Author's response to reviews

Title: OvidSP Medline-to-PubMed search filter translation: a methodology for extending search filter range to include PubMed's unique content

Authors:

Raechel A Damarell (raechel.damarell@flinders.edu.au)
Jennifer J Tieman (jennifer.tieman@flinders.edu.au)
Ruth M Sladek (ruth.sladek@flinders.edu.au)

Version: 2 Date: 14 April 2013

Author's response to reviews: see over
12 April 2013

Dr Roberta Scherer, BMC Medical Research Methodology

Re: Manuscript 2121707813901775. *OvidSP Medline-to-PubMed search filter translation: a methodology for extending search filter range to include PubMed’s unique content [new title]* by Raechel A. Damarell et al.

Dear Dr Scherer,

Thank you for the opportunity to resubmit a revised version of our manuscript. We have now uploaded a revised manuscript.

Please also find following this letter our point-by-point response to the concerns raised by the reviewers and the Associate Editor’s summation of those concerns. We have considered each comment in full and made appropriate changes to the manuscript in order to improve clarity. One significant change includes a new title: *OvidSP Medline-to-PubMed search filter translation: a methodology for extending search filter range to include PubMed’s unique content*

One major concern was that the method could not be reproduced. However, as the point-by-point response details, the discrepancy between our results and the reviewer’s findings was largely due to the fact that the reproduction omitted several significant steps described within the methodology. This difficulty in interpreting our method has highlighted the need for us to include the original strategy as a table in the document for improved clarity. This has now been provided.

In all, we are very grateful to the reviewers for their feedback and believe their suggestions have helped us improve the quality of the manuscript.

We trust we have addressed all issues to your satisfaction. If not, we would be pleased to correspond further on any point of concern.

Yours faithfully,

Raechel Damarell (corresponding author)
Phone: 61 8 8204 4650
Fax: 61 8 82773721
Email: Raechel.damarell@flinders.edu.au
Response to reviewers’ comments

Reviewer 1

Major compulsory revisions

1. The authors claim to have created a validated search filter but the only validation of their filter is on the gold standard set of records with which they created the filter.... In the absence of true validation the authors should replace the use of the adjective ‘validated’ in their paper, unless they are able to test the performance of their filter on another set of relevant records or can show that their filter was really validated.

Authors’ response: This study does not describe a method for creating a ‘search filter.’ The method outlined uses an already validated OvidSP Medline-to-PubMed search filter translation to empirically develop a ‘supplementary search strategy’ for retrieving PubMed’s additional, non-indexed content. The reviewer is therefore correct. This strategy has not been ‘validated’ according to the general understanding of the term in the search filter development field. The misleading reference to validation has therefore been removed (final paragraph of the Background section, p.8).

To prevent the misconception that this paper is about developing a ‘search filter,’ the following paragraph was added to the Discussion section (Methodology assessment, para. 4):

The purpose of this study was to explicitly acknowledge PubMed’s unique content and provide a systematic, reproducible method for accounting for this content in translating a search filter from OvidSP Medline. It was not our aim to develop an additional ‘search filter’ for capturing this content, rather an empirically tested extension of an already validated filter which works in tandem with this filter to incrementally improve retrieval across the entire PubMed system.

While our aim was not to develop a search filter according to the established methodology (as outlined by the CADTH CAI, the ‘So many filters, so little time’ appraisal checklist, and the Jenkins review paper), we were interested in the reviewer’s suggestion that the test set (our ‘lost set’) could be divided into subsets—one for developing and testing and one for validation. We have acknowledged our intention to further explore the possibilities inherent in this suggestion in the Discussion section (Methodology assessment, para. 4):

Although it was outside the scope of this study, a future study might extend the methodology to ‘validating’ this additional component in the traditional search filter development sense, using a ‘gold standard’ set of relevant and non-relevant citations. This approach would make it possible to provide the standard metrics of search performance such as sensitivity, specificity, and precision.
2. The authors should provide the precision of their filter. There are several points during the paper where the authors mention precision, but they do not present the precision (alongside the sensitivity) of their filter. They have the precision data as part of the testing and should present it. This is particularly important because the sensitivity is low – and to accept sensitivity this low (80% of relevant records are not retrieved by the strategy) there must be really high precision to compensate for this.

