doi: 10.1111/hex.12344

Combined verbal and numerical expressions increase perceived risk of medicine side-effects: a randomized controlled trial of EMA recommendations

Peter Knapp PhD,* Peter H Gardner PhD† and Elizabeth Woolf BSc RN‡

*Senior Lecturer, Department of Health Sciences, University of York, York, †Senior Lecturer, Institute of Psychological Sciences, University of Leeds, Leeds and ‡Head of Cancer Web Content, Cancer Research UK, London, UK

Correspondence
Peter Knapp, PhD
Department of Health Sciences
University of York
York YO10 5DD, UK
E-mail: Peter.knapp@york.ac.uk

Accepted for publication
31 December 2014

Keywords: European Medicines Agency, paclitaxel, patient information, risk, side-effect, trial

Abstract

Background The study evaluated European Medicines Agency (EMA) recommendations on communicating frequency information on side-effect risk.

Methods The study used a 2 × 2 factorial trial, with random allocation of information about 10 side-effects of paclitaxel (Taxol) expressed using one of four formats. Recruitment was via the CancerHelpUK website. Information was conveyed using numerical frequency bands (e.g. 'may affect up to 1 in 10 people') or combined verbal terms and numerical bands (e.g. 'common: may affect up to 1 in 10 people'); in addition, the risk qualifier verb was manipulated, with risks expressed either as 'will affect…' or 'may affect…'. Participants then made six side-effect frequency estimates indicated their satisfaction with the information and evaluated the side-effects: how bad; how likely; how risky to health; and their influence on taking paclitaxel.

Results The sample comprised 339 people, of whom 37.5% had cancer. The combined verbal and numerical risk expressions resulted in higher estimates of side-effects, four of which reached statistical significance ($P < 0.05$), and participants also said that side-effects would be more likely. Use of 'may affect' or 'will affect' did not result in differences in any estimates.

Conclusions This is the first evaluation of the full range of combined verbal and numerical risk expressions recommended in EMA guidance; it demonstrates that they can lead to significant risk overestimations when compared to numerical frequency bands alone. The EMA should consider revising its guidance. Government agencies and professional bodies should be cautious about recommendations for risk communication in the absence of empirical evidence.

Background

Effective risk communication about medicines can be difficult but is important, given its influence on patients' perceptions and their willingness to take medicines.[1,2] Within the European Union, a licensed medicine must be supplied with a patient information leaflet (or PIL) including all known adverse effects. European Commission guidance to manufacturers has been to indicate the likely frequency of each side-effect, initially using five verbal terms (from *very common* to *very rare*), each representing a frequency band.[3] The intention was to increase the availability and consistency of information for patients.[4,5] However, use of the terms has been shown to be associated with significant risk overestimates.[6]

The obvious alternative to verbal terms – a numerical expression – is not a straight forward choice, particularly because print space in PILs is limited. Percentages work well for many patients but can be misunderstood, especially when the rate is <1%.[7] Frequency expressions (such as 'may affect *X* in 100 people') are mostly understood,[8] but the many side-effects associated with some medicines makes it unrealistic to state a rate for each of them. Furthermore, rates are mostly calculated from relatively small trial samples or post-marketing surveillance data in which a denominator has been estimated, both of which reduce precision. As a result, frequency bands (such as 'may affect up to 1 in 100 people') may be used to group side-effect lists; this saves space, reduces information load for readers and offers a solution to imprecision in rate estimates.

A more recent requirement from the European Medicines Agency (EMA) has been for risk expressions combining words and frequency bands (e.g. 'Common: may affect up to 1 in 10 people').[9] Such expressions have been supported within the UK National Institute for Health and Care Excellence (NICE) clinical guidance on medicine adherence,[10] which stated that verbal terms to convey risk should not be used without numerical information. In both cases, the guidance was based on consensus and an absence of specific research evidence for the combined risk expression format.

