

People's Understanding of Verbal Risk Descriptors in Patient Information Leaflets: A Cross-Sectional National Survey of 18- to 65-Year-Olds in England

Rebecca K. Webster^{1,3,4}  · John Weinman^{1,2} · G James Rubin^{1,3,4}

Published online: 11 May 2017

© The Author(s) 2017. This article is an open access publication

Abstract

Introduction Evidence suggests the current verbal risk descriptors used to communicate side effect risk in patient information leaflets (PILs) are overestimated.

Objectives The aim was to establish how people understand the verbal risk descriptors recommended for use in PILs by the European Commission (EC), and alternative verbal risk descriptors, in the context of mild and severe side effects.

Methods A cross-sectional online survey was carried out by a market research company recruiting participants aged between 18 and 65 years living in England. Data were collected between 18 March and 1 April 2016. Participants were given a hypothetical scenario regarding the risk of mild or severe medication side effects and asked to estimate how many out of 10,000 people would be affected for each of the verbal risk descriptors being tested.

Results A total of 1003 participants were included in the final sample. The risks conveyed by the EC recommended

verbal risk descriptors were greatly overestimated by participants. Two distinct distributions were apparent for participant estimates of side effect risks: those for 'high risk' verbal descriptors (e.g. 'common', 'likely', 'high chance') and those for 'low risk' verbal descriptors (e.g. 'uncommon', 'unlikely', 'low chance'). Within these two groups, the distributions were near to identical regardless of what adverb (e.g. very, high, fair) or adjective (e.g. common, likely, chance) was used. The EC recommended verbal risk descriptors were more likely to be understood in accordance with their intended meanings when describing severe side effects. Very few demographic or psychological factors were consistently associated with how well participants understood the EC recommended verbal risk descriptors.

Discussion The current verbal risk descriptors used in PILs are ineffective at best and misleading at worst. Discontinuing the use of verbal risk descriptors would limit the likelihood of people overestimating the risk of side effects.

Electronic supplementary material The online version of this article (doi:[10.1007/s40264-017-0542-1](https://doi.org/10.1007/s40264-017-0542-1)) contains supplementary material, which is available to authorized users.

✉ Rebecca K. Webster
rebecca.webster@kcl.ac.uk

¹ NIHR Health Protection Research Unit in Emergency Preparedness and Response, King's College London, London, UK

² Institute of Pharmaceutical Science, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

³ Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

⁴ Department of Psychological Medicine, Weston Education Centre, King's College London, Cutcombe Road, London SE5 9RJ, UK

Key Points

This is the first study using a representative sample to show that people greatly overestimate the intended frequency of the verbal risk descriptors used to label side effect risk in patient information leaflets, especially when describing mild side effects.

Small changes to the wording used in the verbal risk descriptors will not solve this problem; the issue was the same for the three different forms of wording that we tested.

More radical changes (including abandoning the use of verbal risk descriptors) should be considered.

1 Introduction

In Europe, all medicines prescribed or sold over-the-counter must be accompanied by a comprehensive patient information leaflet (PIL), which is required to present the risk of potential side effects in ‘clear and understandable terms for the patient’ [1]. Over 70% of patients receiving a drug for the first time will read the accompanying PIL [2].

Effectively communicating information about the risk of side effects is difficult [3]. In 1998, European Commission (EC) guidelines suggested grouping adverse effects within a PIL according to five frequency bands and using a verbal label for each one [4]. Side effects could be ‘very common’ (experienced by more than one in ten patients), ‘common’ (up to one in ten, ‘uncommon’ (up to one in 100), ‘rare’ (up to one in 1000), or ‘very rare’ (up to one in 10,000). This way of quantifying the risk of adverse reactions was originally suggested by the Council for International Organizations of Medical Sciences (CIOMS) Working Group III in 1995 [5], and therefore is also applicable outside the European Union. Although not based on any empirical evidence [6], these EC recommended verbal risk descriptors have become widely used. Of the 50 most frequently dispensed medicines in England and Wales in 2012, 76% of the PILs used these verbal risk descriptors [7].