Authors’ response: In this manuscript we have used the terms ‘precise’ and ‘precision’ in the sense of ‘highly relevant’ retrieval but not as a metric of filter performance. We understand that this may be considered misleading. Indeed, this study has no basis upon which to measure the precision of the supplementary textword search strategy. (The validated search filter, however, has a precision of 75%. This figure was presented as part of the original study (Damarell 2011), of which this current study is an extension.) In this current study, there is no gold standard set of relevant/irrelevant citations for testing precision, only a test set of all relevant citations. We cannot, therefore, provide any precision data.

As stated in the limitations section of the discussion (second to last paragraph), a post-hoc clinician assessment of the precision of the filter extension would have been a valuable addition to the study. However, this was not possible at the time.

To avoid confusion, the following references to precision and precise retrieval have been omitted from the manuscript:

<table>
<thead>
<tr>
<th>Location in manuscript</th>
<th>Original sentence</th>
<th>Replacement sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract background</td>
<td>We describe a reproducible methodology that uses a validated PubMed search filter translation to create a textword-only strategy for precise retrieval of PubMed’s unique heart failure literature</td>
<td>We describe a reproducible methodology that uses a validated PubMed search filter translation to create a textword-only strategy for retrieving PubMed’s unique heart failure literature</td>
</tr>
<tr>
<td>Objectives (p. 9), objective number 2</td>
<td>Identify high precision textword alternatives ...</td>
<td>Identify high frequency textword alternatives ...</td>
</tr>
<tr>
<td>Discussion: methodology assessment, para. 3</td>
<td>Secondly, the cut-off point of 5% for identifying ‘high frequency’ terms was chosen arbitrarily. Whilst it appears reasonable, it may have inadvertently eliminated some highly precise natural language alternatives.</td>
<td>Secondly, the cut-off point of 5% for identifying ‘high frequency’ terms was chosen arbitrarily. Whilst it appears reasonable, it may have inadvertently eliminated some highly specific natural language alternatives.</td>
</tr>
</tbody>
</table>
Both reviewers have expressed concern that the supplementary search strategy only retrieves 18.8% of the ‘lost set’ of citations and that this is evidence of low sensitivity in the non-indexed subset. It was not a stated objective of this study to develop a highly sensitive search strategy for retrieving non-indexed literature. This is not desirable, or even possible, for several reasons:

1. High sensitivity would inevitably come at a loss of search precision in the non-indexed set. While the non-indexed portion of results is usually much smaller than the indexed portion, the non-indexed citations appear at the top of the list of results. When a large proportion of irrelevant citations occupy the first few pages of results, searchers quickly lose confidence in the efficacy of the search itself.

2. The non-indexed search strategy can be viewed as an interim measure for identifying new literature in advance of MeSH indexing. Once these citations have been indexed with MeSH terms, they become accessible to the validated portion of the search filter.

3. As noted in the results section, very few natural language terms identified by frequency analysis of the lost set could retrieve more than 5% of the lost set citations. The long list of terms with a frequency of 5% or less is evidence of the diffuse nature of natural language and the difficulties associated with improving recall in this set to more than 18.8%.

To make it clearer to readers that the emphasis of this study is on what is gained rather than what is failed to be found, we have added an additional objective, plus a clarifying final paragraph, to the article at the end of background section (p. 9):

Objective 3. [to] use these high frequency textwords to create a textword-only search strategy that identifies a contained, rather than comprehensive, set of relevant non-indexed citations.

This resultant textword search string provides an interim strategy for identifying a subset of relevant citations which are inaccessible to the validated search filter prior to subject indexing. As such, the methodology presented here may constitute an incremental expansion of existing search filter translation methodology.

In future projects we intend to explore several strategies for improving recall/sensitivity in the lost set and their consequential effect on precision (measured post-hoc, outside the test set). For example, rather than rely on search phrases with term adjacency (e.g. ‘left ventricular systolic dysfunction,’ ‘left ventricular diastolic dysfunction’, and ‘LV dysfunction’), an improvement in recall may be achieved by not insisting on adjacency, e.g. ((systolic OR diastolic OR LV) AND left AND ventricular AND dysfunction). A paragraph acknowledging that this may be one way of improving search recall has been added to the Discussion section (methodology assessment, para. 5):