Verbal terms can be defended on the grounds that they are used when people talk about risk or uncertainty (e.g. 'that is very common');[11] furthermore, anecdotal reports are that verbal terms in PILs are liked by some patients, as an alternative to the perceived formality or difficulty of numbers. However, empirical support for patients' understanding of combined verbal and numerical expressions is lacking and a concern is that is the inclusion of the verbal term may 'frame' patients' understanding of the risk expression, leading to overestimation.[6,12] The recommended combined expression was evaluated in a previous, similar study and found some increased risk estimates by comparison with numerical frequency bands only.[13] However, that study did not assess the full range of EMA risk expressions and also had a relatively small sample, which might have masked true effects. Therefore, this study intended to evaluate people's understanding and interpretation of the combined (verbal and numerical) risk expression when compared with numbers only.

Another aspect of the recommended risk expression that might influence interpretations is the verb used to convey uncertainty. The EMA-recommended 'may affect up to 1 in *X* people' includes two indicators of uncertainty: 'up to 1 in *X*' to indicate rate imprecision, and 'may affect' to indicate uncertainty. Arguably, one indicator is redundant and their combined use may confuse. Consequently, this study further evaluated the EMA recommendation by comparing an expression using the double uncertainty indicator ('may affect' and 'up to') with a single uncertainty indicator ('will affect' and 'up to'), for their effect on risk interpretations.

The two variables were combined in a single study, using a factorial trial design, in which each participant was allocated to receive information about potential side-effects of medicines using one of the four forms of risk expression, to test the risk communication approach recommended by the EMA.

Method

Design

The study used a 2 × 2 factorial trial design, evaluating the effects of two risk expression formats and two verbal risk qualifiers (see Tables 1 and 2), to which participants were randomly allocated. For one of the two factors, the *Numerical format* presented the number of people likely to experience a side-effect, for example 'this side-effect will affect more than 1 in 10 people'. The *Combined verbal and numerical format* presented risk as, for example, 'common (will affect more than 1 in 10 people)'. For the second factor, the verbal qualifier was manipulated, so that half the participants were given the risk expression including the words '...will affect more than 1 in *X*...', while the others were given expressions containing '...may affect more than 1 in *X*...'. The factorial 2 × 2 design meant that each participant was allocated one of four risk expressions, as listed in Table 2.

Participants responded to the provided risk information by making six probability risk estimates and completing five Likert scales to assess risk perceptions (see Table 3 and Measures).

Table 2 The four groups formed within the factorial trial

Group 1: Numerical terms only; 'may affect up to...' (n = 91)
Group 2: Combined verbal and numerical expression; 'may affect up to...' (n = 85)
Group 3: Numerical terms only; 'will affect up to...' (n = 77)
Group 4: Combined verbal and numerical expression; 'will affect up to...' (n = 86)

Based on data from previous similar studies,[7,13–15] the factorial study sample size of 318 was determined by setting power at 90%, type 1 error equal to 0.025, to detect a 10% difference in estimates of having *any* side-effect (with an estimated standard deviation of 25.2).

Procedure

The participants had no personal interaction with the researchers; each took part via the Cancer Help UK website (www.cancerhelp.org.uk). On accessing one of two pages on the website (one on drugs commonly used in the treatment of cancer; the other on Taxol specifically), participants received an invitation to take part via a 'pop-up' window. The study recruited participants in the period July 2012 to February 2014.

Table 1 Scenario given to participants and an example of the four allocated formats

In this imaginary situation your doctor has told you that you need to take the medicine Paclitaxel (Taxol) as part of your treatment for cancer.	
Please read the information below about Paclitaxel and answer the questions that follow. You can look at the information again when answering the questions. We are interested in your first thoughts. Please don't spend too long thinking about your answers.	
Paclitaxel has some side effects which differ in terms of the chance of occurring. These include	
Very common: may affect more than 1 in 10 people	Bruising more easily
	Aching joints and muscles
Common: may affect up to 1 in 10 people	Mild skin rash
	Severe anaemia (causing tiredness and breathlessness)
Uncommon: may affect up to 1 in 100 people	Serious allergic reaction
	Blood clots
Rare: may affect up to 1 in 1000 people	Muscle weakness in arms, hands, legs
	Itching
Very rare: may affect up to 1 in 10 000 people	Hearing or sight disturbances
	Dizziness or fits
The example above shows one of four risk expressions: the combined verbal and numerical risk expression, and using the words 'may affect...'	