Since the guidelines were published, several studies have suggested that these verbal risk descriptors are problematic [8, 9]. It has been shown that UK students overestimate the risk associated with each of the verbal risk descriptors [10]. In another UK study, only seven out of 180 participants provided probability estimates for the verbal risk descriptors ‘common’ and ‘rare’ that fell within the EU guideline’s frequency range [11]. Studies with patients have found similar overestimations [12, 13], as has research with physicians, pharmacists and lawyers [14, 15]. In part, estimations seem to depend on side effect type, with ‘mild’ side effects given higher estimations than ‘severe’ side effects when described as ‘common’ or ‘rare’ [11].

Since these findings, the guidelines have been updated and suggest PILs should combine the verbal and numerical expressions (e.g. ‘very common, more than 1 in 10 people’) [16]. However, research has shown that this may not lead to more accurate side effect risk estimates than the verbal format [17] and still leads to significant risk overestimations when compared with numerical frequency bands alone [18]. The continued use of the same ‘very common’ to ‘very rare’ wording appears to still present problems for the public, even despite the addition of numerical information.

These findings are troubling. If patients systematically overestimate the risk of side effects from their medications, this may reduce their adherence and also increase the risk of symptoms occurring as a result of a nocebo effect [19]. Previous studies have limitations, however, often being based on small samples that are not representative of the general population. No study so far has used a national representative sample. Nor has any study sought to identify whether there are demographic or psychological factors that make individuals more or less likely to correctly estimate the numerical risk represented by these verbal descriptors. Psychological characteristics such as beliefs about medicines, optimism or perceived sensitivity to medicines are associated with medicine side effect expectations [20, 21] and may also affect how people perceive the risk information about medication.

In this study, we used a large cross-sectional survey of a representative sample of 18- to 65-year-olds in England in order to:

1. Assess how people interpret the risk associated with the EC recommended verbal risk descriptors and two sets of alternative verbal risk descriptors.
2. Investigate if people interpret the risk associated with verbal risk descriptors differently depending on whether they describe a mild or severe side effect.
3. Determine whether demographic or psychological factors are associated with correctly interpreting the risk implied by the EC recommended verbal risk descriptors.

2 Methods

2.1 Design

We commissioned the market research company Ipsos MORI to conduct an online survey of adults aged between 18 and 65 years living in England. Data collection took place between 18 March and 1 April 2016. This study was approved by the Psychiatry, Nursing and Midwifery Research Ethics Committee at King’s College London (reference HR-15/16-2104).

The same study was used to assess in detail factors associated with patients’ expectations of side effects conveyed by verbal labels of risk, the results of which have been submitted elsewhere.

2.2 Participants

Ipsos MORI recruited participants from an existing panel of people willing to take part in internet surveys living in England (approximately $n = 160,000$). We excluded over

65 s because of a concern that older adults who are members of an internet survey panel are not representative of the general population of older adults [22, 23]. Potential participants were emailed a link to the survey. After providing informed consent and clicking through to begin the survey, participants were allocated to receive questions about either mild or severe side effects. This was decided by the survey software on the basis of which condition had the lowest number of completed responses at that time. Panel participants typically receive points for every survey they complete; for our survey, participants received points equivalent to 75 pence.

2.3 Sample Size

Quotas based on participant age and gender (interlocked), location, and working status were used to ensure that the sample reflected the known demographic profile of adults aged 18–65 in England, according to data from the National Readership Survey [24]. We intended to recruit 1000 participants to provide us with a sample error of about plus or minus 3% at the total sample level.

2.4 Questionnaire Development

Where possible, we used or adapted items that had been previously developed and tested for their reliability and validity. We piloted all items with five members of the general public and rephrased items where necessary to improve clarity.

2.5 Primary Outcome: Verbal Risk Descriptor Probability Estimates

We included several items to assess participant understanding of the verbal risk descriptors. These asked people to consider a PIL for an unnamed drug that stated, for example, that 'nausea is common'. Items then asked participants to estimate how many out of 10,000 people who take the drug would develop that side effect. We chose a consistent denominator for each item to prevent confusion among participants had denominators changed in each question. To prevent the need for participants to give responses of less than one, we decided that a denominator of 10,000 would be needed. Our choice to get participants to use numbers is supported by Schwartz and Woloshin [25], who argue that despite concerns that people do not understand the use of numbers, representative populations are competent in using numbers as a decision aid for choosing between two drugs [26].

