

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)	Words do matter: a systematic review on how different terminology for the same condition influences management preferences
AUTHORS	Nickel, Brooke; Barratt, Alexandra; Copp, Tessa; Moynihan, Ray; McCaffery, Kirsten

VERSION 1 - REVIEW

REVIEWER	Brian J. Zikmund-Fisher University of Michigan, USA
REVIEW RETURNED	17-Oct-2016

GENERAL COMMENTS	<p>This systematic review examines the extant (but nascent) literature on how verbal terms describing a condition influence people's desire for treatment. This question is clearly important, and the authors appropriately frame it in the context of overtreatment. Given that the authors only find studies that meet their inclusion criteria, the value provided by the review is somewhat limited. Nonetheless, I can find relatively little to complain about in its implementation, although I have little detailed knowledge to evaluate the specific methods of a systematic review.</p> <p>My main concern, somewhat ironically, involves the terminology used to represent the core finding of the review. The authors state "Overall this review found that when a more medicalised term is used to describe a condition." Yet, the meaning of "medicalised" is undefined in this paper, and no definition is clearly implied in the review of studies. Is "pre-invasive breast cancer cells" medicalized because it referenced cancer, or because it represented a more precise description of the abnormality referred to in the alternate term "abnormal cells"? Note that this is a different kind of tradeoff than that in the first Scherer article, which contrast a disease term vs. no term at all (i.e., simply a description of symptoms). The second Scherer article finds lower treatment interest with "eye infection" (a medical term to me) vs. "pink eye" which is quite specifically NOT a medical term. Even "hairline fracture" is not obviously less medical than "broken bone". My point is that the review of individual studies appears accurate, but the framing of these results as representing a medicalisation dimension is not, to my perspective, justified based on the studies reviewed. Is this high literacy terms? Greater precision? A disease label vs. a description of outcome? Terms familiar in medicine vs. not? Each of these interpretations suggests different patterns in other disease contexts. The authors should be careful in their interpretation of what these data show and potentially discuss as speculation possible mechanisms.</p> <p>Given the small sample of studies and inconsistent statistical</p>
-------------------------	--

	significance observed, I also have a problem with the use of the term "consistently" in the statement that "This review demonstrates that the terminology used to describe a condition consistently influences patient preferences for treatments."
--	---

REVIEWER	Bjørn Hofmann The Norwegian University of Science and Technology (NTNU) Gjøvik, Norway
REVIEW RETURNED	18-Jan-2017

GENERAL COMMENTS	<p>This systematic review demonstrates that different terminology used to describe the same condition can influence patient's treatment preferences and thereby underscores that language is a powerful tool that has the potential to influence patients' thoughts and actions. The authors also cautiously suggest that changing the terminology in low-risk conditions or conditions with indolent clinical course could be a potential communication strategy in order to shift assumptions that immediate invasive treatments are always needed, allow for better shared decision making between clinicians and patients, and the consideration of more conservative treatment options. The methods and materials are well described. The figure, tables and appendices are informative.</p> <p>Although there are only few identified studies in the field, the study is warranted to draw attention to an important issue.</p> <p>I only have one comment with regards to categorization in medicalised and non-medicalised terms. There may be other dimensions that are important for patients (or proxies), e.g., esthetical aspects. Including such aspects could explain the findings with regard to "pink-eye." This may be an aspect to consider for the discussion and for further research.</p> <p>As a note of professional interest it is surprising that the authors did not find any studies on colorectal cancer screening, where terminology could be crucial.</p>
-------------------------	--

REVIEWER	Teresa Gavaruzzi Department of Developmental Psychology and Socialization, University of Padova, Italy
REVIEW RETURNED	25-Jan-2017

GENERAL COMMENTS	<p>Very clear and well written paper.</p> <p>I recommend adding that all studies were hypothetical in the results section of the abstract.</p> <p>Additionally, as only 6 studies were identified, all hypothetical, I recommend being more cautious in the conclusion (e.g. "this review suggests" instead of "demonstrates").</p> <p>In the discussion, I would mention previous work investigating which characteristics of tumour diagnoses (e.g. current diagnosis vs. predisposition, malignant vs. benign vs. non-tumour) drive preferences for active treatment (Gavaruzzi et al., 2011 MDM). According to the results, it is the malignant nature of the current diagnosis that is key in driving preferences for active treatments over surveillance.</p> <p>Gavaruzzi, T., Lotto, L., Rumiati, R., & Fagerlin, A. (2011). What</p>
-------------------------	---