Table 3 Questions used in the study

1. What do you think is the chance that you will have aching joints and muscles from taking Paclitaxel?
(Questions 2–6 used the following terms instead of 'aching joints and muscles': severe anaemia; serious allergic reaction; itching; dizziness or fits; ANY side effect)
7. Overall how satisfied are you with the information you have just read about the risk of side effects from Paclitaxel?
8. Thinking about the information you have just read, how bad overall are the side effects from Paclitaxel?
9. From the information you have just read, how likely is it that you would have a side effect from Paclitaxel?
10. Thinking about the information you have just read, what do you think is the general risk to health from Paclitaxel?
11. How much would the side effect information you have just read affect your decision to continue taking Paclitaxel?

As the pop-up window was activated, the participant saw one of the four versions of the study pages, the choice determined by an allocation schedule produced by a random number generator in the PERL programming package. The procedure meant that allocation was concealed.

On the first study page, participants were given brief information about the study and told that they could withdraw at any time. Participants who agreed then proceeded to the second page, on which they read a hypothetical scenario and information about 10 potential side-effects of Taxol (with the likelihood expressed using one of the four risk expressions). Participants then completed a series of questions (see Measures), without knowledge of the other study conditions, that is both the participants and researchers were 'blinded'. The final study web page thanked them and gave additional information about both the study and Taxol. A CancerHelp UK helpline phone number was provided and the researchers' email addresses, for use if required. On completing the study, participants were invited to leave an email address if they wanted a summary of the study results. They were then returned to the originating CancerHelp UK web page.

Materials

The ten side-effects were described using lay language, for example 'itching', or a combination of technical and lay language, for example 'severe anaemia (causing tiredness and breathlessness)', see Table 1. The side-effects were chosen so that we used two for each of the five risk incident frequency bands recommended by the European Commission (>10%; 1–10%; 0.1–1%; 0.01–0.1%; <0.01%, which had been assigned the terms, respectively, *very common, common, uncommon, rare, very rare*). The side-effect risk information for Taxol was taken from the CRUK database (dated March 2006). Exact side-effect incident rates were available for only 3 side-effects (bruising more easily 11%, aching joints and muscles 60%, severe anaemia 6%); for the remainder, rates were available within a frequency band.

Measures

Participants were first asked to estimate the probability that they would experience each of five individual side-effects from taking Taxol. (We had selected one side-effect from each of the five frequency categories). The sixth question asked participants to estimate the likelihood of getting *any* side-effect from taking Taxol. Given the available source incidence rate data, it was not possible to calculate either the true or maximal rate of getting *any* side-effect.

Participants then completed five Likert rating scales (see Table 3), each scored in the range 1–6. Participants expressed: how satisfied they were with the presented information, how severe the side-effects were, how likely they would be to experience a side-effect, the general risk to health, and how much the information would affect their decision to continue treatment. Participants could scroll back at any stage to the web page containing the side-effect information.

Participants were then asked for information about themselves (sex, age, location, whether English was their first language, highest educational qualification and why they were looking for information about Taxol).

Analyses

Based on our previous studies using this method, we expected that not all participants would complete all question items. We included participants in the analysis if they had completed at least 5 of the 11 measures (i.e. risk estimates or Likert items). For risk estimates and Likert measures the mean, median and standard deviation statistics were calculated, by allocated format, with variation tested by ANOVA univariate *F*-tests for between-subject effects. For Likert scale measures, means (and standard deviations) were calculated, by allocated format. We also tested for presence of interactions between the two randomly allocated conditions. *Post hoc,* we also conducted the same analyses on the subsample of participants who had cancer.

Tables 5–8 report the means, standard deviations, *F*-values and probability values.

Research ethics

The study was given approval by the Research Ethics committee of the Institute of Psychological Sciences, University of Leeds (Ref 12-0127).