A future study might also investigate the effect of including search statements that incorporate the AND Boolean operator in order to increase retrieval in the Lost Set. We only included phrase constructs in our search strategy which imposes an adjacency condition on search terms, e.g. ‘left ventricular systolic dysfunction’ or ‘LV dysfunction’. The AND operator might serve to broaden the search without too great a cost to search precision, e.g. Left AND (systolic OR diastolic OR LV) AND ventricular AND dysfunction. The use of truncation may further improve retrieval (e.g. ventric* retrieves on ventricular, ventricle, and ventricles).
3. The authors should include a clear notice (a ‘health warning’) to users considering using the filter that 80% of relevant records are not retrieved by the filter and that by using other terms (which the authors can probably suggest based on their frequency analysis) the sensitivity is likely to be much improved (with an inevitable loss in precision).

Authors’ response: As reported in our previous study (Damarell 2011), the validated OvidSP Medline heart failure search filter and its PubMed translation have a high level of recall/sensitivity (between 97.8% and 98.2%, as demonstrated in two validation sets). A notice or warning attached to the PubMed heart failure search filter would only apply to the supplementary search strategy which accompanies the validated search filter, and not the validated search filter itself. As this study shows, when the original four-term search filter was restricted to searching on title and abstract alone, it only failed to retrieve 12% of the literature it previously retrieved. An additional five terms reclaimed 18.8% of this 12%. These are not big numbers considering we are only concerned with the smaller, non-indexed subset of PubMed. As stated in the background section of the article, at any point in time, this subset comprises 2% of PubMed’s total citations. Importantly, this loss is largely temporary. The highly sensitive heart failure search filter will be capable of retrieving a large proportion of these previously irretrievable citations once MeSH indexing takes place.

To avoid the potential misconception that this low sensitivity relates to the substantive ‘search filter,’ rather than the additional ‘textword-only search strategy,’ we have made the validated search filter’s recall/sensitivity more explicit in the Background section (Why translate for PubMed?):

To support clinicians providing end-of-life care to heart failure patients, researchers with the CareSearch project recently developed and validated a heart failure search filter for use in OvidSP Medline[7]. This search filter demonstrates 98% sensitivity and 75% precision within this database.

To highlight the point that this low sensitivity is only a temporary condition, pre-indexing, we have also added the following sentence to the objectives section of the Background:

This resultant textword search string provides an interim strategy for identifying a subset of relevant citations which are inaccessible to the validated search filter prior to subject indexing.

Minor essential revisions

1. The authors should consider their use of adjectives throughout the paper and whether they are justified by the results of the filter, for example: In the abstract conclusions the authors state their PubMed search filter provides ‘comprehensive access to emerging heart failure evidence’. I do not think sensitivity of less than 20% merits the use of the word ‘comprehensive’. The conclusions section of the abstract also uses the term ‘optimal’, which seems unsupported with the sensitivity provided and without any data on precision being provided.

Authors’ response: We agree with the reviewer’s points. While our method acknowledges the PubMed system in its entirety, the overall search filter itself (validated and supplementary components working together) can be said to provide ‘extended’ rather than ‘comprehensive’ access to the heart failure literature. The use of the word ‘optimal’ also relates to ‘extended’ coverage rather than search sensitivity/precision.
In light of these comments, we have therefore made the following changes:

<table>
<thead>
<tr>
<th>Location</th>
<th>Original sentence</th>
<th>Replacement sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract methods (p.2)</td>
<td>This string, restricted to searching on PubMed’s unique content, was then combined with the validated PubMed translation to create a comprehensive PubMed heart failure filter.</td>
<td>This string, restricted to searching on PubMed’s unique content, was then combined with the validated PubMed translation to extend the filter’s performance in this database.</td>
</tr>
<tr>
<td>Abstract conclusions (p.3)</td>
<td>A validated PubMed search filter can be used to develop a supplementary textword-only search strategy for optimal retrieval of PubMed’s unique content.</td>
<td>A validated PubMed search filter can be used to develop a supplementary textword-only search strategy to extend retrieval to PubMed’s unique content.</td>
</tr>
<tr>
<td>Background: Accurate OvidSP Medline to PubMed translation (p. 6)</td>
<td>This strategy, restricted to the non-indexed subset, can then be combined with the translated component for optimal retrieval across the full PubMed database.</td>
<td>This strategy, restricted to the non-indexed subset, can then be combined with the translated component for extended retrieval across the full PubMed database. (Moved to Background: Retrieving PubMed’s unique content, p. 7)</td>
</tr>
<tr>
<td>Method: Phase 4 (p.11)</td>
<td>The Lost Set was then used to identify natural language terms that might serve as supplementary search filter terms, optimising search filter performance across the full PubMed database.</td>
<td>The Lost Set was then used to identify natural language terms that might serve as supplementary search filter terms, extending search filter performance across the full PubMed database.</td>
</tr>
<tr>
<td>Method: Phase 5 (p.13)</td>
<td>This strategy, limited to searching PubMed’s unique content by the addition of ‘NOT Medline[sb]’, is combined with the validated PubMed translation to form a comprehensive PubMed search filter for heart failure.</td>
<td>This strategy, limited to searching PubMed’s unique content by the addition of ‘NOT Medline[sb]’, is combined with the validated PubMed translation to extend the filter’s retrieval of the heart failure literature.</td>
</tr>
</tbody>
</table>
2. The authors should define key terms early in the paper, e.g.