Participants were asked about either mild side effects ('headache' or 'nausea') or severe side effects ('seizure' or 'difficulty breathing') depending on which condition they

had been assigned to. We used the EC recommended verbal risk descriptors 'very common', 'common', 'uncommon', 'rare' and 'very rare' supplemented with the terms 'very uncommon' and 'extremely rare'. We also included other terms used in the risk communication literature [27–29], which could be combined into similar scales based on likelihood (very likely, likely, somewhat unlikely, fairly unlikely, very unlikely, extremely unlikely) and chance (very high chance, high chance, fair chance, low chance, very low chance, extremely low chance). Verbal risk descriptors were presented in a random order to participants.

2.6 Demographic Factors

We asked participants about their age, gender, ethnicity, highest level of education, employment status and whether they or anyone in their household had a long-standing illness, disability or infirmity.

2.7 Psychological Factors

We adapted the single-item literacy screener [30] assessing health literacy and asked participants to rate how often they needed help reading PILs, and included one item that asked how often they read PILs when taking a new medication. Both were rated from 1 ('never') to 5 ('always'). We used one item from the Health Anxiety Inventory [31] to assess health anxiety, which asked participants to select one of four statements describing their feelings over the past 6 months. These ranged from 1 'I do not worry about my health' to 4 'I spend most of my time worrying about my health'.

We assessed optimism using the Revised Life Orientation Test [32], which consists of six items (plus four filler items), giving a total score from 5 to 30, with higher scores indicating higher optimism. Participant beliefs about medicines were assessed using the overuse and harm general subscales of the Beliefs about Medicines Questionnaire (BMQ) [33], providing scores from 5 to 20 for each subscale, with higher scores indicating higher perceived overuse or harm. Finally, how sensitive participants thought they were to medicines was assessed using the Perceived Sensitivity to Medicines scale [34], giving a score from 5 to 25, with higher scores indicating higher perceived sensitivity.

2.8 Analysis

We grouped side effect frequency estimates given by participants into the same bands as those suggested for use by the EC guidelines [4] and produced histograms to show the

frequency with which each band was selected for each verbal descriptor.

We carried out a series of Mann–Whitney *U* tests to test if participants' median estimates differed between mild and severe side effects.

We used a series of multinomial logistic regressions to test if any demographic or psychological characteristics were associated with participants under-, over- or correctly estimating the numerical risk associated with the EC recommended verbal risk descriptors. Participants' numerical risk estimates were first recoded as an underestimate, overestimate or correct estimate for each descriptor. For the verbal risk descriptors of very common and very rare (where it is not possible to overestimate or underestimate, respectively), binary logistic regressions were carried out instead. For each of the regressions, all demographic variables and side effect types (mild or severe) were added to the regression in one block, and each psychological variable was added on its own, controlling for the previously entered variables.

For all analyses, answers of 'don't know' or 'prefer not to say' were excluded. A maximum of 3% of participants answered 'don't know' for any one question where this was an option, and this was 1% for 'prefer not to say'. All analyses were carried out using SPSS 22. Because the frequency of participants' side effect numerical estimates for each of the verbal risk descriptors did not change by more than 0.2% when using data weighted by age, gender, region and working status, we used unweighted data for our analyses.

3 Results

3.1 Sample Characteristics

A total of 1003 participants completed the survey and were included in the final sample (see Fig. 1 for response rates). Demographic information for the participants is given in Table 2.

3.2 People's Interpretation of the Verbal Risk Descriptors

Figure 2 shows the frequency of participants' numerical risk estimates for each verbal descriptor. Two distinct distributions were apparent; those for 'high risk' verbal descriptors, which portrayed a side effect as very common, common, very likely, likely, very high chance, high chance, or fair chance and those for 'low risk' verbal descriptors, which portrayed a side effect as uncommon, rare, unlikely and so on. Within these two groups, the distributions were near to identical regardless of what

