract is not explanatory of the study by itself. Perhaps, the limit of words have made author to present a shortened version, but this version, is very difficult to understand. Perhaps someone belonging to ICBP is familiarizing oneself with the methodology of the study, but not readers in general. The following comments refer to the abstract. The objective is not completely described: 'This study investigates factors that may contribute to cancer survival differences across these jurisdictions' but you do not specify what kind of factors. In Design section you should add the target, for example: A validated survey administrated via internet to PCP…' You do not mention the phase of the vignette previously to 'Primary and secondary outcome measures', thus, the phrase 'Analysis compared the cumulative proportion of PCPs in each jurisdiction opting to investigate or refer at each phase for each vignette' is not clear for the reader. In the article summary authors said Response rates were sub-optimal and respondents were not totally representative of the PCPs in all jurisdictions. I suggest considering the inclusion of the response rate in the abstract.

VERSION 1 – AUTHOR RESPONSE

We agree with most points raised by the peer reviewers and have set out below how we have addressed each specific point (numbered throughout).

This paper reports on a significant and heroic research effort to discern differences in referral thresholds between primary care practitioners in six different countries. The methods used are innovative and, at the current time, probably represent as good an effort at could be made at investigating this important question. The authors do, perhaps, gloss over the main limitation of their work - namely are they showing cause or effect? The clear implication as argued is that GPs referral thresholds differ in differ countries, and that GPs in countries with poorer survival are too choosy in whom to refer. By extent lowering those thresholds (i.e. weakening the gate-keeper function) in countries with poorer survival will improve early detection. The reality, however, is likely that GP referral behaviour is dictated by the resources that they have available to them, such that the most

effective measures are likely to be targeted at health service organisation rather than GP behaviour. I think this point could be reflected in the conclusions of the paper. A second, and obvious point, is that there are likely to be subtle confounders that are not discussed here - for example poorer outcomes may just reflect better IT systems, and one suspects that availability of private routes to referral and diagnosis are likely to have had an impact on the reported results. Both these points could be acknowledged more explicitly. That said, this is a significant and important piece of work and deserves to be published.

I have the following specific comments:

1) Abstract:
The objectives should be more specific to the aims of the study rather than the somewhat vague - "the study investigates factors that may contribute to cancer survival differences....."

Updated the sentence to read: The International Cancer Benchmarking Partnership (ICBP) is a collaboration between six countries and 12 jurisdictions with similar primary care led health services. This study investigates primary care physician behaviour and systems that may contribute to the timeliness of investigating for cancer and subsequently cancer survival differences.
2) Introduction:
The presentation of references supporting the notion that longer delays mean poorer outcomes is a little one-sided. The introduction should reflect the notion that it is far from clear that there is a strong relationship between moderate delays and outcome.

Updated the sentence to read: There is some evidence that delay between presentation and diagnosis (the diagnostic interval)[7] is associated with poorer outcomes[8-11] but the factors involved are complex and the strength of the relationship is unclear.

Included the new reference to [11]: Neal RD, Tharmanathan P, France B, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. British Journal of Cancer 2015; 1-16 DOI:10.1038/bjc.2015.48

3)"Public awareness of signs and symptoms and beliefs about cancer appear to be quite similar...." this is a striking comment to not be supported by a reference or further justification which it would be good to see.

Reference 11 supports this statement as noted.

4) As per my comment above, it would be good to the introduction acknowledge that the most powerful predictor of GPs readiness to investigate or refer symptoms possibly indicating cancer are likely to be health-service related and not to be intrinsic characteristics. It seems unlikely that GPs in different countries would vary in judgements that a patients's symptoms might warrant cancer.

Updated sentence and added the relevant reference: There are many system factors that will influence a PCPs decision to act including guidelines, access to investigations, culture of collaboration between primary and secondary care and these will all contribute to PCP behaviour [15].

Methods:
4a) It would be good to have a little bit more detail about how the online survey was administered. Was a standard approach used? Could there be differences in the way this was done between jurisdictions which might have affected the response/engagement with the project.
This is provided in supplementary table 1 (see also response to point 7 below).

5) A few sentences on how the survey was validated would be useful, even though a reference is provided. It would reassure the reader that this wasn't just simple face validity.

We developed an online survey of PCPs exploring differences in their behaviours, attitudes, knowledge and skills relating to cancer diagnosis. Development involved iterative discussion with

international partners at every stage of development. The overall validation was initially undertaken in England with two rounds of validation using a cognitive interviewing technique with PCPs following completion of the draft survey. Validation of the completed survey was tested in all jurisdictions, particularly to ensure that translation had not altered meaning. There were two questions relating to access to tests and internal consistency was measured by comparison of the answers to these questions. The development and validation has been described in detail elsewhere[15].

6) It seems a pity that respondents were aware that this study was linked to cancer. This might have reduced the differences seen in a blinded study and be more meaningful. Perhaps the authors could speculate on this?