Results

Participants

In total 783 people activated the pop-up study window, of whom 422 did not proceed and entered no data. 339 (of the remaining 361) people entered sufficient data for inclusion in the study and had been allocated as follows: 91 to numerical terms only, 'may affect…'*;* 85 to verbal and numerical terms, 'may affect…'; 77 to numerical terms only, 'will affect…'; and 86 to verbal and numerical terms, 'will affect…', see Table 2. The 339 participants' characteristics are summarized in Table 4. Three-quarters of those who took part were women, and participants had a wide range of ages (16–80) with a median of 49. Most participants (78.6%) came from the UK and almost all had English as their first language. More than one-third (37.5%) were people with cancer; in addition, around a quarter (26.5%) of participants had a close relative or friend with cancer, and 16.2% were health professionals. Almost two-thirds had either a university degree or a professional qualification. The four trial allocations had similar proportions of

Table 4 Sample characteristics (n = 339)

Gender	231 (76.7%) female; 70 (23.3%) male; n = 38 not stated
Age	Mean 48.5 (SD 13.1); median 49; range 16–80. n = 33 not stated
Location	239 (77.3%) UK; 70 (22.7%) non-UK (29 USA; 15 Australia; 6 New Zealand; 5 India; 4 Canada; 3 Republic of Ireland; 2 Belgium; 1 participant each from Austria, Denmark, France, Italy, Malaysia, Switzerland); 30 not stated
English as first language	289 (93.2%) Yes; 21 (6.8%) No; n = 29 not stated
Reason for visiting the web page	24 (7.7%) Currently taking Taxol 61 (19.7%) Have cancer but not taking Taxol 19 (6.1%) Have cancer and previously taken Taxol 12 (3.9%) Have cancer and about to start taking Taxol 82 (26.5%) Have a close relative or friend with cancer 50 (16.2%) Health-care professional 61 (19.7%) None of the above 30 Not stated
Highest educational qualification	9 (2.9%) No formal qualification 19 (6.1%) GCSE/O Level/qualification typically gained at age 16 36 (11.6%) A Level/qualification typically gained at age 18 134 (43.4%) University degree 101 (32.7%) Professional qualification 10 (3.2%) Other; 30 Not stated

participants on each of the demographic variables.

Rating scale responses and risk estimates

The risk estimates and Likert scales were assessed in univariate analyses of variance. In all analyses, plots of residuals (i.e. the error terms) were approximately normally distributed. When asked to rate the chance of the five individual or any side-effects, participants who received the combined verbal and numerical risk expression made higher estimates on all, although only four reach statistical significance. The relative increase in risk frequency estimates ranged from 31% to 225% (see Table 5). When estimating the chance of having any side-effect, the difference between 18.7% (numerical only) and 31.1% (combined verbal and numerical) is statistically significant and represents a 66.3% relative increase. A 2 × 2 MANOVA on the six risk estimates found a statistically significant main effect for combined verbal and numerical terms vs. numerical terms only ($F = 2.77$, d.f. = 6, $P = 0.013$), no effect of 'may' vs. 'will' ($F = 0.383$, d.f. = 6, $P = 0.889$) and no interaction effect ($F = 0.599$, d.f. = 6, $P = 0.731$).

Levels of variance on the risk estimates among the sample were high (see standard deviations in Tables 5 and 6), as noted in previous studies.[7,13–15] Variance was higher in risk estimates from the combined verbal and numerical terms, although the differences were not tested statistically.