	<p>makes a tumor diagnosis a call to action? On the preference for action vs. inaction. Medical Decision Making, 31 (2), 237-244.</p> <p>I would also add some considerations on how feasible it would be in practice to change the terminology for low-risk, screen detected conditions. For example, if the less or non-medicalised term was used by health professionals when talking to patients, but the medical records/documents use the medical term, this could backfire. Indeed, as shown in McCaffery et al., 2015 study, this could increase concern, worry and preference for invasive treatment.</p>
--	--

REVIEWER	<p>Adrian Barnett Queensland University of Technology Australia</p> <p>I am on a grant application with Kirsten McCaffery, but I feel I was able to review the paper fairly.</p>
REVIEW RETURNED	22-Feb-2017

GENERAL COMMENTS	<p>This was an interesting review in a relatively new area of research (all the papers found were recent). The review was pre-registered and appeared to be well conducted. The paper was clearly written and the results well reported.</p> <p>At a number of points in the text and tables there was a dichotomy between results that were and were not statistically significant. This is often not a useful dichotomy and the more interesting difference here may be the clinical difference. For example, even a relatively small increase of 5% more respondents preferring surgery could translate into many more surgeries when scaled up to a country.</p> <p>Minor comment</p> <ul style="list-style-type: none"> - page 10, multivariate means multiple dependent variables (e.g., jointly modelling diastolic and systolic blood pressure), the correct term here is "multiple variable logistic regression" - Table 4, final row. I'm not sure the comment about multiple comparisons is needed given that all the p-values were small and unlikely to change much with any adjustment.
-------------------------	---

REVIEWER	<p>Dongfeng Wu School of Public Health and Information Sciences University of Louisville, USA</p>
REVIEW RETURNED	14-Mar-2017

GENERAL COMMENTS	<p>From the letter of the Editor, "I invite you to review this manuscript with a particular emphasis on the statistical methods and analyses used." But I found that there was no statistical method, nor analysis ever carried out in this review article. It listed 6 publications to support their main argument: "the terminology used to describe a condition consistently influence patient preferences for treatments and related outcomes. However, these 6 published work don't quite agree with each other and the studies did not focus in the same disease. One of the study (McCaffery et al 2015) found NO significant differences in treatment preferences, while the other 5 studies got an opposite conclusion. Among the 6 studies, 3 was on</p>
-------------------------	--

	ductal carcinoma in situ, 1 on gastroesophageal reflux disease, 1 on conjunctivitis, 1 on bony fracture. I would suggest that the authors should focus on one disease, instead of spreading out over different diseases. And if there were contradicted results in the studies, a meta-analysis is necessary for a review article. Based on these findings, I would suggest to have a major revision before it can be accepted.
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

This systematic review examines the extant (but nascent) literature on how verbal terms describing a condition influence people's desire for treatment. This question is clearly important, and the authors appropriately frame it in the context of overtreatment. Given that the authors only find studies that meet their inclusion criteria, the value provided by the review is somewhat limited. Nonetheless, I can find relatively little to complain about in its implementation, although I have little detailed knowledge to evaluate the specific methods of a systematic review.

My main concern, somewhat ironically, involves the terminology used to represent the core finding of the review. The authors state "Overall this review found that when a more medicalised term is used to describe a condition." Yet, the meaning of "medicalised" is undefined in this paper, and no definition is clearly implied in the review of studies. Is "pre-invasive breast cancer cells" medicalized because it referenced cancer, or because it represented a more precise description of the abnormality referred to in the alternate term "abnormal cells"? Note that this is a different kind of tradeoff than that in the first Scherer article, which contrast a disease term vs. no term at all (i.e., simply a description of symptoms). The second Scherer article finds lower treatment interest with "eye infection" (a medical term to me) vs. "pink eye" which is quite specifically NOT a medical term. Even "hairline fracture" is not obviously less medical than "broken bone". My point is that the review of individual studies appears accurate, but the framing of these results as representing a medicalisation dimension is not, to my perspective, justified based on the studies reviewed. Is this high literacy terms? Greater precision? A disease label vs. a description of outcome? Terms familiar in medicine vs. not? Each of these interpretations suggests different patterns in other disease contexts. The authors should be careful in their interpretation of what these data show and potentially discuss as speculation possible mechanisms.

Thank you for your comments on our paper. We agree that the meaning of medicalised needs to be more clearly defined in the paper. We have now added in a sentence in the 'Synthesis of results' section in the results to help define this. Please see page 14.

"In this analysis medicalised or precise terminology refers to language that describes the condition either using medical terms that healthcare professionals commonly used or that described the condition in a more specific way (when compared to the comparator term)."

