BMJ Open

Words do matter: a systematic review on how different terminology for the same condition influences management preferences

Brooke Nickel,[1] Alexandra Barratt,[1] Tessa Copp,[1] Ray Moynihan,[1,2] Kirsten McCaffery[1]

To cite: Nickel B, Barratt A, Copp T, *et al*. Words do matter: a systematic review on how different terminology for the same condition influences management preferences. *BMJ Open* 2017;**7**:e014129. doi:10.1136/bmjopen-2016-014129

► Prepublication history and additional material for this paper are available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2016-014129).

Received 3 September 2016
Revised 20 May 2017
Accepted 31 May 2017

[1]Wiser Healthcare, Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia
[2]Centre for Research in Evidence-Based Practice, Bond University, Gold Coast, Queensland, Australia

Correspondence to
Ms. Brooke Nickel; brooke.nickel@sydney.edu.au

ABSTRACT

Objectives Changing terminology for low-risk, screen-detected conditions has now been recommended by several expert groups in order to prevent overdiagnosis and reduce the associated harms of overtreatment. However, the effect of terminology on patients' preferences for management is not well understood. This review aims to synthesise existing studies on terminology and its impact on management decision making.

Design Systematic review.

Methods Studies were included that compared two or more terminologies to describe the same condition and measured the effect on treatment or management preferences and/or choices. Studies were identified via database searches from inception to April 2017, and from reference lists. Two authors evaluated the eligibility of studies with verification from the study team, extracted and crosschecked data, and assessed the risk of bias of included studies.

Results Of the 1399 titles identified, seven studies, all of which included hypothetical scenarios, met the inclusion criteria. Six studies were quantitative and one was qualitative. Six of the studies were of high quality. Studies covered a diverse range of conditions: ductal carcinoma in situ (3), gastro-oesophageal reflux disease (1), conjunctivitis (1), polycystic ovary syndrome (1) and a bony fracture (1). The terminologies compared in each study varied based on the condition assessed. Based on a narrative synthesis of the data, when a more medicalised or precise term was used to describe the condition, it generally resulted in a shift in preference towards more invasive managements, and/or higher ratings of anxiety and perceived severity of the condition.

Conclusions Different terminology given for the same condition influenced management preferences and psychological outcomes in a consistent pattern in these studies. Changing the terminology may be one strategy to reduce patient preferences for aggressive management responses to low-risk conditions.

Trial registration number PROSPERO: CRD42016035643.

Strengths and limitations of this study

- This is the first systematic review to synthesise the evidence on how different terminology given for the same condition impacts patients' management preferences.
- Only a small number of studies have examined this research question and were included in the review.
- Due to the variability of terms and outcomes assessed, authors were unable to conduct a meta-analysis and pool the effects of the data.
- All studies included were hypothetical; therefore patients facing a real diagnosis may respond differently.

INTRODUCTION

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 communication factors including the medical terminology clinicians use to diagnose and describe conditions to patients.

Overdiagnosis of several medical conditions and associated overtreatment is now widely accepted[3 4] and can have serious implications for patients, healthcare systems and society.[5 6] Numerous approaches are beginning to be proposed to help combat overdiagnosis and overtreatment, including various communication strategies.[7] Changing the terminology for medical conditions may be one communication strategy to mitigate the effect of overdiagnosis and overtreatment as it has the potential to influence the effect of diagnosis and labelling of a condition, and influence patients' decision making about management. It may encourage both patients and clinicians to more carefully consider conservative management options.

In particular, cancer terminology is one area where use of different terminology may greatly influence management decision making. 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] Lesions with low malignant potential are common (such as low-risk ductal carcinoma in situ (DCIS), low-risk papillary thyroid cancer and low-grade prostate cancer), and with the advent and increasing use of various screening technologies, these indolent lesions and their precursors are now frequently clinically identified. This identification can lead to the condition being overdiagnosed and in turn overtreated.

Changing the terminology for these low-risk, screen-detected conditions has now been recommended by several expert groups—including a National Institutes of Health state-of-the-science conference panel and a National Cancer Institute working group—in order to prevent overdiagnosis and encourage more consideration of less invasive management options such as active surveillance.[9–11] However, the effect of terminology on patients' willingness to accept more conservative management options is not well understood. This systematic review aims to synthesise existing studies on terminology for medical conditions and its impact on management decision making, and associated psychosocial outcomes.

METHODS

Protocol and registration

The review's protocol is registered with PROSPERO (an international prospective register of systematic reviews), registration number: CRD42016035643.

Review question

How do different terminologies given to the same condition influence treatment or management decision making?

Search strategy

A comprehensive list of search terms was developed (see online supplementary appendix 1) with consultation from an information specialist and a search of relevant databases (Medline, Pre-Medline, EMBASE, PyscINFO, Cinhal and PubMed) was conducted from inception to April 2017. The returned search results were screened by title and abstract independently by two researchers (BN and TC) for irrelevant articles, review papers, editorials or commentaries and duplicates. An eligibility checklist was developed (see online supplementary appendix 2) to guide the selection of appropriate studies. Decisions regarding inclusion and exclusion of studies were then made independently by two researchers (BN and TC) and disagreements discussed. Any further disagreement or uncertainty was discussed and verified by two additional researchers (AB and KM). A hand search of reference lists of included studies as well as papers recommended through personal communication were also examined for relevant studies.

