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)	Validating the use of Hospital Episode Statistics data and comparison of costing methodologies for economic evaluation: an end of life case study from the Cluster randomised triAl of PSA testing for Prostate cancer (CAP)
AUTHORS	Thorn, Joanna; Turner, Emma; Hounsome, Luke; Walsh, Eleanor; Down, Liz; Verne, Julia; Donovan, Jenny; Neal, David; Hamdy, Freddie; Martin, Richard; Noble, Sian

VERSION 1 - REVIEW

REVIEWER	Chunhuan Lao Waikato Clinical Campus, University of Auckland
REVIEW RETURNED	02-Feb-2016

GENERAL COMMENTS	This paper covers an interesting topic, validating the accuracy of NHS reference costs and HES data for costing inpatient resource using the medical records as comparison. However, there are more problems with the paper. Title: 1. Why does routine data mean? From the main text, it seems it means HES data rather than medical records. It would be better to specify it. Abstract: 1. Line 25: Do you mean PSA testing? 2. Line 34 is confusing. Better to rewrite it. Introduction 1. HES data was introduced in the second paragraph. It would be helpful to introduce medical records and cost data from finance department, and to explain the connections and differences among the three datasets. Methods 1. The methods included a lot of details about the CAP trial. Some are not relevant to this study, e.g. Line 40-41. 2. The estimated costs cannot be considered as end-of-life cost. Was the exclusion of men who died within one year after the prostate cancer diagnosis just to limit the number of patients included? 3. The third set data, cost data from trust finance department, was confusing. It was not included to validate the accuracy of HES data. it is better to be removed? 4. The statistic methods that generating the p values need to be specified. The difference might be significant with one method, but might not be significant with another one. The costs from both

	datasets need to be compared in pairs. Results 1. Since it is not a representative sample, the characteristics of this cohort is not useful for the audiences. 2. It is not necessary to present both figure 1 and table 1. They are showing the same results. The only difference is the p values in Table 1. 3. It is strange that p value in months 11 was 0.2, but the mean cost from HES data was almost 3 times the cost from medical records. Please check and explain. Discussion and conclusion 1. The research purpose in the last paragraph in the discussion was not consistent with those in the abstract and in the introduction. This study seems to investigate both the costing methods and the data sources. 2. The cost estimated from HES data and NHS reference cost was validated by the medical records and NHS reference cost. However, the latter (especially the costing method applied) has not been proven to be reliable. Comparing the numbers of day of hospital stay from both datasets might be more convincing.

REVIEWER	Steven Zeliadt University of Washington, Seattle, Washington USA
REVIEW RETURNED	05-Feb-2016

GENERAL COMMENTS	I greatly appreciate this manuscript as scrutiny of cost sources is an important issue. While I don't disagree generally with the authors conclusions, I am more troubled by the size of the differences than they appear to be. As an 8% bias is not meaningless. The presentation is clear and does allow the reader to drawn his own conclusions. There are a few minor points of clarification that would strengthen the manuscript. 1. P.8, line 33. "Mean imputation was used where mapping was not possible (e.g., where the ward type was unknown or unclear in the medical records it was replaced by the mean)." Later the authors indicate that these were imputed for 4.7% of missing episodes. A bit more detail would be helpful about imputation approach, and how sensitive the findings are to these imputed costs. 2. The authors indicate this was done: p.9, line 44. "A manual inspection of inpatient events occurring in medical record review and HES records was conducted, with the aim of identifying missing events in either dataset." But I may have missed the results – was the conclusion that there was complete congruence except for the 11 men with events in HES, and 7 in medical records review that were missing other records? Were there any differences in what events were recorded – e.g. one source could have recorded an invasive procedure while other could have recorded something different? Or details such as imaging events or inpatient medications were missing?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Chunhuan Lao
Waikato Clinical Campus, University of Auckland

Please leave your comments for the authors below
This paper covers an interesting topic, validating the accuracy of NHS reference costs and HES data for costing inpatient resource using the medical records as comparison. However, there are more problems with the paper.

Title:
1. Why does routine data mean? From the main text, it seems it means HES data rather than medical records. It would be better to specify it.
The title has been amended to reflect the fact that the study is based on HES data

Abstract:
1. Line 25: Do you mean PSA testing?
The word 'testing' has been added to correct the sentence.