Table 5 All participants' responses (mean, SD), according to verbal and numerical format vs. numerical alone

	Format					
	Combined verbal and numerical expression		Numerical term only		Relative risk estimate increase: verbal + numerical > numerical only	ANOVA *F*-value (*P*)
	M (SD)	*n*	*M* (SD)	*n*		
Side-effect risk estimates (actual %)						
Chance of aching joints & muscles (60%)	23.1 (25.4)	135	13.2 (15.7)	135	+75.0%	**14.90 (<0.001)**
Chance of severe anaemia (6%)	14.7 (17.0)	145	11.2 (10.7)	138	+31.2%	**4.17 (0.042)**
Chance of serious allergic reaction (0.1–1%)	4.2 (9.5)	144	2.3 (8.3)	142	+82.6%	3.04 (0.082)
Chance of itching (0.01–0.1%)	5.2 (15.4)	143	1.6 (8.6)	133	+225%	**5.74 (0.017)**
Chance of dizziness or fits (<0.01%)	3.3 (11.4)	139	1.4 (9.6)	131	+135.7%	2.14 (0.149)
Chance of ANY side-effect	31.1 (30.6)	137	18.7 (19.4)	132	+66.3%	**15.48 (<0.001)**
Likert scale items						
Satisfaction with side-effect information	4.4 (1.3)	159	4.3 (1.4)	155		1.49 (0.22)
How bad overall are Taxol side-effects	3.4 (1.0)	159	3.4 (1.1)	156		0.24 (0.88)
Likelihood of having a side-effect	3.8 (1.4)	159	3.4 (1.4)	155		**5.74 (0.017)**
General risk to health from Taxol	3.2 (1.0)	158	3.1 (1.0)	156		0.91 (0.34)
Impact of information on decision to take Taxol	2.7 (1.5)	158	2.8 (1.5)	155		0.49 (0.48)

Bold text indicates statistical significance ($p < .05$).

Table 6 All participants' responses (mean, SD), according to 'may affect' or 'will affect' risk qualifier terms

	Format					
	'May affect' risk qualifier		'Will affect' risk qualifier		Relative risk estimate increase:	
	M (SD)	*n*	*M*	(SD) *n*	'may affect' > 'will affect'	ANOVA *F*-value (*P*)
Side-effect risk estimates (actual %)						
Chance of aching joints & muscles (60%)	18.9 (23.2)	144	17.2 (19.7)	126	*+9.9%*	0.44 (0.51)
Chance of severe anaemia (6%)	13.3 (15.2)	150	12.6 (13.4)	133	*+12.3%*	0.17 (0.68)
Chance of serious allergic reaction (0.1–1%)	3.6 (10.2)	148	2.8 (7.4)	136	*+24.9%*	0.52 (0.47)
Chance of itching (0.01–0.1%)	3.4 (12.7)	145	3.5 (12.8)	131	−3.7%	0.01 (0.94)
Chance of dizziness or fits (<0.01%)	2.2 (10.5)	143	2.6 (10.7)	127	−16.3%	0.10 (0.75)
Chance of ANY side-effect	25.9 (27.4)	144	23.9 (25.3)	125	*+8.4%*	0.38 (0.54)
Likert scale items						
Satisfaction with side-effect information	4.4 (1.3)	165	4.3 (1.3)	149		0.20 (0.66)
How bad overall are Taxol side-effects	3.5 (1.1)	165	3.2 (1.0)	160		**5.26 (0.022)**
Likelihood of having a side-effect	3.7 (1.4)	165	3.6 (1.4)	149		0.60 (0.43)
General risk to health from Taxol	3.2 (1.0)	164	3.1 (1.0)	150		0.26 (0.61)
Impact of information on decision to take Taxol	2.9 (1.4)	164	2.7 (1.5)	149		1.51 (0.22)

Bold text indicates statistical significance ($p < .05$).

There were few differences between the groups on the five Likert scale measures, other than on the estimated likelihood of having a side-effect. Participants who saw the combined risk expressions thought that side-effects were more likely to occur (their mean estimate of 3.8 was 0.4 higher than the numerical-only group); the difference is statistically significant.

The use of the terms 'may affect' and 'will affect' produced very similar rates on the six estimates of the chance of side-effects (see Table 6). A similar picture was evident on the Likert ratings, with no differences apparent on four of the five estimates and just one ('How bad overall are the side-effects?') being rated higher in the 'may affect' group (means 3.5 vs. 3.2; $P = 0.022$).