- sensitivity, precision.
- ‘high level evidence’ (background)
- significantly

Authors’ response:
- Sensitivity was defined at the start of the article in the background section (under heading: validated search filters). A definition of precision has now been included in the same section.
- What is meant by ‘high level evidence’ has been clarified in the opening paragraph of the background section by modifying the sentence to ‘... a burgeoning volume of high level evidence, specifically randomised controlled trials and systematic reviews[1]’.
- Significantly – this term has been used several times in the article without meaning to imply statistical significance. It has therefore been replaced within the following sentences without loss of meaning:
  - Frequency analysis of these citations identified only three terms significantly closely related to the concept ...
  - Only a small number of these terms occurred with significantly relatively high frequency across the entire Lost Set to warrant consideration...
  - More significantly Furthermore, by far the largest proportion of this unique content comprises citations ...

3. The authors should be less categorical in their statements and use more cautious phrasing, for example using ‘might’, ‘may’, ‘can’. For example, on page 7 the authors suggest there are many outdated terms in MeSH and provide one example. In the absence of clear citable evidence, it might be more accurate to note that there can or well may be outdated terms rather than implying (without much more evidence) that there are many.

Authors’ response: The following sentence in the second-to-last paragraph of the background section has been modified to: ‘Similarly, articles employing new terminology to describe emerging concepts may, on occasion, continue to be indexed with imprecise, out-dated MeSH terms for a considerable period of time...’ A few other instances of categorical expression were identified and modified throughout the manuscript.

4. The authors cite their own paper – it would be more helpful to busy readers to provide a few sentences summarising the key messages of the paper rather than expecting the reader to obtain a further paper.

Authors’ response: The methodology for establishing equivalency between the OvidSP Medline heart failure filter and its PubMed translation is now described within the current study article (Background section: Accurate OvidSP Medline to PubMed translation). A few sentences describing the heart failure filter have also been added to the ‘Why translate for PubMed?’ section of the background.
5. On p8 (second paragraph) although there may not be many papers reporting on PubMed search filters, the development of filters using words from the titles and abstracts of records (and using frequency analysis) is core to much search filter development in the past 20 years (and is reported in the filter papers cited by the authors).

Authors’ response: The crux of our paper is the presentation of a systematic approach to OvidSP Medline-to-PubMed search filter translation which explicitly acknowledges the non-indexed subset. As frequency analysis is purely the means to this end, we have revised this sentence to change the emphasis: To the best of our knowledge, this systematic and explicit acknowledgement of PubMed’s non-indexed subset constitutes an innovative and incremental addition to existing search filter translation methodology.

6. The authors should show all of the strategies described on page 9, stage by stage, so that the reader can follow the development.

Authors’ response: Table 2 has been included which shows the creation of the Lost Set: study phases 1-3.

7. On p10 Concordance is a ‘program’ rather than a ‘programme’.

Authors’ response: ‘Programme’ has been changed to ‘program’.

8. On p15 (para 1) the authors discuss some publication types they regard as not yielding ‘high level research evidence’ (undefined). I acknowledge that, for busy searchers, these types of information may not be considered very useful, but I note that for other searchers such as systematic reviewers, letters and comments can have value because they sometimes are the source of results and can also provide clues to studies which have been conducted. It would be helpful to allude to this complexity in improving the precision of searching for some searchers.