2. Line 34 is confusing. Better to rewrite it.
The line has been rewritten to provide clarity: "11 men (3.8%) had events identified in HES that were all missing from medical record review, while seven men (2.4%) had events identified in medical record review that were all missing from HES. "

Introduction
1. HES data was introduced in the second paragraph. It would be helpful to introduce medical records and cost data from finance department, and to explain the connections and differences among the three datasets.

The paragraph has been amended to clarify the distinction between the collection of resource-use data and the application of costs (the choices are therefore HES or medical records, and NHS reference costs or data from finance departments).
"There are necessary tradeoffs between research resources available and the accuracy of the costs that can be obtained. Typically, health economists have collected information on resources consumed by extracting data from medical records or by asking patients themselves to supply data. Similarly, cost data relating to resource use identified within a secondary care trial is often obtained by asking finance departments in NHS trusts to supply data; over half the HTA studies that reported an economic evaluation used costs from local sources 9. However, these can be time-consuming and costly processes, and may not represent the best use of research resources. Increased access to routinely collected data in the UK, such as the NHS reference costs and the Hospital Episode Statistics (HES) dataset in England 10 11, is affording researchers opportunities for obtaining costs in healthcare more readily. For example, HES data can be used to identify relevant resource use, while the NHS reference costs can potentially be used in place of data from finance departments. Each hospital inpatient….."

Methods
1. The methods included a lot of details about the CAP trial. Some are not relevant to this study, e.g. Line 40-41.

The amount of detail included has been reduced.

2. The estimated costs cannot be considered as end-of-life cost. Was the exclusion of men who died within one year after the prostate cancer diagnosis just to limit the number of patients included?
We excluded men who died less than a year following a diagnosis of prostate cancer to ensure that the final-year cost profiles were based on a homogenous sample; including men who were diagnosed only six months prior to death, for example, would have affected the costs in the earlier months. Although the estimated costs include only inpatient costs and cannot therefore be considered as end-of-life costs (as we note on page 16), we felt that a final year inpatient cost profile was still of value.

3. The third set data, cost data from trust finance department, was confusing. It was not included to validate the accuracy of HES data. it is better to be removed?
Originally, we planned to use the finance department data for validation purposes – the poor response prevented us from doing so in a meaningful way. However, we believe the finance department data are interesting and worth retaining to help other scientists avoid wasting time. We have amended to the title to reflect the fact that there is also an element of general costing methodology reported in the paper: “Validating the use of Hospital Episode Statistics data and comparison of costing methodologies for economic evaluation: an end of life case study from the Cluster randomised triAl of PSA testing for Prostate cancer (CAP)”. We have also updated the objectives in the abstract: “Objectives: To evaluate the accuracy of routine data for costing inpatient resource use in a large clinical trial and to investigate costing methodologies.”

4. The statistic methods that generating the p values need to be specified. The difference might be significant with one method, but might not be significant with another one. The costs from both datasets need to be compared in pairs.
The relevant information has been added to the Methods section: “Two-sample paired ttests were used to derive p values for the differences between costs; a 1% significance level is appropriate for the multiple tests carried out on these data.”

Results
1. Since it is not a representative sample, the characteristics of this cohort is not useful for the audiences.
We acknowledge that the sample is not representative; however, we believe that the characteristics remain important to allow readers to interpret results in context.

2. It is not necessary to present both figure 1 and table 1. They are showing the same results. The only difference is the p values in Table 1.
Table 1 has been removed, along with the associated text in the results section.

3. It is strange that p value in months 11 was 0.2, but the mean cost from HES data was almost 3 times the cost from medical records. Please check and explain.
The high mean HES cost in month 11 prior to death was driven by a single man with a long stay beyond the trimpoint; the resulting high variance means that the difference between the HES and medical record data is consistent with chance. This is explained on page 10: “The wider variation in month 11 prior to death arose from a single man with a very lengthy stay beyond the trim point in the HES dataset.”