Inclusion and exclusion criteria

Studies were included in the review if they empirically measured treatment or management decision making as an anticipated or experienced outcome of being given or told one terminology versus another or others for a specific condition. Studies were excluded if they were reviews, editorials or commentaries or if they assessed participants less than 18 years of age making a decision for themselves (see online supplementary appendix 2).

Quality assessment and data extraction

All studies that met the inclusion criteria were appraised for study quality by two authors (BN and TC) independently using two separate tools, one for quantitative studies and one for the qualitative study. The quantitative studies were appraised using a modified version of the Cochrane Collaboration's tool for assessing risk of bias, which was developed by study authors (table 1). As the quantitative studies included in this review were not clinical trials, study authors adapted the Cochrane Collaboration's tool for assessing risk of bias to appraise the studies for items relating to study design, study setting, study validity and analysis. Study validity included allocation of participants to different terminology, blinding of participants to the study hypotheses and reporting and measurement bias. The qualitative study was appraised using criteria adapted from the Consolidated Criteria for Reporting Qualitative Studies (COREQ) framework[12] and results appear in online supplementary appendix 3.

Two categories of study quality were identified by study authors according to each study's methodological characteristics. In high-quality studies (lower risk of bias), the majority of criteria were fulfilled and done well, while in low-quality studies (higher risk of bias), the majority of criteria were not done or done poorly.

Data from the final studies included in the review were extracted independently into a standardised template by two authors (BN and TC). Any discrepancies were discussed and resolved by the entire research team. For studies where the data was not reported or was unclear, authors were contacted and responded for confirmation of details of the data. Authors were also asked to confirm the extracted data to minimise any reporting bias. All authors responded. 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 results.[13]

RESULTS

Initial search results identified 1399 papers. After removal of duplicates and screening by title and abstract, 20 studies from the search and 1 study identified from reference lists underwent full-text reviews. Seven studies that reported the impact of different terminology for the same condition on treatment or management decision

Table 1 Criteria for assessment of study quality for quantitative studies

	Information extracted
Study design	Independent sample or paired sample design If independent: whether groups were randomised?
Study setting	Community sample, clinics, hospital, other
Selection bias	Independent sample: Were participants allocated to each group randomly? Were samples similar in terms of important characteristics? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described? Paired sample: Were terms randomised? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described?
Performance bias	Were participants kept blind to the study hypothesis?
Attrition bias	Was attrition or exclusions (lost to follow-up) reported?
Reporting bias	Examination of selective reporting Did the study have a protocol?
Measurement bias	Exposure variable—describe the term used Outcome measures—how were outcomes measured? Were they validated or referenced? Confounders—reporting of additional measures and/or demographics
Analysis	Were appropriate statistical tests used to analyse data and report results?

making were identified (figure 1). Three studies reported on DCIS,[14–16] two on common childhood conditions (gastro-oesophageal reflux disease (GERD), conjunctivitis),[17 18] one on polycystic ovary syndrome (PCOS)[19] and one on a bony fracture.[20] Six studies reported quantitative findings[14 15 17–20] and one study reported qualitative findings.[16] The key characteristics of these studies and their methods are summarised in table 2. All studies were hypothetical and involved various samples of community members who were not currently and/or previously diagnosed with the condition assessed. Of the six quantitative studies, five involved a randomised experimental design, with two using a paired sample design (cross-over) and three using an independent sample design (2×2 factorial). The qualitative study was linked to one of the quantitative studies,[15] as women from the study were asked if they would be willing to participate in an additional qualitative interview for further investigation of the topic.

Overall, the majority of studies were of higher quality with a lower risk of bias; however, one study did have a high risk of bias as the quality of reporting was low[20] (table 3). Higher quality studies had defined study populations and settings, had low selection bias and described the data collection, methods and analysis appropriately. The qualitative study included in the review was rigorous and had appropriate methods to reach its research objectives. Researchers and the target population were defined, the sampling strategy was explained and data collection methods and data analysis were appropriate and well documented.

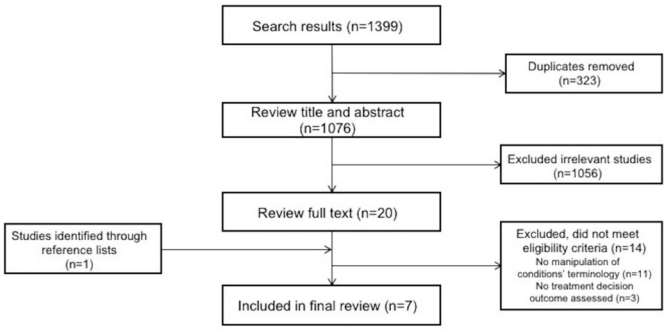

Figure 1 Flow diagram of study selection.

Since the strength of the review is the diversity of included studies, results of the studies are summarised individually by study quality category, with data on management outcomes summarised in table 4. Following this is a narrative synthesis of the effect of terminology (more medicalised or precise terms vs less or non-medicalised terms) on management preferences (invasive management preference vs non-invasive management preference). Importantly for each study, we identify the classifications of management preferences and terminology and provide justification where applicable (see online supplementary appendix 4).