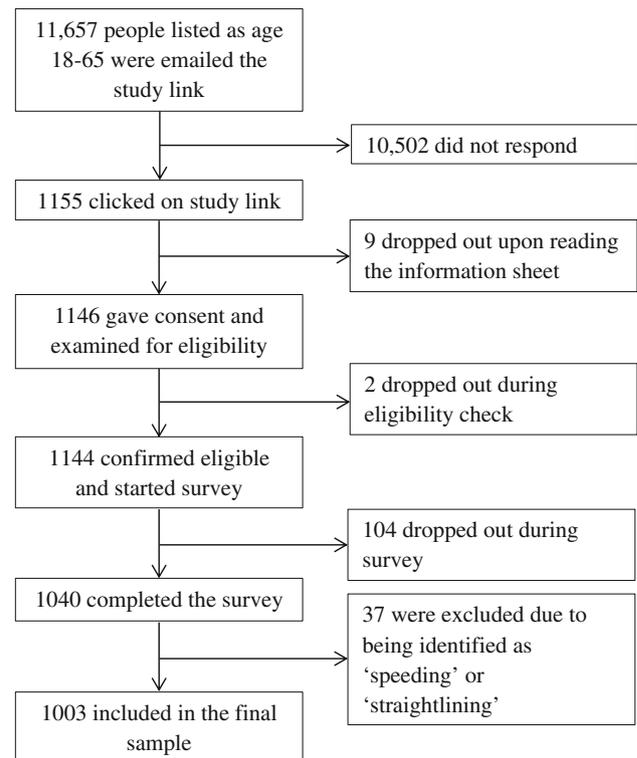


Fig. 1 Participant flow through the survey

adjective (e.g. common, likely, chance) or adverb (e.g. very, high, fair) was used.

For the 'high risk' verbal descriptors, most participants (84.4% and 81.2%, respectively) thought 'very common' and 'common' meant a risk of 1001–10,000 per 10,000 patients (i.e. more than one in ten) for mild side effects, with similar percentages seen for 'very likely' (83.8%), 'likely' (83.2%), 'very high chance' (85.2%), 'high chance' (84.0%), and 'fair chance' (74.7%). This pattern repeated itself for severe side effects, with the majority (56.3–71.2%) of participants giving estimates that corresponded to a risk of 1001–10,000 (more than one in ten). For 'low risk' verbal descriptors, the majority (61.8–76.7%) of participants provided estimates of 101–1000 (up to one in ten) or 11–100 (up to one in 100) for mild side effects, with this dropping to 43.5–70.6% for severe side effects.

3.3 The Effect of the Severity of Side Effects on People's Interpretation of the Verbal Risk Descriptors

Participants' numerical risk estimates were lower for each verbal descriptor when it described severe side effects compared to mild side effects (all *p* values <0.001, see Fig. 3).