Updated the strengths and weaknesses to read: In addition, respondents were aware the survey was part of a study linked to cancer for ethical reasons; responses might have been different in a blinded study. However, this bias will tend to underestimate the possible variation and our results are thus minimum estimates of the correlation between readiness to investigate and survival. T

7) Linked to above, the differing sampling and approach method could be a major source of bias. More detail needs to be provided about the different approaches uses so that we can be reassured on this point. The wide variation in response rates (5% - 45%) underpins this point!

This is provided in supplementary table 1 and we have added a sentence in the text to highlight this point: 4a and 7) While variation in sampling methods and approaches might be expected to introduce sample bias and to affect response rates, there are no observable trends that would suggest that this is true. (Supplementary Table 1).

8) Why were out-of-hours practitioners included. OOH practitioners would not usually refer non-emergency symptomatic patients in the UK. Is this different elsewhere?
The survey was intended to include anyone involved in primary care and there was an option for respondents to identify whether they worked OOH. Our description of the sample reflected the dissemination / inclusion criteria, however none of our respondents were OOH PCPs.
We have amended the sentence under participants to re: Participants were PCPs working predominantly in clinical practice, including locums: retired PCPs, and those in training were not eligible.

Results:
9) Was it legitimate to include the Northern Irish data even though that jurisdiction fell beneath the recruitment target? A brief comment is needed.
Inserted a new comment to read: Northern Ireland was included for completeness, even though their recruitment fell below target, acknowledging that confidence intervals would be wider than anticipated. Having collected the data, scientifically it would have been worse to omit it.

Discussion:
10) I would like to see some reflection on my points above in the limitations section: what was really being measured here - GP behaviour or health service provision - and what can we learn from this study about the link between them.
The comments above were: The clear implication as argued is that GPs referral thresholds differ in different countries, and that GPs in countries with poorer survival are too selective in whom to refer. By extent lowering those thresholds (i.e. weakening the gate-keeper function) in countries with poorer survival would be expected to improve early detection. The reality, however, is likely that GP referral behaviour is dictated by the resources that they have available to them, such that the most effective measures are likely to be targeted at health service organisation rather than GP behaviour (though the latter will need to change in recognition of system change). I think this point could be reflected in the conclusions of the paper.
Updated to read: The readiness of PCPs to act consists of personal attributes (for example knowledge and attitudes about cancer as well as perceptions about the role of PCPs) and system features (for example guidelines, availability of tests/referral and waiting time for results). The analysis did not identify which of these factors are likely to be the main influences on the PCP readiness to act.

11) A second, and obvious point, is that there are likely to be subtle confounders that are not

discussed here - for example poorer outcomes may just reflect better IT systems, and one suspects that availability of private routes to referral and diagnosis are likely to have had an impact on the reported results. Both these points could be acknowledged more explicitly.
Inserted a new sentence to read: Other hidden confounders may have influenced the results, but they are unlikely to have been major

12) The paragraph beginning "Those responding were not wholly representative..." should be rewritten I think to pick up on the risk that the recruitment methods have introduced confounding factors that the analysis will not control for e.g. prevalence of private insurance, diagnostic centres; or differential use of incentives to take part in the study.
We contest this - private insurance is a major issue in countries with primarily primary care driven health services and no country had diagnostic centres at the time.
Re differential use of incentives we have added: Evidence from Australia suggests that while respondents had more positive views about cancer compared to non-responders, the magnitude of this difference is the same irrespective of incentives (conditional or otherwise) [27].

13) Within the existing paragraph the statement that "it is unlikely that selection bias will operate in different directions in each country" is a brave one. It very well might, given the non-uniform recruitment methods so this needs a closer justification and explanation. Furthermore, why should this study have picked up the very best and most interested GPs? I can't see there is any justification for believing that. Furthermore, why should the best and most interested GPs investigate or refer earlier? The opposite might be true.
We agree that this was speculations and presented hypothesis. Therefore, we have deleted the rest of this paragraph.

14) A clearer statement of "where next" should be provided. The current one is a little vague and all-encompassing. What should be the first targets? What is the role for routine health service data in future efforts? Should this be used to validate the current findings as a first step to ensure the validity of the methods?

Amended to read: The study supports the ecologic findings that there is a correlation between the health care system and the way GPs perform clinical diagnosis. Therefore, it seems appropriate to perform studies testing whether changed access to investigations will also change the GPs' readiness. Further, we need studies on the outcome of different access to investigations like stage distribution and cohort studies on survival and mortality.

Overall: This is a large and complex study addressing a fascinating question in an innovative way. Despite, this, there are obvious limitations, some of which are not given due prominence in the current report. I believe, however, with appropriate attention to these the manuscript is an important addition to the literature, and raises some key issues for further investigation.

Reviewer Name María-Dolores Chirlaque
Institution and Country Murcia Cancer Registry.
Regional Health Council, Murcia, Spain.
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
The present manuscript on primary care practitioners practice and specific cancer survival like outcome is an interesting and current issue to study. The draft, in general, is complete, well analysed, and the discussion addresses the main points.

15) The design of the study is ecological because analyze correlation between readiness of PCPs to investigate symptoms of cancer and cancer survival. This aspect is not deal with along the draft and ecological bias should be describe, analyzed and take into account like possible factor influencing in the results.