Results from individual higher quality studies

Copp *et al* 2017 (PCOS)

This study on the influence of the term 'PCOS' found that when young women were given the term 'PCOS' in a hypothetical scenario of a doctor's visit, they had significantly higher intention to have an ultrasound compared with women who were given the term 'hormonal imbalance' (mean=6.62 vs mean=5.76, $F(1, 176)=4.63$, p=0.033). The study also found that those who received the term 'PCOS' perceived their hypothetical condition to be more severe (17.7 vs 15.82, $F(1, 176)=5.65$, p=0.019) and had lower self-esteem compared with women who were not

Table 2 Characteristics of included studies

Study	Year	Country	Disease focus	Study design	Study sample and setting	Study aims	Terms manipulated	Primary outcomes measured	Other outcomes measured
Copp *et al*	2017	Australia	PCOS	Randomised experimental design (2×2x(2) factorial design)	181 female university students	To test the impact of the PCOS disease label on intention to undergo an ultrasound and psychosocial outcomes	'Polycystic ovary syndrome' versus 'hormonal imbalance'	Intention to have an ultrasound (10 point scale: 1 = definitely will not to 10 = definitely will)	Negative affect, self-esteem, perceived severity of condition, credibility of GP and interest in a second opinion
McCaffery *et al*	2015	Australia	DCIS	Randomised experimental design (cross-over design)	269 healthy women from a community sample	To examine whether the use of terminology including the term cancer to describe DCIS increased hypothetical level of concern and treatment preferences	'Abnormal cells' versus 'pre-invasive breast cancer cells'	Treatment preferences: treatment versus watchful waiting, measured on a five-point Likert scale (definitely prefer treatment, probably prefer treatment, prefer to do nothing, probably prefer watchful waiting, definitely prefer watchful waiting)	Level of concern measured on a five-point Likert scale (Extremely concerned–Not concerned at all)
Nickel *et al**	2015	Australia	DCIS	Semistructured qualitative interviews	26 women from a community sample	To understand how different proposed terminologies for DCIS affect women's perceived concern and management preferences	'Abnormal cells' versus 'pre-invasive breast cancer cells' 'ductal carcinoma in situ' 'ductal intraepithelial neoplasia' 'indolent lesions of epithelial origin'	Women's qualitative responses to terminologies with and without the cancer term on level of concern and management preferences	
Omer *et al*	2013	USA	DCIS	Randomised experimental design (cross-over design)	394 healthy women with no history of breast cancer from a hospital patient registry	To examine how women respond to terminology for DCIS without the cancer term	'Non-invasive breast cancer', 'breast lesion', 'abnormal cells'	Treatment preferences (choice between surgery, medication, active surveillance)	
Scherer *et al*	2013	USA	GERD	Randomised experimental design (2×2 factorial design)	175 parents aged 18 years or older presenting at a primary care paediatric clinic	To determine if the disease label GERD influences parents' perceived need to medicate an infant	'GERD' versus no label 'this problem'	Parents' interest in giving their infant medication (three measures using six-point Likert scale: No, definitely not–Yes, definitely)	Perception of illness severity (three measures using five-point Likert scale: worry = Not at all worried–Very worried, serious = Not at all serious–Very serious, sick = Strongly disagree–Agree) Appreciation of medication offered (one measure using five-point Likert scale: No, definitely not–Yes, definitely)
Scherer *et al*	2015	USA	Conjunctivitis	Randomised experimental design (2×2 factorial design)	159 parents aged 18 years or older presenting at a primary care paediatric clinic	To determine whether the 'pink eye' term would influence parents' beliefs about the condition and their interest in using antibiotics	'Pink eye' versus 'eye infection'	Parents' decision to give their child antibiotics measured on a six-point Likert scale (No, definitely not–Yes, definitely)	Targeted beliefs about contagiousness measured on a five-point Likert scale (Not at all contagious–Very contagious) Parents ability to send their child to day care measured on a five-point Likert scale (Not at all likely–Very likely)

Continued

Table 2 Continued

Study	Year	Country	Disease focus	Study design	Study sample and setting	Study aims	Terms manipulated	Primary outcomes measured	Other outcomes measured
Azam *et al*	2010	UK	Bony fracture	Cross-sectional survey	100 adult patients presenting at an emergency department	To assess the way different terms used to describe a fracture affect the understanding a patients has of it including the perceived severity of the injury and how the patient expects to be treated	'A crack in the bone' 'A broken bone' 'A fracture' 'A hairline fracture' 'A greenstick fracture'	Expected treatment (choice between heals on own, sling, cast, operation)	Perceived severity on a 10-point scale (one being minimally problematic to 10 being a very serious injury)

*Included qualitative study.
DCIS, ductal carcinoma in situ; GERD, gastro-oesophageal reflux disease; GP, general practitioner; PCOS, polycystic ovary syndrome.

given the term 'PCOS' (25.86 vs 27.56, $F(1, 176)=4.74$, p=0.031). After women received information about the potential of PCOS overdiagnosis in a second scenario, both intention and perceived severity decreased, regardless of the term given (both p<0.001).

The study also found a significant three-way interaction between the term 'PCOS', information about ultrasound reliability and overdiagnosis information ($F(1, 176)=4.23$, p=0.041), where that for those who did not receive the term 'PCOS', intention was significantly lower for those who received information about the unreliability of ultrasounds compared with those who received no information. For women who received the term 'PCOS', however, intention was high, even when told that the ultrasound was unreliable. This difference disappears after information about overdiagnosis is given in the second scenario, suggesting the provision of overdiagnosis information reduces the effect of the term.