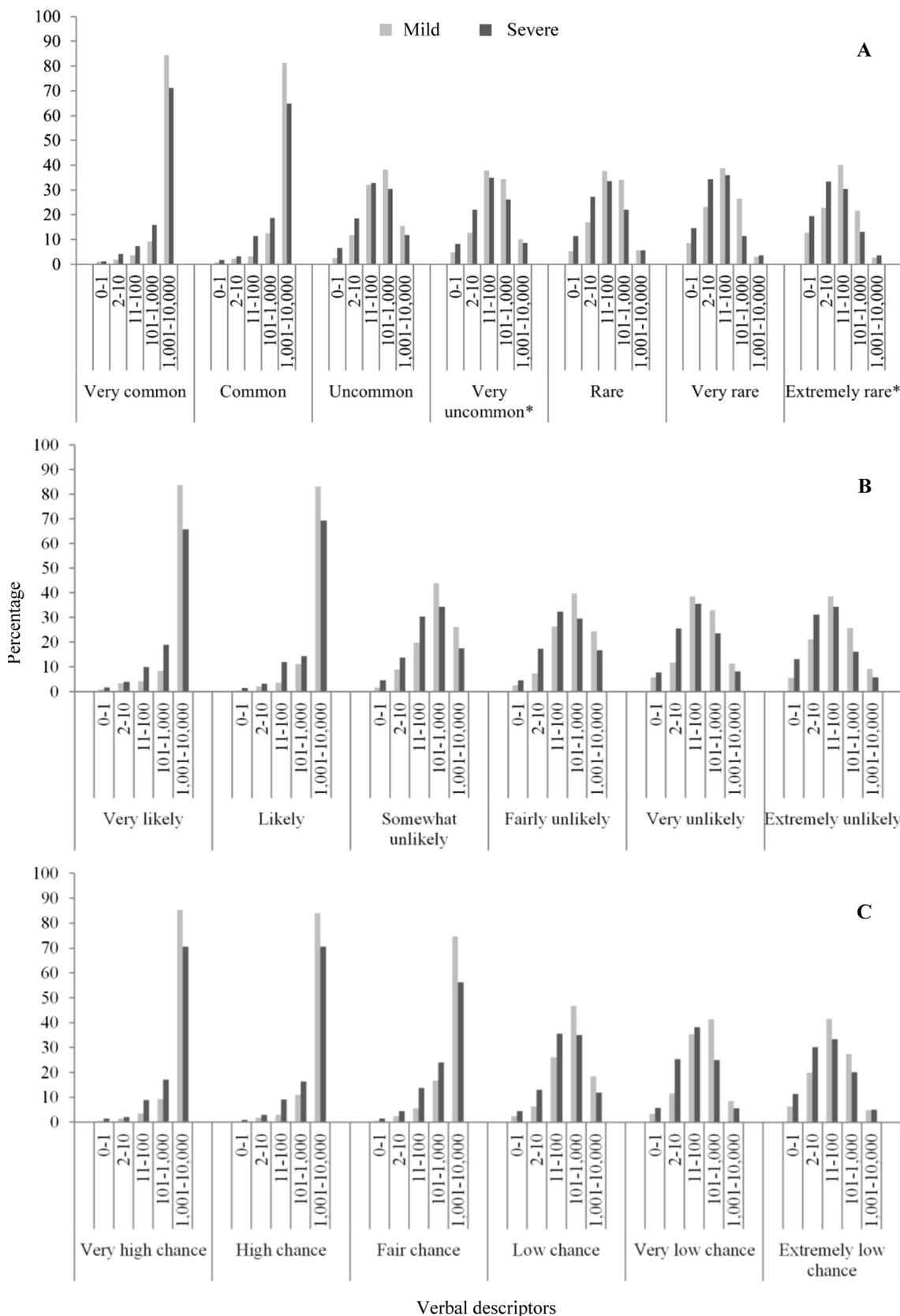


Fig. 2 Participants' estimates of the meaning for each verbal risk descriptor: **a** European Commission recommended verbal descriptors; **b** likely verbal descriptors; **c** chance verbal descriptors. *Added in for this study

3.4 The Association Between Demographic and Psychological Factors on People's Numerical Estimates for the EC Recommended Verbal Risk Descriptors

The proportions of participants giving correct, over- or underestimates for the EC recommended verbal risk descriptors are shown in Table 1. Table 2 shows the association between demographic and psychological variables with participants' numerical estimates. Older participants and those from ethnic minorities were generally less likely to overestimate the numerical risk of the verbal risk descriptors. Participants with no academic qualifications were 63% less likely than participants with university degrees to give correct estimates for 'very common', and 56% less likely than participants with degrees to overestimate 'common'. Participants who had someone in their household with a long-term illness or disability were 129% more likely than those without to underestimate 'rare'. By far the most influential factor was whether the descriptor related to mild or severe side effects. In general, mild side effects were more likely to be overestimated than severe side effects.

Most psychological characteristics had no association with whether participants estimated the numerical risk in accordance with the EC frequency bands. Optimism showed a small but significant effect for two verbal risk descriptors, with each 1-point increase in optimism resulting in participants being 4% less likely to overestimate 'uncommon' and 6% less likely to underestimate 'rare' descriptors. Belief about the harm of medicines also showed a small but significant effect for the 'very common' descriptor, with each 1-point increase in harm score resulting in participants being 6% less likely to give correct estimates. Finally, for each 1-point increase in health illiteracy, participants were 24% less likely to provide a correct estimate for 'very common'. Similarly, for each 1-point increase in health illiteracy, participants were 27% less likely to overestimate 'common' compared with estimating it in accordance with the EC frequency bands.

3.5 Post-hoc Analyses

As an additional analysis, we tested how much variance could be explained by entering the predictors altogether in one model. For predicting correct or incorrect estimates for the verbal risk descriptors 'very common' and 'very rare' using binary logistic regression, both models were a good fit for the data, with both Hosmer and Lemeshow tests being non-significant (both p values > 0.068). However, using Nagelkerke's R^2 , the models only explained 10.3% of the variance in estimates for 'very common' and 3.1% of the variance in estimates for 'very rare'. Similarly, for

predicting correct, over- or underestimates for the verbal risk descriptors 'common', 'uncommon' and 'rare' using multinomial logistic regressions, the models were a good fit for the data, with all χ^2 tests being non-significant (all p values > 0.232). Again, however, using Nagelkerke's R^2 , the models only explained 12.3% of the variance in estimates for 'common', 10.6% of the variance in estimates for 'uncommon' and 8.7% of the variance for 'rare' outcomes.