Authors’ response: The reviewer makes a good point. The two sentences relating to this point have been removed from the manuscript.

Reviewer 2

1. I am not sure whether the title is correct. As I understand the manuscript well it is on identify non-indexed abstracts in Pubmed and not translating a search filter

Authors’ response: We acknowledge that the original study title was misleading. The title has now been changed to: OvidSP Medline-to-PubMed search filter translation: a methodology for extending search filter range to include PubMed’s unique content

2. In the Methods section of the Abstract it says that the PubMed version consists of three MeSH terms, I can only find one in Table 1. This seems not to be correct. Please change.

Authors’ response: The search filter includes the MeSH term Ventricular dysfunction, Left. However, both Heart failure and Cardiomyopathy exist as MeSH terms in the MeSH thesaurus and are included in the search filter as MeSH terms by virtue of the [tw] search tag. The [tw] search tag forces a search on title/abstract and MeSH search fields. The [tw] search tag was intentionally assigned to these two terms as they constituted high frequency, high performing MeSH terms as well as title/abstract terms in the original search filter development study.
The fourth term, *left ventricular ejection fraction*, however, does not search as a MeSH term as it does not exist as a MeSH term in the MeSH thesaurus. It is only searched in the title and abstracts fields. The [tw] tag is still applied to it here for the sake of consistency, although it would be more appropriate/accurate to apply the [tiab] search tag.

A note has been added to Table 1 to explain the [tw] search tag and its relationship to MeSH term searching.

3. **With the "translation" you miss 81.2% of the abstracts (Results section). So this translation is not good enough, and should be improved.**

Authors’ response: The translated heart failure search filter itself does not miss 81.2% of abstracts. As reported in a previous study and referenced in this paper, the heart failure search filter has a high sensitivity of 98% in the indexed subset of PubMed (this fact has now been added to the background section).

This study is about developing a textword-only version of the validated search filter which will be appended to it for the purpose of extending retrieval to PubMed’s non-indexed subset (which comprises only 2% of total PubMed citations). We can find no existing studies that provide a reproducible methodology for systematically extending search filter retrieval to this subset. Therefore, this study arguably makes an incremental improvement to existing search filter translation methodology.

If search filters are not routinely accounting for this unique content in their development, this study demonstrates an 18.8% retrieval gain in the non-indexed subset, rather than a failure to retrieve 81.2% of citations. Furthermore, as stated in the objectives section, most of the 81.2% of citations presently not retrieved by the supplementary search strategy will eventually be retrieved by the validated search filter once MeSH indexing is complete. The loss is therefore not critical as it constitutes a temporary situation.

An additional sentence emphasising this gain, rather than loss in retrieval has been added to the objectives part of the Background section: *This resultant textword search string provides an interim strategy for identifying a subset of relevant citations which are inaccessible to the validated search filter prior to subject indexing.*

4. **I do not agree with the Conclusions of the authors. One should not use only [tiab] in PubMed.**

Author’s response: In the final PubMed heart failure search strategy (table 4), the [tiab] tag is only attached to search terms in the supplementary textword-only search strategy. As the article details, the sole purpose of the supplementary textword-only search strategy is to extend retrieval to PubMed’s non-indexed subset. It is in fact restricted to this subset by the final search element “NOT Medline[sb]”. As citations in the non-indexed subset have no MeSH terms assigned to them, the [tiab] is appropriate in this context. No other search tag can improve retrieval in this set where there are only titles and abstracts to search on.

5. **Nowadays PubMed has up to 22 million citations. Please correct.**

Authors’ response: The article was referring to the 19 million citations in the NLM’s Medline database (now increased to 20 million, according to the NLM website), rather than PubMed’s 22 million citations. It is precisely this underlying difference in size, and the need to account for
PubMed’s additional content that makes it important to adapt search filters translated from OvidSP Medline.

To make it clear that Medline is being referred to in this paragraph; the additional PubMed figure has also been provided in the manuscript.

6. Page 6, section Timelines of evidence. the authors state that PubMed also consists of book chapters etc. For evidence these items are less important, as books can be out of date. Please skip this. It does not fit with the arguments given in Section Why translate.

Authors’ response: We agree that the paragraph is redundant. It has been removed.