McCaffery *et al* 2015 (DCIS)

McCaffery and colleagues' study on the impact of DCIS terminology on treatment preference (immediate treatment vs watchful waiting) and women's level of concern found no significant differences in treatment preference between arm A (women who were given the term 'abnormal cells' first and then were given the term 'preinvasive cancer cells') and arm B (women who were given the term 'preinvasive cancer cells' first and then were given the term 'abnormal cells') with 33% and 40% of women respectively favouring treatment (p=0.23). However, 18% of women in arm A who were initially given the 'abnormal cells' terminology changed their preference to treatment when the terminology was switched to 'preinvasive breast cancer cells' while only 6% changed to watchful waiting (p=0.008). In contrast, there were no significant changes in treatment preference in arm B when the terminology was switched the other way (9% vs 8% changed their stated preference).

Similarly, this study found that there was no significant difference between arms with regard to level of concern with 49% and 44% of women indicating that they would be extremely concerned p=0.600. However, when the alternative term was used, women in arm A (who were initially given the 'abnormal cells' term and then were given the term 'preinvasive breast cancer cells') were significantly more likely to report increased concern than women in arm B (67% vs 52%, p=0.001).

Findings from this study were also supported by the included qualitative study conducted with a subset of women.[16]

Nickel *et al* 2015 (DCIS)

This qualitative study investigated in-depth how different proposed terminologies to describe DCIS affected women's treatment preferences and psychological outcomes by conducting semistructured interviews with women of varying education, cancer screening experience and with no history of a DCIS diagnosis. Findings

Table 3 Risk of bias summary for quantitative studies

Study	Study design	Study setting	Selection bias	Performance bias	Attrition bias	Reporting bias	Measurement bias	Analysis
Copp 2017[19]	+	+	+	?	+	+	+	+
McCaffery 2015[15]	+	+	+	?	+	+	+	+
Omer 2013[14]	+	?	+	?	?	+	+	+
Scherer 2013[17]	+	+	+	?	?	+	+	+
Scherer 2015[25]	+	+	+	?	?	+	+	+
Azam 2010[20]	–	+	–	?	?	–	–	–

+ Low risk of bias (done well)
? Unclear or unknown (not reported)
= High risk of bias (not done or done poorly)

demonstrated that overall women preferred a diagnosis of DCIS to be communicated using terminology that did not include the term cancer, as women generally exhibited stronger negative reactions when the cancer term was used to describe DCIS compared with when a non-cancer term was used. Although concern seemed to be high overall, women displayed a high level of interest in watchful waiting when it was described to them in a hypothetical scenario as a safe and effective option, and told that they could proceed to treatment in the future if necessary.

Omer *et al* 2013 (DCIS)

This study also examined the impact of DCIS terminology on women's treatment preferences by comparing three identical scenarios, with the only difference being the term used to described DCIS ('non-invasive cancer', 'breast lesion' and 'abnormal cells'). Treatment options presented were surgery, medication and active surveillance. All participants saw all scenarios, with the order of scenarios varied across participants. When DCIS was described as a non-invasive cancer, 53% of participants preferred non-surgical options, whereas 66% chose non-surgical options when it was described as breast lesion and 69% chose non-surgical options when it was described as abnormal cells (p≤0.001). 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, CI 1.34 to 3.34, p=0.001), lesion (OR 1.96, 1.20 to 3.19, p=0.001) and abnormal cells (OR 1.63, 1.01 to 2.67, p=0.048).

Scherer *et al* 2013 (GERD)

This study on the influence of the term 'GERD' (vs no term or label given) on parents' preferences for medication for their infant found that parents who received the term 'GERD' in the scenario were more interested in medication than parents who did not receive that term ($F(1, 165)=6.95$, $p<0.01$). To assess parental interest in antibiotics, study authors combined three highly intercorrelated questions involving parent's interest in medication: "Will you give your infant this medicine?", "Do you think your infant needs the medicine your doctor offered?" and "Do you think that the medicine will help your infant get better?"

The study also found a significant interaction between the term GERD and report of medicine ineffectiveness ($F(1, 165)=4.52$, $p<0.05$), as parents who received a GERD diagnosis were interested in medicating the infant, even if they were told that the medications are likely ineffective. By contrast, parents not given a diagnosis were interested in medication only when they were not given information about medication effectiveness, thus able to assume that the medications are effective.

All parents in the study were also asked (using a 0–5 Likert scale with labelled end-points for each question) whether they were worried about their infant's health (mean=2.28, SD=1.30), thought the condition was somewhat serious (mean=2.12, SD=1.19) and were relatively unlikely to describe their infant as being sick (mean=1.87, SD=1.45). These answers were not influenced, however, by the presence or absence of the term 'GERD' (all p>0.12).

Scherer *et al* 2015 (conjunctivitis)

A similar study design conducted by Scherer and colleagues did not find an initial difference between the term 'pink-eye' and 'eye infection' on parents' preference to medicate their infant. However, when symptoms were referred to as an 'eye infection', information about antibiotic ineffectiveness significantly reduced interest in using medication ($F(1, 62)=14.67$, $p\leq0.001$). By contrast, when parents were told that the symptoms were 'pink eye', interest in antibiotics was not reduced by information about antibiotic ineffectiveness ($F(1, 74)=0.93$, $p=0.33$).