4 Discussion

4.1 Summary of Main Findings and Interpretation

There are several key findings from our work. First, the verbal risk descriptors recommended by the EC and commonly used in PILs are not accurately interpreted by members of the public as meaning the same things as the associated numerical expression. As shown in previous studies [10, 12, 14, 15], people widely overestimate what they mean in terms of frequencies. This overestimation of risk is important, making people less inclined to take their medication [10, 35] and leading to potentially self-fulfilling expectations of symptoms [36]. This overestimation generalises to other verbal risk descriptors not recommended by the EC; simply tweaking the wording of the verbal risk descriptors in use seems unlikely to resolve this problem. Instead, the issue may be a fundamental mismatch between how we are attempting to communicate (with five different divisions of frequency) and how people understand risk. Our data suggest people view verbal risk descriptors as meaning either 'likely' or 'not likely'—all descriptors are mentally reinterpreted as one of those two versions.

Second, it is hard to establish who will interpret the risk information in accordance with the corresponding EC frequency bands. Even combining all our predictors into one model did little to explain the variance in estimates across the different verbal risk descriptors. Only age and side effect type showed a consistent effect. This follows previous research that has shown that older people have lower perceptions of risk in general [37]. The finding that mild side effects were particularly likely to be overestimated has also been shown previously [11] and may reflect the influence of the availability of a heuristic [38]; given that people will have had more experience with headaches than seizures, it seems likely that people will find it easier to recall an example of a headache, elevating their perceived likelihood. Ethnicity, education and health illiteracy did have an effect, but only for the higher verbal risk descriptors. White participants, participants with academic qualifications and those who were more health literate were more likely to overestimate these high verbal risk