Discussion and conclusion
1. The research purpose in the last paragraph in the discussion was not consistent with those in the abstract and in the introduction. This study seems to investigate both the costing methods and the

data sources.
The paragraph has been improved to read: “The purpose of this study was to investigate different costing methods that could be employed in clinical trials, and to validate the use of the HES inpatient dataset in particular; the costs derived should not, therefore, be interpreted as end of life costs. The study was restricted to inpatient resource use 18; however, informal caregiving 33, chemotherapy 34 and hospice care 35 may also be substantial in the end of life period. Derivation of an overall cost of end of life prostate cancer care is planned for future work.”

2. The cost estimated from HES data and NHS reference cost was validated by the medical records and NHS reference cost. However, the latter (especially the costing method applied) has not been proven to be reliable. Comparing the numbers of day of hospital stay from both datasets might be more convincing.

We were interested in the impact of the different costing methodologies on the costs associated with resource use in a clinical trial. Therefore, in this context, we believe the costs are more relevant than the number of nights spent in hospital. One of the conclusions of the study is that neither method truly represents a gold standard.

Reviewer: 2

Steven Zeliadt
University of Washington, Seattle, Washington USA

Please leave your comments for the authors below
I greatly appreciate this manuscript as scrutiny of cost sources is an important issue. While I don’t disagree generally with the authors conclusions, I am more troubled by the size of the differences than they appear to be. As an 8% bias is not meaningless. The presentation is clear and does allow the reader to drawn his own conclusions.

There are a few minor points of clarification that would strengthen the manuscript.

1. P.8, line 33. “Mean imputation was used where mapping was not possible (e.g., where the ward type was unknown or unclear in the medical records it was replaced by the mean).” Later the authors indicate that these were imputed for 4.7% of missing episodes. A bit more detail would be helpful about imputation approach, and how sensitive the findings are to these imputed costs.
Investigation of the sensitivity to the imputed costs has been conducted, and is now reported in the manuscript: “Sensitivity analysis indicated that this result was not sensitive to the imputed value used in the medical record review costing; all confidence intervals remained overlapping.”
Further details have also been added to the Methods section. “Where the ward type was unknown or unclear in the medical records, mapping to service codes was not possible. Mean imputation was used in these cases: a mean cost per night (weighted by the number of nights) was derived from all episodes with available cost data, and applied to events with missing costs. Sensitivity to the imputed value was explored by using the maximum and minimum costs. ”

2. The authors indicate this was done: p.9, line 44. “A manual inspection of inpatient events occurring in medical record review and HES records was conducted, with the aim of identifying missing events in either dataset.” But I may have missed the results – was the conclusion that there was complete congruence except for the 11 men with events in HES, and 7 in medical records review that were missing other records? Were there any differences in what events were recorded – e.g. one source could have recorded an invasive procedure while other could have recorded something different? Or

details such as imaging events or inpatient medications were missing?
The aim of the manual inspection was to identify missing events in either dataset. There were many differences between the events in the two datasets; however, because the definitions of events were different, it was not possible to reliably ascertain whether an event recorded in one dataset was truly missing in the other or whether it had simply been recorded differently (possibly for the reasons you suggest). Therefore, we concentrated on those occasions where a man had no events recorded at all in one or other of the datasets because there was no ambiguity – for example, if a man had 5 events in HES but none were picked up through the medical records it is clear that those events are truly missing from the medical records (and vice versa). The lack of congruence beyond the 11 men with all records missing in medical records and 7 men with all records missing in HES has now been clarified: “Therefore, events in the datasets were not congruent; minor differences in dates and lengths of stay were also common. However, there was no ambiguity where men had zero events recorded in either one or the other dataset.”

VERSION 2 – REVIEW

REVIEWER	CHUNHUAN LAO Waikato Clinical Campus The University of Auckland New Zealand
REVIEW RETURNED	15-Mar-2016

GENERAL COMMENTS	I am happy with the revision and the responses from the authors.

REVIEWER	Steven Zeliadt, PhD University of Washington, USA
REVIEW RETURNED	26-Mar-2016

GENERAL COMMENTS	The manuscript reads very well and the findings are clearly explained. The level of detail and additional explanation of the results seems appropriate and I found the revised clarifications to be helpful. The conclusions and the authors interpretations match the findings quite well in my opinion. I appreciate the inclusion of the failed efforts with the finance department and feel it is very good to include this detail in the manuscript to help future investigators considering this approach.