7. Page 8. Issue1 in Methods section. Here the authors state that they identified citations in a closed dataset. This suggest a kind of gold standard, while only a set of retrieved abstracts is meant. Please rewrite.

Authors’ response: We acknowledge that this could be interpreted this way. The term ‘closed set’ has been removed without any associated loss in meaning.

8. It is for me not clear why the authors apply the Medline subset limit (page 9). When I repeated the search of the authors, adding of the Medline subset to the search did not give a difference in identified citations. Next to that one can assume that citations from 2010 are indexed.

Authors’ response: It was important to restrict the search to citations indexed with MeSH terms (i.e. the Medline subset) to demonstrate the effect on retrieval when the search terms previously able to retrieve on MeSH terms can no longer retrieve based on MeSH term availability (the difference between phase 1 and phase 2 of the study). At the time we ran our study (21 April 2011), a large number of 2010 citations had still not been indexed. One may assume from your replicated search that these citations have all now been indexed.

9. Page 12 No search date is given. Repeating the search give different data now. May be they can give a complete search history.

Authors’ response: the search date (21 April 2011) had been inadvertently removed from the latest draft. It has now been reinstated in the methods section, and as part of the table title showing the search strategy itself within PubMed (table 2).

10. I am curious why no truncation is used, and why the ATM of Pubmed is not mentioned in the manuscript. And as said before Medline[sb] does not do anything.

Authors’ response: The methodology had not intentionally overlooked or excluded truncation as a way of improving search sensitivity. It would have been employed if any of the candidate search terms lent themselves to it. Plural forms and terms with variant endings were not identified in the course of our frequency analysis. However, as truncation may be relevant to future applications of this methodology, we have referred to it as a potential tool for improving search retrieval in the Discussion (methodology assessment, para. 5): ‘The use of truncation may further improve retrieval (e.g. ventric* retrieves on ventricular, ventricle, and ventricles).’

ATM was not relevant to this study as the objective was to identify high frequency title and abstract terms for building a search strategy for the non-indexed subset of PubMed. As ATM relates to MeSH term mapping, it is not a relevant consideration in this context.
11. Page 13. When I understand the authors well they miss 81.2% of the Lost Set Citations. This is too much. Therefore their “translation” is not good enough. As written above no truncation is used (or wildcards). Using only [tiab] is not good as you will miss the ATM possibility of PubMed.

It was not a stated objective of this study to develop a highly sensitive search strategy for retrieving non-indexed literature. This is not desirable or even possible for several reasons:

1. High sensitivity would inevitably come at a loss of search precision in the non-indexed set. While the non-indexed portion of results is usually much smaller than the indexed portion, the non-indexed citations appear at the top of the list of results. When a large proportion of irrelevant citations occupy the first few pages of results, searchers quickly lose confidence in the efficacy of the search itself.

2. The non-indexed search strategy can be viewed as an interim measure for identifying new literature in advance of MeSH indexing. Once these citations have been indexed with MeSH terms, they become accessible to the validated portion of the search filter.

3. As noted in the results section, very few natural language terms identified by frequency analysis of the lost set could retrieve more than 5% of the lost set citations. The long list of terms with a frequency of 5% or less is evidence of the diffuse nature of natural language and the difficulties associated with improving recall in this set to more than 18.8%.

As stated earlier, the emphasis of this study is on what is gained rather than what is failed to be found. Therefore, an additional objective plus a clarifying final paragraph have been added to the end of background section:

Objective 3. [to] use these high frequency textwords to create a textword-only search strategy that identifies a contained, rather than comprehensive, set of relevant non-indexed citations.

This resultant textword search string provides an interim strategy for identifying a subset of relevant citations which are inaccessible to the validated search filter prior to subject indexing.

The points about truncation and ATM have been addressed above.

12. In the light of above remarks (point 10) I do not agree with the authors in the Discussion section. It is nice to identify 5 new terms for the heart failure strategy, but you miss 81.2% of the data. And this is due to using [tiab] in stead of [tw]. My advise is never use [tiab] alone in a search as you may miss citations.

Authors’ response: The inability to retrieve the 81.2% was due to the fact that natural language is widely variable and lacks the precision and specificity of MeSH terms. A very long search strategy comprising a much larger number of diffuse textwords would not have significantly improved retrieval much above the 18.8% level. Furthermore, a longer search string would have the potential to reduce search precision outside the test set to a remarkable degree.