Study authors also measured (using a 0–5 Likert scale with labelled end-points) parents' perceptions about contagiousness and belief that their child could attend child care and found that parents who received the term 'pink-eye' thought that the symptoms were

Table 4 Key findings from individual studies

Study	Primary outcome	Primary outcome by terminology		Primary outcome statistical significance*	Other outcomes
		Medicalised term	Non-medicalised term		
Copp 2017 (n=181)[19]	Intention to have an ultrasound	**Polycystic ovary syndrome (n=90)** mean=6.62 (on a 10-point Likert scale where 1=Definitely will not to 10=Definitely will)	**Hormonal imbalance(n=91)** mean=5.76 (on a 10-point Likert scale where 1=Definitely will not to 10=Definitely will)	Primary outcome significant (p=0.033). Women who received 'PCOS' term in the scenario had significantly higher intentions to have an ultrasound than those who received the 'hormonal imbalance' term $F(1, 176)=4.63$, p=0.033. After women received information on overdiagnosis, both intention and perceived severity decreased, regardless of the terminology of the condition (both p<0.001).	Self-esteem: Women's self-esteem was significantly lower for those given the term 'PCOS' than those given the term 'hormonal imbalance' $F(1, 176)=4.74$, p=0.031. Perception of severity: Women who were given the term 'PCOS' had significantly higher perceived severity about the condition than those given the term 'hormonal imbalance' $F(1, 176)=5.64$, p=0.019. Negative effect, credibility of doctor and interest in a second opinion: No difference between terms (all p>0.05).
McCaffery 2015 (n=269)[15]	Treatment preference	**Preinvasive breast cancer cells (n=128)** 40% (51) prefer treatment 60% (77) prefer watchful waiting Change in terminology 41% (52) prefer treatment 59% (76) prefer watchful waiting	**Abnormal cells (n=141)** 33% (47) prefer treatment 67% (94) prefer watchful waiting Change in terminology 45% (63) prefer treatment 55% (78) prefer watchful waiting	Primary outcomes not significant; however, change in terminology shows a significant difference. There were no significant differences in treatment between arm A (women who were given the term 'abnormal cells' first and then were given the term 'preinvasive cancer cells') and arm B (women who were given the term 'preinvasive cancer cells' first and then were given the term 'abnormal cells') with 33% and 41% of women respectively favouring treatment, p=0.23. In arm A, 18% of women changed their preference to treatment while only 6% changed to watchful waiting (p=0.008). No significant treatment preferences were observed in arm B (9% vs 8%, p>0.99).	Level of concern: There was no significant difference between arms with 49% and 44% of women indicating they would be extremely concerned p=0.600. However, when the alternative term was used, women in arm A ('abnormal cells' terminology first and then 'preinvasive cancer cells' terminology) were significantly more likely to report increased concern than women in arm B ('preinvasive cancer cells' terminology first and then 'abnormal cells' terminology), 67% vs 52%, p=0.001.
Omer 2013 (n=394)[14]	Treatment preference	**Cancer** 47% (186) surgery 20% (79) medication 33% (129) active surveillance	**Lesion** 34% (136) surgery 18% (70) medication 48% (188) active surveillance Abnormal cells 31% (124) surgery 21% (82) medication 48% (188) active surveillance	Primary outcome significant (p≤0.001) When DCIS was described as a 'non-invasive cancer' 53% (208 of 394) of participants preferred non-surgical options, whereas 66% (258 of 394) chose non-surgical options when it was described as 'breast lesion' and 69% (270 of 394) chose non-surgical options when it was described as 'abnormal cells' (p≤0.001).	

Continued

Table 4 Continued

Study	Primary outcome	Primary outcome by terminology: Medicalised term	Primary outcome by terminology: Non-medicalised term	Primary outcome statistical significance*	Other outcomes
Scherer 2013 (n=175)[18]	Parents' interest in medicating their infant (three items pooled and mean reported)	**GERD term (n=87)** mean=2.51 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	**No label (n=88)** mean=2.04 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Primary outcome significant (p<0.01) Parents who received the 'GERD' term in the scenario were more interested in medication than parents who did not receive that term, $F(1, 165)=6.95$, p<0.01. Parents not given the 'GERD' term were interested in medication only when they were not given information about medication effectiveness; therefore, they were allowed to assume that the medications are effective $F(1, 165)=4.52$, p<0.05.	Perception of illness severity: Findings were not influenced by the presence or absence of the 'GERD' term (all p>0.12). Appreciation of medication offer: Parents were least appreciative of medication when they were told that there medication was ineffective and were not given the 'GERD' term $F(1, 165)=7.16$, p<0.01
Scherer, 2015 (n=158)**[25]	Parents' decision to give their child antibiotics	**Pink-eye (n=82)** mean=2.43 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	**Eye infection (n=76)** mean=2.32 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Primary outcome not significant; however, when adjusted for information of medication effectiveness, outcomes are significant. Parents were less willing to give their child antibiotics when symptoms were referred to as an 'eye infection' information about antibiotic ineffectiveness significantly reduced $F(1, 62)=14.67$, p≤0.001. By contrast, when parents were told that the symptoms were 'pink eye' interest in antibiotics was not reduced by information about antibiotic ineffectiveness $F(1, 74)=0.93$ p=0.33.	Beliefs about contagiousness: Parents who received the 'pink-eye' term thought that the symptoms were significantly more contagious than parents who received the 'eye infection' term, $F(1, 137)=11.21$, p=0.001. Feelings towards sending child to day care: Parents who received the 'pink-eye' term were less likely to believe that their child would be allowed to go to child care than parents who received the 'eye infection' term, $F(1, 141)=9.70$, p=0.002.
Azam 2010 (n=100)[20]	Expected treatment	**Broken bone** 3% heals on own 39% sling 56% cast 2% operation **Greenstick fracture** 47% heals on own 25% sling 22% cast 6% operation **Fracture** 14% heals on own 44% sling 36% cast 6% operation	**Crack in the bone** 24% heals on own 57% sling 13% cast 6% operation **Hairline fracture** 45% heals on own 29% sling 26% cast 0% operation	t-tests which demonstrate significance: Crack versus break=p<0.0001 Crack versus hairline fracture=p<0.0001 Crack versus greenstick fracture=p<0.0001 Break versus hairline fracture=p=0.0001 Break versus greenstick fracture=p<0.0001 Fracture versus hairline fracture=p<0.0001 Fracture versus greenstick fracture=p<0.0001	Patients perspectives of severity: (mean/median score, no SDs given): 3.28/3 for 'a crack in the bone' 3.58/3 for 'a hairline fracture' 4.95/5 for 'a fracture' 5.28/5 for 'a greenstick fracture' 6.64/7 for 'a broken bone'