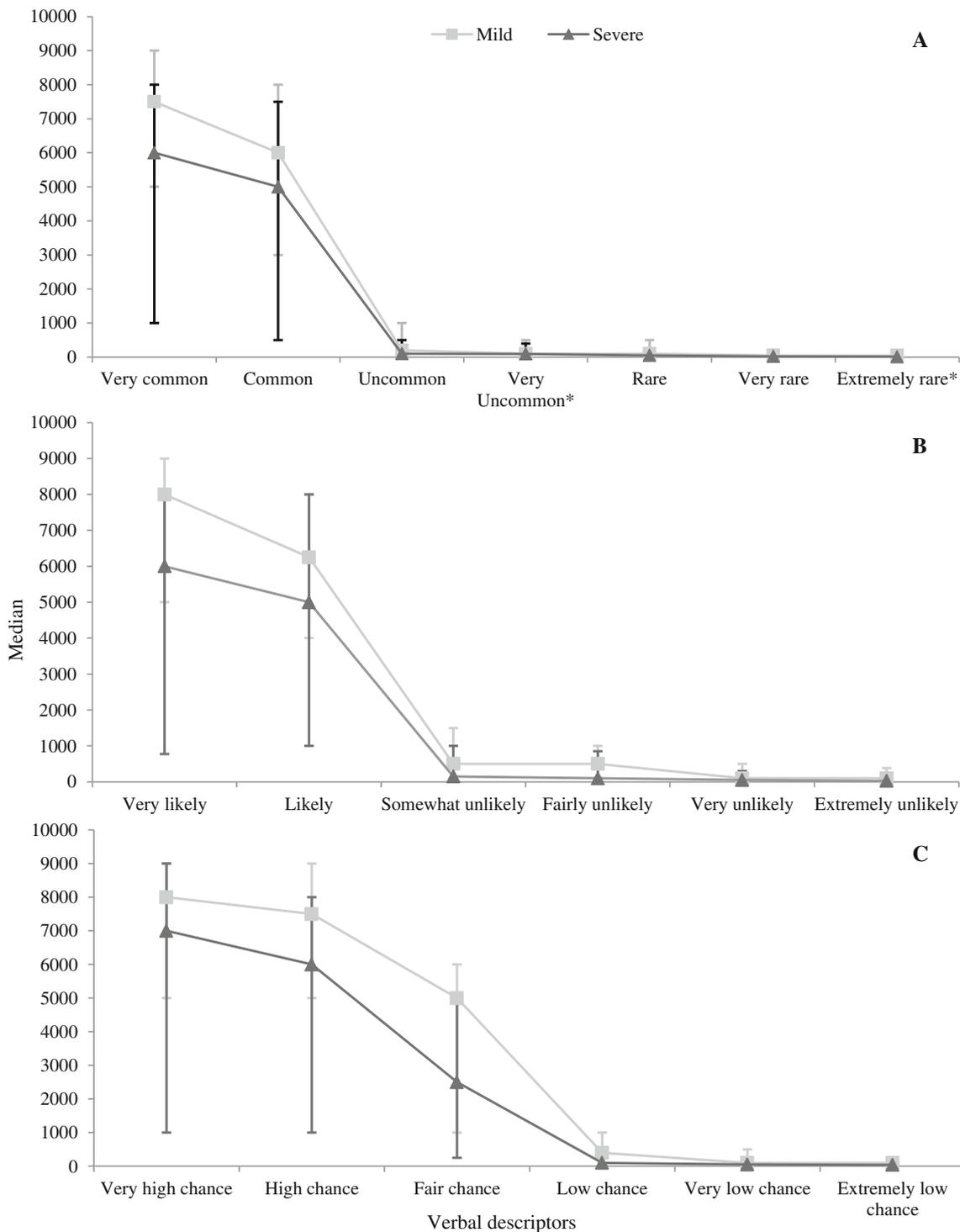


Fig. 3 Median estimates out of 10,000 given for each verbal risk descriptor: **a** European Commission recommended verbal descriptors; **b** likely verbal descriptors; **c** chance verbal descriptors. *Added in for this study. Bars represent the interquartile range

descriptors. It is possible that white participants are more likely to overestimate than participants from ethnic minorities, as it has been shown that they are prescribed more medications [39] and as such may have more

experience with medication side effects. It is surprising that participants with academic qualifications and those with higher health literacy were more likely to overestimate, and we are unsure why this might be.

Table 1 Frequency of how many people provided numerical risk estimates for each EC recommended verbal risk descriptor that were correct or incorrect according to the corresponding EC frequency bands

Verbal risk descriptor	Guess	Mild, <i>n</i> (%)	Severe, <i>n</i> (%)	Total, <i>n</i> (%)
Very common	Correct	427 (84.4)	354 (71.2)	781 (77.9)
	Incorrect (under)	79 (15.6)	143 (28.8)	222 (22.1)
Common	Under	31 (6.1)	82 (16.5)	113 (11.3)
	Correct	64 (12.6)	93 (18.7)	157 (15.7)
	Over	411 (81.2)	322 (64.8)	733 (73.1)
Uncommon	Under	71 (14.0)	125 (25.2)	196 (19.5)
	Correct	162 (32.0)	163 (32.8)	325 (32.4)
	Over	273 (54.0)	209 (42.1)	482 (48.1)
Rare	Under	27 (5.3)	57 (11.5)	84 (8.4)
	Correct	86 (17.0)	135 (27.2)	221 (22.0)
	Over	393 (77.7)	305 (61.4)	698 (69.6)
Very rare	Correct	44 (8.7)	72 (14.5)	116 (11.6)
	Incorrect (over)	462 (91.3)	425 (85.5)	887 (88.4)

EC European Commission

4.2 Implications for Side Effect Reporting Guidelines and Clinical Practice

Verbal risk descriptors have long been favoured for the presentation of side effect risk on the basis that they can break up long lists of side effects into more manageable sections based on frequency and convey the uncertainty of risk, and because some people are more comfortable with verbal than numerical information [40]. Current guidelines suggest PILs should combine the verbal risk descriptors with the numerical expressions [16]; however, we argue this is not enough, as it still leads to overestimation [17, 18]. If combined expressions are used, it remains important to use the correct verbal risk descriptor that is interpreted by people in the same way as the numerical expression that is associated with it. However, our survey has shown that verbal risk descriptors as a whole mislead rather than inform, leading readers to greatly overestimate their risk of side effects. We suggest that PILs should abandon the use of these verbal risk descriptors and instead side effect risk should be grouped under numerical frequency bands only. As well as having implications for PILs, the results of this survey also point out the need for clinical practitioners to reassure patients that side effects are much less likely than patients think. In addition, as mild side effects were overestimated more than severe, we suggest that practitioners may wish to focus in particular on correcting misperceptions about the likelihood of mild side effects.