We have also added in the word 'precise' into the abstract and throughout the manuscript and have changed the headings in Table 5 to again make this clearer.

While we agree that the different terminologies included in the studies in this review may have different mechanisms driving treatment preferences, we believe that the classification of the terminology is important to understand what links preferences and that terminology that is more medicalised or more precise (as now defined in the paper) is an important driver of this. In saying this, we have tried our best to clarify this in the limitations section of the discussion and have softened the conclusion in relation to this comment. We have also revisited the categorisation of the Azam paper (on bony injuries) and have updated the results and the explanation of this categorisation in the

limitation section of the discussion. This did not affect the results. Please see page 17.

“The delineation between more medical/precise and less or non-medical/precise was challenging in some studies. For example, it seemed clear that ‘pre-invasive breast cancer cells’ was a more medicalised term than ‘abnormal cells’. In contrast, it was not as clear what it was exactly about the term ‘pink eye’ (a more precise term to describe the condition) that elicited stronger parental preferences for antibiotics when compared with the term ‘eye infection’. Other aspects important to parents in this study (e.g. aesthetical aspects), may have influenced management preferences. The author team therefore made explicit judgements about which terms were deemed more medicalised or precise and which were not, as well as what managements were considered invasive and what were considered non-invasive. These decisions were guided from the aims and outcomes of the studies, and followed categories used in the original studies, with the exception of the Azam paper where authors were guided by information on the precise medical terminology healthcare professionals use to describe a bony injury (including ‘broken bone’, ‘fracture’, ‘greenstick fracture’, ‘hairline fracture’) (see Appendix 4).”

Given the small sample of studies and inconsistent statistical significance observed, I also have a problem with the use of the term "consistently" in the statement that "This review demonstrates that the terminology used to describe a condition consistently influences patient preferences for treatments."

We agree and have removed the term “consistently” from the statement above in the manuscript. Please see page 18.

“This review suggests that the terminology used to describe a condition can influence patient preferences for management and related outcomes.”

Reviewer: 2

This systematic review demonstrates that different terminology used to describe the same condition can influence patient’s treatment preferences and thereby underscores that language is a powerful tool that has the potential to influence patients’ thoughts and actions.

The authors also cautiously suggest that changing the terminology in low-risk conditions or conditions with indolent clinical course could be a potential communication strategy in order to shift assumptions that immediate invasive treatments are always needed, allow for better shared decision making between clinicians and patients, and the consideration of more conservative treatment options. The methods and materials are well described. The figure, tables and appendices are informative. Although there are only few identified studies in the field, the study is warranted to draw attention to an important issue.

I only have one comment with regards to categorization in medicalised and non-medicalised terms. There may be other dimensions that are important for patients (or proxies), e.g., esthetical aspects. Including such aspects could explain the findings with regard to “pink-eye.” This may be an aspect to consider for the discussion and for further research.

We agree with your comment on the categorization of the pink-eye study and that there may be other dimensions that are important for patients to consider. We have tried to further clarify the categorization issue throughout the manuscript based on your comment and Reviewer 1’s related comment, and have also tried to better describe this issue in the limitations section in the discussion and in the conclusion. Please see changes on page 17.

“The delineation between more medical/precise and less or non-medical/precise was challenging in some studies. For example, it seemed clear that ‘pre-invasive breast cancer cells’ was a more

medicalised term than 'abnormal cells'. In contrast, it was not as clear what it was exactly about the term 'pink eye' (a more precise term to describe the condition) that elicited stronger parental preferences for antibiotics when compared with the term 'eye infection'. Other aspects important to parents in this study (e.g. aesthetical aspects), may have influenced management preferences. The author team therefore made explicit judgements about which terms were deemed more medicalised or precise and which were not, as well as what managements were considered invasive and what were considered non-invasive. These decisions were guided from the aims and outcomes of the studies, and followed categories used in the original studies..."

As a note of professional interest it is surprising that the authors did not find any studies on colorectal cancer screening, where terminology could be crucial.

Although the terminology in colorectal cancer screening could be crucial to screening and/or treatment decisions, we did not find any studies on colorectal cancer terminology which met the inclusion criteria for our study.

Reviewer: 3

Very clear and well written paper.

I recommend adding that all studies were hypothetical in the results section of the abstract.

We have now added that all studies included hypothetical scenarios into the results section of the abstract. Please see page 2.

"Of the 1399 titles identified, 7 studies, all of which included hypothetical scenarios, met the inclusion criteria."

Additionally, as only 6 studies were identified, all hypothetical, I recommend being more cautious in the conclusion (e.g. "this review suggests" instead of "demonstrates").