*As reported by original study authors.
**One case from total sample missing.
DCIS, ductal carcinoma in situ; GERD, gastro-oesophageal reflux disease; PCOS, polycystic ovary syndrome.

significantly more contagious (mean=4.04, SD=1.30) than parents who received the term 'eye infection' (mean=3.29, SD=1.63) ($F(1, 137)=11.21$, p=0.001). Parents who received the term 'pink-eye' were also less likely to believe that their child would be allowed to go to child care (mean=0.35, SD=0.90) than parents who received the term 'eye infection'(mean=0.93, SD=1.13) ($F(1, 141)=9.70$, p=0.002).

Individual results from the lower quality study

Azam *et al* 2010 (bony fracture)

This study found that patients' management expectations and perceptions of bony injuries differ based on the terminology used to describe the injury. Patients expected more invasive managements (operation or cast) when a more medicalised term was used to describe the injury compared with a less invasive treatment (sling or heals on own), with 58% of patients expecting invasive treatments for the term 'broken bone', 42% for 'fracture', 28% for 'greenstick fracture' and 26% for hairline fracture. In contrast, patients given the less medicalised term had a lower preference for invasive treatments, with 19% of patients choosing an invasive treatment for 'crack in the bone'. This study also found that patients perceived the injury to be more severe when a more medicalised term was used to describe the injury (average mean severity score out of 10; broken bone=6.64, greenstick fracture=5.28, fracture=4.95, hairline fracture=3.58, crack in the bone=3.28).

Synthesis of results

Table 5 summarises preferences for more invasive management option by type of terminology across the studies. Generally, there was a pattern in the same direction showing that when more medical or precise term was used, it resulted in a greater preference or interest in more invasive managements, whether this be intention to have an ultrasound for a PCOS diagnosis, surgical treatment for DCIS, increased interest in potentially ineffective medication or an operation or cast for a bony injury, although these differences did not always reach statistical significance. In contrast, when participants were given a less or non-medical/precise term, a higher proportion of participants chose a non-invasive management option. 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 with the comparator term). Studies varied as some found a significant effect of terminology on management decision making while others found a significant interactions effect or within-person effect and psychological outcome effects (eg, perceived severity, level of concern).

DISCUSSION

This is the first systematic review which synthesises the evidence on how different terminology given for the same condition impacts management preferences. Overall, the review demonstrates that different terminology used to describe the same condition can influence patient's management preferences. Results indicate that when a more medical or more precise term was used to describe a condition, people tended to have stronger preferences for more invasive management options. 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.

The terminology given during a diagnosis to patients is important, as when a healthy individual becomes a patient, they can immediately become more vulnerable to the words their clinicians use.[21] Using words that generate fear or anxiety cause patients to have more difficulty making informed decisions and becoming an active participant in their care.[22] Different terms used to describe

Table 5 Summary of preferences for more invasive management option by type of terminology*

Study	More medicalised or more precise term/s (%)	Less medicalised or less precise term/s (%)	Difference (%)	p Value
Copp 2017[19]	70	53	17	>0.05§
McCaffery 2015[15]	40	33	7	0.23
Omer 2013[14]	47	32.5	14.5	<0.001
Scherer 2013[18]	74	67	7	>0.1§
Scherer 2015[25]	60	58	8	>0.1§
Azam 2010[20]	39 (4 operation, 35 cast)†	19 (6 operation, 13 cast)‡	20	<0.025§

*Combined data where applicable and mean percentages reported, see online supplementary appendix 4 for explicit justification of categorisation of terminology.
†Broken bone, fracture, greenstick fracture, hairline fracture.
‡Crack in the bone.
§Calculated significance using raw (Copp and Scherer) and published (Azam) data based on our classification of which terms were more medicalised.

the same condition can have a direct influence on how patients understand their diagnosis and how threatening they perceive it to be.[23] It has also been shown that the use of interpretive terminology (eg, including the words positive or negative or using a metaphor),[24 25] the terminology used to describe a treatment choice,[26] describing a condition with plain language terminology as compared with jargon[27] and the severity of the characteristics of the diagnosis[28] can have an influence on medical decision making. Furthermore, a recent survey of medical students found that students were more likely to classify condition synonyms as a 'disease' if the term used to describe it was medical.[29] Together with findings from this review, these studies show that language is a powerful tool that has the potential to influence patients' thoughts and actions.