The [tiab] tag is the only search tag appropriate if creating a search strategy designed to retrieve PubMed’s non-indexed content. The [tw] tag will not improve sensitivity of retrieval in any way in this subset as there is no subject heading (MeSH) field to retrieve on, which is the major difference between the [tw] and [tiab] tags. We agree with the reviewer that the [tiab] tag would not generally be employed in a PubMed search designed to search the whole of PubMed. However, the use of
[tiab] in the methodology to stimulate the non-indexed subset of PubMed, was a crucial aspect of our methodology.

13. When I understand the authors well the new search filter as described in Table 3 should give 41+5995=6036 citations. I have repeated the search in PubMed and identified more citations than the original search filter as described in Table 1. So there is something wrong or in the data or in the description. I attach the search history of what I have done. And this I miss in this manuscript, a flow chart with data, complete search history which can clarify a lot. Also I got other data than the authors (9772 and 9216 citations in the first and second dataset, with a difference of 556 citations).

Authors’ response: In duplicating the search, the reviewer has omitted several key steps in the methodology:

- **Ventricular dysfunction, Left** should have been inverted to **Left ventricular dysfunction** (Method, phase 1, p. 10)
- The [mh:noexp] tag attached to **Left ventricular dysfunction** needed to be converted to the [tiab] tag (Method, phase 2, p. 10). This is the crucial part of the method. It identifies the quantity of literature that is not retrievable by left ventricular dysfunction as a title/abstract search term prior to MeSH indexing. As the study results show, more than 50% of citations indexed with the Left ventricular dysfunction MeSH term do not include this same term in their title or abstract but have a range of ways of describing the same concept, eg. ‘LV dysfunction’, ‘Left ventricular systolic dysfunction’. By not converting the search tag, the reproduction has failed to retrieve the bulk of the lost literature.
- The replicated search did not apply the English language limit (Method, phase 1).
- Furthermore, the numbers will naturally have increased due to retrospective indexing.

We reran the search on 13 April 2011 and have reproduced the results here. These results show that although the numbers have changed, the proportions have stayed the same across the space of two years. The Lost Set still comprises 12% (1153/9330) of citations.

**Proof for creating the ‘Lost set’ of citations within PubMed (as described in phases 1-3 of method section). Run on 13 April 2013**

<table>
<thead>
<tr>
<th>Search</th>
<th>Query</th>
<th>Items found</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>#1 NOT #2</td>
<td>1153</td>
</tr>
</tbody>
</table>

We acknowledge that the search strategy must not be as clear as it needs to be in order to be reproducible and transparent. We have therefore included the original search strategy as Table 2 in the methods section.
12. Page 14, last four lines, there it is stated that only citations with an abstract were taken into account. I am not sure whether this is mentioned in the Methods Section. If not please adjust.

Authors’ response: This was mentioned in the Methods sections, phase 1. Here the full range of limits applied to the search are outlined (i.e. citations with abstracts, English language, publication date of 2010, and Medline indexed subset only).

13. I do not understand the translation from the search strategy mentioned in table 1 and the one from table 3. I tried the strategy from Table 3 and identified more citations than the one from Table 1 (9811), and not less. And what I have written before why not using truncation. May be I miss something or do not understand issues. If so, than the manuscript is not clear enough.

Authors’ response: The full and final PubMed heart failure search strategy provided in table 3 (now table 4) will inevitably retrieve more citations across PubMed as this search uses the OR operator to combine the validated PubMed search translation (table 1) with the supplementary textword-only search strategy. As this paper has stressed, the validated PubMed search strategy from table 1 retrieves well from the indexed subset of PubMed but needs the addition of the 9-term textword-only component to extend its range into the non-indexed set.

14. Table 2. Please change "own subset" in something else. I did not understand what was meant. May be "original dataset". Or at least clarify this subset.

Authors’ response: “Own subset” has been changed to “individual subsets” on table 2 (now table 3) in the hope that this improves clarity.

15. All in all a nice piece of work, but I am not content with the results as described in the Results and Discussion. It is good to have clarified why citations are not identified when using a different search strategy, but the consequences are interpreted as far as I can see, not in a right manner. I cannot repeat their methods and their outcome. And further I am not happy with the final proposed search strategy, see arguments above.