We have now changed the term "demonstrates" to be "suggests" in the conclusion. Please see page 18.

"This review suggests that the terminology used to describe a condition can influence patient preferences for management and related outcomes."

In the discussion, I would mention previous work investigating which characteristics of tumour diagnoses (e.g. current diagnosis vs. predisposition, malignant vs. benign vs. non-tumour) drive preferences for active treatment (Gavaruzzi et al., 2011 MDM). According to the results, it is the malignant nature of the current diagnosis that is key in driving preferences for active treatments over surveillance. Gavaruzzi, T., Lotto, L., Rumiati, R., & Fagerlin, A. (2011). What makes a tumor diagnosis a call to action? On the preference for action vs. inaction. *Medical Decision Making*, 31 (2), 237-244.

Thank you for the suggestion. We have now added the above referenced study into the discussion section of our manuscript. Please see page 15.

"It has also been shown that the use of interpretive terminology (e.g. including the words positive or negative, or using a metaphor) 23 24, the terminology used to describe a treatment choice²⁵, describing a condition with plain language terminology as compared to jargon²⁶ and the severity of the characteristics of the diagnosis²⁷ can have an influence on medical decision making."

I would also add some considerations on how feasible it would be in practice to change the

terminology for low-risk, screen detected conditions. For example, if the less or non-medicalised term was used by health professionals when talking to patients, but the medical records/documents use the medical term, this could backfire. Indeed, as shown in McCaffery et al., 2015 study, this could increase concern, worry and preference for invasive treatment.

Thank you again for the suggestion. We have now discussed that changing terminology would be difficult in practice and would take a systems approach at all levels, however, we believe could still be feasible as demonstrated by a recent change to the terminology of non-invasive encapsulated follicular variant of papillary thyroid carcinoma (EFVPTC). This terminology was changed to highlight the true nature of the tumour, lessen the emotional and psychological burden associated with the term "cancer" and potential reduce unnecessary overtreatment. Please see page 16.

"Changing the terminology of low-risk conditions may be difficult in practice as a systems level approach would need to be taken to ensure that all healthcare professionals implemented the new terminology. Although, it would be feasible as demonstrated by the recent change to the terminology of the "non-invasive encapsulated follicular variant of papillary thyroid carcinoma" (EFVPTC) to be "non-invasive follicular thyroid neoplasm with papillary-like nuclear features" (NIFTP) in order to highlight the true nature of the tumour, lessen the emotional and psychological burden associated with the term "cancer" and potentially reduce overtreatment^{33 34}."

Reviewer: 4

This was an interesting review in a relatively new area of research (all the papers found were recent). The review was pre-registered and appeared to be well conducted. The paper was clearly written and the results well reported.

At a number of points in the text and tables there was a dichotomy between results that were and were not statistically significant. This is often not a useful dichotomy and the more interesting difference here may be the clinical difference. For example, even a relatively small increase of 5% more respondents preferring surgery could translate into many more surgeries when scaled up to a country.

Thank you very much for raising this important point. We agree and have now added this into the first paragraph discussion section of the manuscript (please see page 14-15) and have added in the percentage difference information into a column in Table 5.

"Although not all of the studies included in our review had results which were statistically significant in relation to preferences for more invasive managements, at a population level these trends may represent a clinically important difference. For example, a relatively small increase in the number of people preferring surgery in these studies could translate into significantly more surgeries across a larger population."

Minor comment

- page 10, multivariate means multiple dependent variables (e.g., jointly modelling diastolic and systolic blood pressure), the correct term here is "multiple variable logistic regression"

Thank you. We have now changed to the correct term. Please see page 12.

"Although women with a previous history of cancer (other than breast cancer) and women with high socioeconomic status more frequently chose surgery in univariate analyses, high numeracy was the only independent predictor of preference for surgical treatment in the multiple variable logistic regression model for all three terms: cancer (OR 2.11, 1.34-3.34 CI, p=0.001), lesion (OR 1.96, 1.20-3.19, p=0.001), abnormal cells (OR 1.63, 1.01-2.67, p=0.048)."

- Table 4, final row. I'm not sure the comment about multiple comparisons is needed given that all the p-values were small and unlikely to change much with any adjustment.

Thank you. We have now removed the comment about multiple comparisons in Table 4.