Our findings are also in line with the results of other research on the effect of labelling a condition in relation to the social implications it may have for the individual, rather than its effect on medical decision making. Research has found that disease labels can result in various emotional, cognitive and physical consequences.[30 31] Most notably, studies which have examined the effect of hypertension labelling have found that giving the label of hypertension to individuals (compared with giving no label) resulted in increased self-reported illness and absenteeism from work and a significant increase in blood pressure.[30 32 33] In the conjunctivitis study by Scherer and colleagues included in this review, the term 'pink-eye' was perceived as being more contagious, and parents were less likely to believe their child could go to childcare, compared with parents who received the condition described as an 'eye infection'. Additionally, in the PCOS study, women had significantly lower levels of self-esteem when the term 'PCOS' was used to describe their condition compared with when it was described as a hormonal imbalance.[19]

Using more medicalised or precise medical terminology and disease labels may also reduce a patient's sense of self-control, and therefore increase perceived severity and uptake of medications. The more medical the term or label, the less control a person may feel over the situation, increasing their perceived severity about the condition and creating a perception that more invasive interventions are warranted. This was also shown in the GERD, conjunctivitis and PCOS studies where there was a high level of interest in medical intervention when a precise medical term or label was given, even when participants were told that the medication or test was ineffective.[17 18]

This study was limited by the small number of included studies in the review, which also included three studies that were conducted by members of the review team. However, in light of recent and ongoing evidence of overtreatment across a number of conditions and suggestions from leading global medical bodies to change the terminology of low-risk conditions,[10 11] this review addressed an important and timely under-researched question.

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.[34 35]

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 synthesised the results narratively. Since a quantitative synthesis of the data was not possible, findings may not be generalisable. Furthermore, the delineation between more medical/precise and less or non-medical/precise was challenging in some studies. For example, it seemed clear that 'preinvasive 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 (eg, 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 online supplementary appendix 4). We note the Azam paper was of lower quality, which made interpretations about the terms used more difficult. Nonetheless, it was very clear that in all of the studies, the use of different terminology for the same condition tended to elicit different responses to management preferences as well as psychosocial outcomes. Further research could usefully explore in more detail what characterises terms that elicit stronger preferences for more invasive managements.

Importantly, all studies included in this review were hypothetical and were not presented to patients randomised to receive a diagnosis, and management preferences were assessed instead of actual management decision making. Patients facing real decisions may respond differently to those in the studies; however, it is likely that in real life situations, these effects may be even more pronounced. It is likely that patients would be more anxious than participants in current studies,

therefore more susceptible to judgement biases that result from more medicalised terminology or labels.

Against a backdrop of recent evidence of overdiagnosis and overtreatment across a number of conditions[5 36–38] and the potential physical and psychological impacts this may have for the patient, understanding how different terminology given for the same medical condition may influence patients' management preferences is important. This review suggests that the terminology used to describe a condition can influence patient preferences for management and related outcomes. Although further evidence is needed to help better understand precisely why some terms elicited stronger preferences for more invasive management, this review helps support the calls for changing the terminology of conditions where the risk of progression is low. Changing the terminology in low-risk conditions or conditions with indolent clinical course may be a potential communication strategy to help shift assumptions that immediate invasive treatments or tests are always needed, allow for better shared decision making between clinicians and patients and the consideration of more conservative management options.

Acknowledgements The authors would like the thank Phillippa Bourke for her help with the literature search and Dr Kevin McGeechan for his comments on the statistical aspects of the paper.

Contributors BN, AB and KM contributed to study concept and design. BN conducted the systematic literature search. BN and TC evaluated the eligibility of studies, performed the study quality assessment, extracted the data and conducted the evidence synthesis. BN, AB, TC, RM and KM interpreted the findings. BN drafted the manuscript and all authors critically reviewed and approved the manuscript.

Funding This paper was written with support from the Wiser Healthcare CRE grant from the National Health and Medical Research Council (NHMRC) #1104136. BN was supported by the Sydney Catalyst Research Scholar Award. KM was supported by an NHMRC fellowship (1029241). The funders had no role in the design or conduct of the study; in the collection, analysis and interpretation of the data; or in the preparation or approval of the manuscript.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; AB and RM are co-chairs of the steering group for the Preventing Overdiagnosis scientific conferences; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data available.