In the discussion of the methods we have now included a section: This study used an ecologic

outcome which makes the risk of an ecologic fallacy important. We do not know whether the correlation with readiness and survival is causal or simply an indicator of e.g. delivering better medical quality in general or another relation. Therefore, this study raises the hypotheses of an interaction between the GPs' readiness and the system in which they perform and adds to the science pointing towards this important implication.

16) More update results on survival have been published by Allemani C et al (Global surveillance of cancer survival 1995-2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2)) in Lancet (Nov 2014) including results up to 2009; You mention, like a weaknesses in the discussion, that the latest comparative survival data for the three cancers in participating jurisdictions is from 2007. A suggestion is to review this draft and to test whether it provides novel results.
This is a good point and if there had been more updated and timely data on survival we should use them. We have considered using Concord data to re-run the analysis however:
• While it provides updated survival data, with outcomes to 2009, there is no significant difference in survival trends between 2007 (from ICBP Module 1) and the new Concord data
• Outcomes to 2009 are still earlier than the period of our survey, so we would have the same limitation
• Our study uniquely gives detail for Canadian provinces, Australian states or devolved countries within the UK – Concord has only reported national data so far
• The inclusion criteria for each cancer in concord is slightly different – e.g. colon and rectal cancer are separated out in Concord, in ICBP they are considered together

17) Aspects related to the readiness of PCPs to investigate or refer for suspected cancer have been widely argued in the present draft. However, aspect related to 1 year survival and 5 year survival conditional on surviving at least 1 year (conditional 5 year survival) would need some clarifications. 5-year relative survival conditional on 1-year survival is considered as helpful for patients and clinicians. The difference between 1 and 5 years relative survival is that 1-year directly provides information on early survival, while the conditional 5-year survival goes five years into the future and four years back again. Thus, 1-year RS is indirectly related to the proportion of late stage tumours. A considerable proportion of patients with advanced disease die during the first year, and those who survive have a more uniform stage distribution. Differences between conditional and unconditional relative survival are lower for cancers with relatively good prognoses, like colorectal, and higher for cancer with unfavourable prognosis like lung.
We absolutely agree with the reviewer and this was indeed the reason we included both survival measures. We can also add, that another reason was the importance of possible comorbidity as e.g. lung cancer is associated with a higher proportion of comorbidity which will affect the survival. We have now added a section where we shortly describe the reasons as indicated by the reviewer. We used 1 year and 5/1 year survival on the advice of Prof Michel Colman from the Cancer Survival Group at the London School for Hygiene and Tropical Medicine as delay in diagnosis will affect both to an extent. He says the main effect on delay is in 1 year survival and that is why we used this as main outcome measure and 5/1 as supplementary.

18) Other point is to compare with relative survival instead of observed survival. What type of survival have you used in the present draft like outcome? This issue has not been explained in methods.

We did use relative survival and agree with the reviewer that is the optimal.

19) Factor on process of care like primary care structure or PCPs characteristics have been correctly addressed in the present draft. However, factors of the own patients or tumours (demographics, comorbidities, stage/grade, histology, etc.) are important determinant of overall survival and could be commented in the discussion.

We agree that a lot of other factors influence the survival of cancer patients. However, we have no particular reason to believe that the readiness of PCPs in a country would be biased by different overall mortality, demographics and definitely not the later stage and histology. So in this particular study only survival will be affected and the factors are thus not confounders in that sense.

20) I miss a specific paragraph mentioning that an ethical committee had approved the study or that participants gave consent to participate.

We have added a paragraph on ethics and consent in the methods section and we can refer to supplementary table 3 which lists all ethics approvals for ICBP Module 3.

21) The abstract is not explanatory of the study by itself. Perhaps, the limit of words have made author to present a shortened version, but this version, is very difficult to understand. Perhaps someone belonging to ICBP is familiarizing oneself with the methodology of the study, but not readers in general. The following comments refer to the abstract. The objective is not completely described: 'This study investigates factors that may contribute to cancer survival differences across these jurisdictions' but you do not specify what kind of factors.
Addressed in 1) above

22) In Design section you should add the target, for example: A validated survey administrated via internet to PCP…'
Amended to read: A validated survey administered to primary care physicians (PCPs) via the internet set out in two parts: direct questions on primary care structure and practice relating to cancer diagnosis, and clinical vignettes, assessing management of scenarios relating to the diagnosis of lung, colorectal or ovarian cancer.

23) You do not mention the phase of the vignette previously to 'Primary and secondary outcome measures', thus, the phrase 'Analysis compared the cumulative proportion of PCPs in each jurisdiction opting to investigate or refer at each phase for each vignette' is not clear for the reader.
Given word count it is not possible to add more. This is fully presented and discussed in the methods paper which we hope to be able to refer to.

24) In the article summary authors said Response rates were sub-optimal and respondents were not totally representative of the PCPs in all jurisdictions. I suggest considering the inclusion of the response rate in the abstract.
Updated to include (ranging from 5.5% to 45.6%)