We tested for interactions between the two manipulated variables (combined verbal and numerical vs. numerical alone; 'may affect' vs. 'will affect') on all measures, as a check for independence of the two manipulations. One of the eleven interaction statistics was significant: estimated frequency of dizziness ($F = 4.17$; $P = 0.042$). All other interaction statistics had probabilities of $P > 0.05$, suggesting that overall the two manipulated variables produced effects that were independent of each other.

Ratings of participants with cancer

We analysed the risk estimates and Likert measures for the subsample comprising participants with cancer ($n = 116$). The results show the same pattern as for the whole study sample

Table 7 Participants with cancer (n = 116) responses (mean, SD), according to combined vs. numerical formats

	Format					
	Combined verbal and numerical expression		Numerical term only		Relative risk estimate increase: verbal + numerical > numerical only	ANOVA
	M (SD)	*n*	*M* (SD)	*n*		*F*-value (*P*)
Side-effect risk estimates (actual %)						
Chance of aching joints & muscles (60%)	28.4 (29.2)	52	12.5 (16.1)	41	+127.2%	**9.86 (0.002)**
Chance of severe anaemia (6%)	16.9 (19.3)	52	12.1 (13.9)	43	+39.7%	1.87 (0.175)
Chance of serious allergic reaction (0.1–1%)	6.4 (12.7)	48	2.3 (7.6)	43	*+178.3%*	3.44 (0.07)
Chance of itching (0.01–0.1%)	9.9 (22.6)	53	1.1 (5.0)	39	*+800.0%*	**5.64 (0.02)**
Chance of dizziness or fits (<0.01%)	7.0 (16.8)	49	1.5 (9.5)	40	*+366.7%*	3.39 (0.07)
Chance of ANY side-effect	38.0 (33.9)	49	17.6 (19.7)	41	*+115.9%*	**11.55 (0.001)**
Likert scale items						
Satisfaction with side-effect information	4.4 (1.3)	62	4.5 (1.3)	54		0.65 (0.80)
How bad overall are Taxol side-effects	3.5 (0.9)	62	3.4 (1.2)	54		0.26 (0.61)
Likelihood of having a side-effect	4.0 (1.3)	62	3.7 (1.6)	54		1.40 (0.24)
General risk to health from Taxol	3.2 (0.9)	62	3.2 (1.1)	54		0.14 (0.91)
Impact of information on decision to take Taxol	2.8 (1.6)	62	2.4 (1.3)	53		2.35 (0.13)

Bold text indicates statistical significance (p <.05).

Table 8 Participants with cancer (n = 116) responses (mean, SD), according to 'may affect' or 'will affect' risk qualifier terms

	Format					
	'May affect' risk qualifier		'Will affect' risk qualifier		Relative risk estimate increase: 'may affect' > 'will affect'	ANOVA
	M (SD)	*n*	*M* (SD)	*n*		*F*-value (*P*)
Side-effect risk estimates (actual %)						
Chance of aching joints & muscles (60%)	24.3 (28.2)	49	18.2 (21.9)	44	+33.5%	1.33 (0.25)
Chance of severe anaemia (6%)	15.5 (18.0)	49	13.9 (16.4)	46	+11.5%	0.20 (0.66)
Chance of serious allergic reaction (0.1–1%)	5.4 (12.5)	47	3.4 (8.4)	44	+58.8%	0.79 (0.38)
Chance of itching (0.01–0.1%)	5.3 (16.2)	49	7.2 (20.0)	43	−26.4%	0.25 (0.62)
Chance of dizziness or fits (<0.01%)	4.0 (12.4)	47	5.2 (16.2)	42	−23.1%	0.17 (0.68)
Chance of ANY side-effect	30.5 (30.8)	45	26.9 (29.3)	45	+13.4%	0.34 (0.56)
Likert scale items						
Satisfaction with side-effect information	4.6 (1.3)	57	4.3 (1.3)	59		1.46 (0.23)
How bad overall are Taxol side-effects	3.4 (1.1)	57	3.4 (1.0)	59		0.04 (0.85)
Likelihood of having a side-effect	3.9 (1.4)	57	3.8 (1.5)	59		0.30 (0.86)
General risk to health from Taxol	3.2 (1.0)	57	3.2 (1.0)	59		0.02 (0.89)
Impact of information on decision to take Taxol	2.6 (1.4)	57	2.5 (1.5)	58		0.52 (0.82)