REFERENCES

1. Schofield PE, Butow PN, Thompson JF, *et al*. Psychological responses of patients receiving a diagnosis of Cancer. *Ann Oncol* 2003;14:48–56.
2. Arora NK. Interacting with Cancer patients: the significance of physicians' communication behavior. *Soc Sci Med* 2003;57:791–806.
3. Hoffman JR, Cooper RJ. Overdiagnosis of disease: a modern epidemic. *Arch Intern Med* 2012;172:1123–4.
4. Welch HG, Schwartz LM, Woloshin S. *Over-diagnosed: making people sick in the pursuit of health*. Boston: Beacon Press, 2011.
5. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ* 2012;344:e3502.
6. Heath I. Overdiagnosis: when good intentions meet vested interests--an essay by Iona Heath. *BMJ* 2013;347:f6361.
7. McCaffery KJ, Jansen J, Scherer LD, *et al*. Walking the tightrope: communicating overdiagnosis in modern healthcare. *BMJ* 2016;352:i348.
8. Dunn BK, Srivastava S, Kramer BS. The word "cancer": how language can corrupt thought. *BMJ* 2013;347:f5328.
9. Esserman LJ, Thompson IM, Reid B, *et al*. Addressing overdiagnosis and overtreatment in Cancer: a prescription for change. *Lancet Oncol* 2014;15:e234–e242.
10. Allegra CJ, Aberle DR, Ganschow P, *et al*. National Institutes of Health State-of-the-Science Conference statement: diagnosis and management of ductal carcinoma in Situ September 22-24, 2009. *J Natl Cancer Inst* 20102010;102:161–9;102:161–9.
11. Esserman LJ, Thompson IM, Reid B. Overdiagnosis and overtreatment in Cancer: an opportunity for improvement. *JAMA* 2013;310:797–8.
12. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–57.
13. Schünemann HJ, Oxman AD, Higgins JPT, *et al*; *Presenting results and 'Summary of findings' tables — results without meta-analyses*, 2011.
14. Omer ZB, Hwang ES, Esserman LJ, *et al*. Impact of ductal carcinoma in situ terminology on patient treatment preferences. *JAMA Intern Med* 2013;173:1830–1.
15. McCaffery K, Nickel B, Moynihan R, *et al*. How different terminology for ductal carcinoma in situ impacts women's concern and treatment preferences: a randomised comparison within a national community survey. *BMJ Open* 2015;5:e008094.
16. Nickel B, Barratt A, Hersch J, *et al*. How different terminology for ductal carcinoma in situ (DCIS) impacts women's concern and management preferences: A qualitative study. *Breast* 2015;24:673–9.
17. Scherer LD, Finan C, Simancek D, *et al*. Effect of "Pink Eye" Label on Parents' Intent to Use Antibiotics and Perceived Contagiousness. *Clin Pediatr* 2016;55.
18. Scherer LD, Zikmund-Fisher BJ, Fagerlin A, *et al*. Influence of "GERD" label on parents' decision to medicate infants. *Pediatrics* 2013;131:839–45.
19. Copp T, McCaffery K, Azizi L, *et al*. Influence of the disease label 'polycystic ovary syndrome' on intention to have an ultrasound and psychosocial outcomes: a randomised online study in young women. *Hum Reprod* 2017;32:1–9.
20. Azam N, Harrison M. Patients' perspectives on injuries. *Emerg Med J* 2011;28:601–3.
21. Bedell SE, Graboys TB, Bedell E, *et al*. Words that harm, words that heal. *Arch Intern Med* 2004;164:1365–8.
22. Cassell EJ, Leon AC, Kaufman SG. Preliminary evidence of impaired thinking in sick patients. *Ann Intern Med* 2001;134:1120–3.
23. Chadha NK, Repanos C. Patients' understanding of words used to describe lumps: a cross-sectional study. *J Laryngol Otol* 2006;120:125–8.
24. Zikmund-Fisher BJ, Fagerlin A, Keeton K, *et al*. Does labeling prenatal screening test results as negative or positive affect a woman's responses? *Am J Obstet Gynecol* 2007;197:528.e1–528.e6.
25. Scherer AM, Scherer LD, Fagerlin A. Getting ahead of illness: using metaphors to influence medical decision making. *Med Decis Making* 2015;35:37–45.
26. Rot I, Ogah I, Wassersug RJ. The language of prostate Cancer treatments and implications for informed decision making by patients. *Eur J Cancer Care* 2012;21:766–75.
27. Pieterse AH, Jager NA, Smets EM, *et al*. Lay understanding of common medical terminology in oncology. *Psychooncology* 2013;22:1186–91.
28. Gavaruzzi T, Lotto L, Rumiati R, *et al*. What makes a tumor diagnosis a call to action? on the preference for action versus inaction. *Med Decis Making* 2011;31:237–44.
29. Erueti C, Glasziou P, Mar CD, *et al*. Do you think it's a disease? a survey of medical students. *BMC Med Educ* 2012;12:19.
30. Pickering TG. Now we are sick: labeling and hypertension. *J Clin Hypertens* 2006;8:57–60.

31. Hoyt CL, Burnette JL, Auster-Gussman L. "Obesity is a disease": examining the self-regulatory impact of this public-health message. *Psychol Sci* 2014;25:997–1002.
32. Haynes RB, Sackett DL, Taylor DW, *et al*. Increased absenteeism from work after detection and labeling of hypertensive patients. *N Engl J Med* 1978;299:741–4.
33. Rostrup M, Kjeldsen SE, Eide IK. Awareness of hypertension increases blood pressure and sympathetic responses to cold pressor test. *Am J Hypertens* 1990;3(12 Pt 1):912–7.
34. Nikiforov YE, Seethala RR, Tallini G, *et al*. Nomenclature Revision for Encapsulated Follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol* 2016;2:1023-9.
35. Patel KN. Noninvasive encapsulated follicular variant of papillary thyroid "Cancer" (or Not): Time for a Name Change. *JAMA Oncol* 2016;2:1005-6.
36. Independent UK Panel on breast Cancer screening. the benefits and harms of breast Cancer screening: an independent review. *Lancet* 2012;380:1778–86.
37. Moyer VA. On behalf of the U.S. Preventive Services Task Force. screening for prostate Cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012;157:120–34.
38. Brito JP, Morris JC, Montori VM. Thyroid Cancer: zealous imaging has increased detection and treatment of low risk tumours. *BMJ* 2013;347:f4706.