Reviewer: 5

From the letter of the Editor, "I invite you to review this manuscript with a particular emphasis on the statistical methods and analyses used." But I found that there was no statistical method, nor analysis ever carried out in this review article. It listed 6 publications to support their main argument: "the terminology used to describe a condition consistently influence patient preferences for treatments and related outcomes. However, these 6 published work don't quite agree with each other and the studies did not focus in the same disease. One of the study (McCaffery et al 2015) found NO significant differences in treatment preferences, while the other 5 studies got an opposite conclusion. Among the 6 studies, 3 was on ductal carcinoma in situ, 1 on gastroesophageal reflux disease, 1 on conjunctivitis, 1 on bony fracture. I would suggest that the authors should focus on one disease, instead of spreading out over different diseases. And if there were contradicted results in the studies, a meta-analysis is necessary for a review article. Based on these findings, I would suggest to have a major revision before it can be accepted.

Thank you for taking the time to review the paper. The aim of the review was to synthesise existing studies on terminology and its impact on treatment decision making across all conditions. We chose to focus on all medical conditions since we were interested in understanding the general implications for how terminology influences people's desire for medical management, which has broader implications for overtreatment and overuse. Focusing on only one condition would not allow us to understand the heterogeneity of this question. Furthermore, the literature in this area is scant, as demonstrated by the small number of studies included in the review. In the strengths and limitations section, the methods section and the limitations section of the discussion, we state that the heterogeneity of the studies and their respective outcome measures did not support the pooling of results. Please see support text below:

"Strengths and Limitations of this Study:

- Due to the variability of terms and outcomes assessed, authors were unable to conduct a meta-analysis and pool the effects of the data" (pg.4)

"Results from the studies were synthesised in a narrative form, as the heterogeneity of the studies and their respective outcome measures did not support pooling of results¹³." (pg.8)

"Due to the variability of the terms and outcomes assessed in the included studies, authors were unable to conduct a meta-analysis and pool the effects of the data, and therefore a definite synthesis of results of all studies was not possible." (pg.16-17)

VERSION 2 – REVIEW

REVIEWER	Brian Zikmund-Fisher University of Michigan, United States
REVIEW RETURNED	02-May-2017

GENERAL COMMENTS	The authors did a nice job of responding to the comments from the multiple reviewers. I have no further concerns.
-------------------------	---

REVIEWER	Teresa Gavaruzzi
-----------------	------------------

	DPSS, University of Padova, Italy
REVIEW RETURNED	11-May-2017

GENERAL COMMENTS	<p>Second part of first paragraph of introduction: this is what you aim to show (see next to the last two sentences of introduction).</p> <p>p.5 line 43-45 well understood by whom? patients? please specify</p> <p>Table 3 is listed in the text before Table 2 (page 8 line 5)</p>
-------------------------	---

REVIEWER	<p>Adrian Barnett Queensland University of Technology</p> <p>I was on an unsuccessful grant application in 2016/17 with authors Moynihan and McCaffery.</p>
REVIEW RETURNED	02-May-2017

GENERAL COMMENTS	<p>My few comments from the previous review have been well answered.</p> <p>Minor comment - Table 2, typo "269 health women" should be "healthy"</p>
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

The authors did a nice job of responding to the comments from the multiple reviewers. I have no further concerns.

Thank you.

Reviewer: 3

Second part of first paragraph of introduction: this is what you aim to show (see next to the last two sentences of introduction).

Thank you for your comment. We agree that there is some overlap in the introduction and have now removed the last sentence of the first paragraph of the introduction. Please see page 5.

“Medical encounters can be challenging and confronting for patients, especially when they are faced with a management decision. Clinical communication and language is an important aspect of a medical encounter as it influences patients’ understanding of their diagnosis and management options[1,2]. Decisions about treatments or tests may be influenced by various factors including the medical terminology clinicians use to diagnose and describe conditions to patients.”

Reviewer: 4

My few comments from the previous review have been well answered.

Minor comment

- Table 2, typo "269 health women" should be "healthy"

Thank you for pointing out this typo in table 2. We have now changed it to “healthy”.

p.5 line 43-45 well understood by whom? patients? please specify

To make this point clearer we have now indicated that it is well accepted by cancer experts, researchers and clinicians. Please see page 5.

“The term ‘cancer’ is understandably frightening for people to hear and can influence their thought and action[8], but it is now well accepted by cancer experts, researchers and clinicians that a range of conditions which include indolent to fast-growing lesions are labelled as cancer[9].”

Table 3 is listed in the text before Table 2 (page 8 line 5)

Thank you for pointing this out. We have now removed the sentence with reference to Table 3 in the Methods section. This sentence discusses the results of the risk of bias assessment and should be only be included in the Results section and not be included in both sections. Please see deleted text on page 8.