(see Tables 7 and 8), with greater risk estimates made by those given the combined verbal and numerical risk expressions, although fewer comparisons reached statistical significance (most likely due to a lack of statistical power). The participants with cancer made similar ratings on the Likert measures. On the 'may affect' and 'will affect' comparisons, there were no differences between the groups in risk frequency estimates or Likert measures.

Discussion

In this first evaluation of all the European Medicines Agency's recommended risk expression terms, the use of combined verbal and numerical risk expressions led to significant overestimations of risk associated with the drug paclitaxel (Taxol), compared to numerical-only expressions. The pattern was repeated among a subsample, comprising participants with cancer, although the subsample size was small and differences in this group were mostly not statistically significant. Use of the recommended expression '..may affect ...' and an alternative expression '..will affect...' resulted in similar risk estimates.

The study methods allowed participants to track back to view the risk expressions when making their ratings, so reducing any concern that ratings were affected by recall bias. Furthermore, the remote, computer-stored location of the trial meant that it had methodological strengths, such as concealment of allocation, blinding of outcome measurement and a lack of contamination between the tested interventions. The study presented a hypothetical scenario, asking people to estimate their behaviour and risk perceptions, which is a potential weakness. However, the legal requirement to provide side-effect information in PILs means that a study testing different risk expressions in real PILs would be hard to achieve. Furthermore, the sample was recruited via web pages containing information, help and advice on medicines used in cancer and most participants themselves had cancer or had a relative or friend with cancer, increasing the external validity of the study. Three-quarters of the sample were university graduates and/or had a professional qualification, making it untypical of the wider populations served by EMA recommendations. But there seems no reason to think that this aspect would explain the pattern of results; indeed, a sample with more typical population levels of health literacy and numeracy may have produced greater overestimations of risk associated with the use of verbal terms.[16,17]

The use of the five verbal risk expression terms (*very common* to *very rare*) in PILs appears problematic, as their combination with numerical terms (in this case frequency bands) resulted in significantly increased perceived risks; this might impact on patients' decisions on medicine taking. It is also notable that the variation in risk estimates is very high, confirming the difficulties in achieving a shared understanding of risk expressions, whatever format is used. The levels of variance are also greater in the combined expression group than in the numerical-only group, illustrating the differing interpretation of verbal quantifier terms – what has been termed the elasticity of language. Similar effects have been noted in probabilistic statements about climate change, a very different setting with much higher chances of occurrence. In that case, verbal terms (ranging from *virtually certain* to *exceptionally unlikely*) were interpreted in a highly variable way by people both within and between 24 countries and 17 language groups.[18] This illustrates a further problem of verbal terms, such as those recommended by international organizations such as the EMA: it can be hard to achieve equivalence across languages.

It is notable that the mean risk estimates of participants in the numerical-only group were also higher than the actual risks on 4 of the 5 tested side-effects, suggesting that there should be further evaluation of the use of frequency bands and possible alternatives. However, there are three points to note. The first is that overestimation may be a statistical artefact resulting from low incident rates and therefore a greater potential for people to over-estimate rather than the opposite. Secondly, risk perception is notably heterogeneous[19] and achieving universal (and shared) understanding of any risk expression may be unrealistic. Lastly, the process of obtaining reports of adverse effects associated with medicines means that incident rates are necessarily imprecise, particularly for those that occur infrequently; communicating the *fact* of imprecision and potentially also the reasons for it, may be an important aspect of risk communication to medicine users.

Further studies are needed to evaluate the EMA-recommended risk expressions in other

medicines and, if replication of this study is achieved, it would suggest that the combined verbal and numerical terms should be withdrawn from recommendations. The information that is provided to patients, particularly from credible sources such as government agencies and professional bodies, can impact on their attitudes and their behaviour, in this case what they think about their medicines and whether or not they will take them. It is essential therefore that recommendations made about patient information are based on evidence of interpretation and understanding.

Source of funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The participants have no potential conflict of interests to declare.

Acknowledgements

We thank Dave Horton for his help in data collection for this study.

Trial Registration

ISRCTN39432352.

References

1 Berry DC, Gillie T, Banbury SP. What do patients want to know about their medicines? An empirical study *Expert Systems with Applications*, 1995; **8**: 419–428.
2 Ziegler DK, Mosier MC, Buenaver M, Okuyemi K. How much information about adverse effects of medication do patients want from physicians? *Archives of Internal Medicine*, 2001; **161**: 706–713.
3 European Commission. *A Guideline on the Readability of the Label and Package Leaflet of Medicinal Products for Human Use*. Brussels: EC Pharmaceuticals Committee, 1998.
4 Medicines and Healthcare products Regulatory Agency (MHRA). *Always Read the Leaflet – Getting the Best Information With Every Medicine*. London: The Stationery Office, 2005.
5 Carrigan N, Raynor DK, Knapp P. Adequacy of patient information on adverse effects: an assessment of patient information leaflets in the UK. *Drug Safety*, 2008; **31**: 305–312.
6 Knapp P, Raynor DK, Berry D. Comparison of two methods of presenting risk information to patients about the side effects of medicines. *Quality and Safety in Healthcare*, 2004; **13**: 176–180.
7 Knapp P, Raynor DK, Woolf E, Gardner PH, Carrigan N, McMillan B. Communicating the risk of side effects to patients: an evaluation of UK regulatory recommendations. *Drug Safety*, 2009; **32**: 837–849.
8 Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal*, 2004; **327**: 741–744.
9 European Commission. A Guideline on Summary of Product Characteristics (SmPC): Annex 1, 2009. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Template_or_form/2009/10/WC500004368.pdf, accessed 3 June 2014.
10 Nunes V, Neilson J, O'Flynn N *et al. Clinical Guidelines and Evidence Review for Medicines Adherence: Involving Patients in Decisions About Prescribed Medicines and Supporting Adherence*. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners, 2009.
11 Wallsten TS, Budescu DV, Zwick R, Kemp SM. Preferences and reasons for communicating probabilistic information in numerical or verbal terms. *Bulletin of the Psychonomic Society*, 1993; **31**: 135–138.
12 Berry DC, Knapp P, Raynor DK. Provision of information about drug side-effects to patients. *Lancet*, 2002; **359**: 853–854.
13 Knapp P, Gardner PH, Raynor DK, Woolf E, McMillan B. Perceived risk of tamoxifen side effects: a study of the use of absolute frequencies or frequency bands, with or without verbal descriptors. *Patient Education & Counseling*, 2010; **79**: 267–271.
14 Knapp P, Gardner PH, Carrigan N, Raynor DK, Woolf E. Perceived risk of medicine side effects in users of a patient information website: a study of the use of verbal descriptors, percentages and natural frequencies. *British Journal of Health Psychology*, 2009; **14**: 579–594.
15 Knapp P, Gardner PH, McMillan B, Raynor DK, Woolf E. An evaluation of a combined (percentage and frequency) risk expression to communicate medicine side effects. *International Journal of Pharmacy Practice*, 2013; **21**: 226–232.
16 Kondilis BK, Kiriaze IJ, Athanasoulia AP, Falagas ME. Mapping health literacy research in

the European Union: a bibliometric analysis. *PLoS One*, 2008; **3**: e2519.
17 Galesic M, Garcia-Retamero R. Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, 2010; **170**: 462–468.
18 Budescu DV, Por H-H, Broomell SB *et al*. The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, 2014; **4**: 508–512. doi:10.1038/NCLIMATE2194.
19 Teigen KH, Brun W. Verbal expressions of uncertainty and probability. In: Hardman D, Macchi L (eds) *Thinking: Psychological Perspectives on Reasoning, Judgement and Decision Making*. London: John Wiley, 2003: 125–145.