4.3 Strengths and Weaknesses, and Future Research

This study is strengthened by its large sample size and the fact that it was demographically representative of 18-

to 65-year-olds in the English population. While it is possible to question the validity of the data as it is unknown if online participants read the questions properly or if they were distracted with other tasks whilst completing the study [41], this issue may not be as big as suspected [42], and was partly offset by our exclusion of participants for ‘straightlining’ or ‘speeding’. It is limited, however, in terms of selection bias, as we do not know whether members of market research panels are psychologically representative of the general population in terms of attitudes to medicines and their risk of side effects.

It is possible that the finding that psychological variables poorly predict participants’ estimates might be due to a lack of quality in the measures used to capture these variables. This is unlikely for optimism, belief about medicines, and perceived sensitivity to medicines, which were measured using well validated scales; however, health anxiety, health literacy and PIL reading behaviour were assessed by modifying validated scales and creating bespoke items for this study.

The response mode we chose for participants when estimating the numerical risk of the verbal risk descriptors also could have affected our results. Participants were asked to give a number out of 10,000; however, past research has suggested that open-ended questions such as this are more susceptible to risk overestimation compared with questions that require selecting an answer from a few different response options [43]. We chose this method to make it easier for participants to express small probabilities, and to allow participants to give their exact thoughts rather than having to choose from select options covering a broad range of answers. Nevertheless, we would be interested in future research to see how the results differ

Table 2 Demographic and psychological factors associated with how well participants guessed the numerical risk estimate for very common, common, uncommon, rare and very rare

Variable	No. (%) or median (IQR)	Verbal risk descriptors							
		Very common, correct; adjusted odd ratio (95% CI)	Common, underestimated; adjusted odd ratio (95% CI)	Common, overestimated; adjusted odd ratio (95% CI)	Uncommon, underestimated; adjusted odd ratio (95% CI)	Uncommon, overestimated; adjusted odd ratio (95% CI)	Rare, underestimated; adjusted odd ratio (95% CI)	Rare, overestimated; adjusted odd ratio (95% CI)	Very rare, correct; adjusted odd ratio (95% CI)
Demographic									
Gender ^a									
Male	492 (49.1)	0.97 (0.71–1.32)	1.00 (0.61–1.64)	1.12 (0.78–1.60)	0.74 (0.52–1.07)	1.15 (0.86–1.54)	1.32 (0.79–2.21)	1.33 (0.97–1.82)	0.71 (0.48–1.07)
Female	511 (50.9)	c	c	c	c	c	c	c	c
Age ^a	41.0 (22.0)	0.99 (0.97–0.998)	1.01 (0.99–1.03)	0.99 (0.98–1.01)	1.01 (0.99–1.02)	0.97 (0.96–0.98)	0.99 (0.97–1.01)	0.98 (0.97–0.99)	1.00 (0.98–1.02)
Ethnicity ^a									
Ethnic minorities	107 (10.7)	0.60 (0.37–0.97)	0.72 (0.33–1.55)	0.50 (0.30–0.84)	0.87 (0.48–1.59)	0.75 (0.46–1.19)	0.98 (0.44–2.20)	0.84 (0.51–1.41)	1.19 (0.64–2.20)
White	886 (88.3)	c	c	c	c	c	c	c	c
Employment ^a									
Not working	280 (27.9)	0.98 (0.67–1.39)	0.81 (0.46–1.41)	0.83 (0.56–1.24)	0.89 (0.59–1.34)	0.77 (0.55–1.07)	0.74 (0.40–1.35)	0.88 (0.61–1.25)	0.94 (0.59–1.51)
Working	723 (72.1)	c	c	c	c	c	c	c	c
Education ^a									
No qualifications	44 (4.4)	0.37 (0.19–0.74)	0.60 (0.19–1.89)	0.44 (0.21–0.93)	0.63 (0.25–1.62)	0.97 (0.47–1.97)	0.94 (0.18–4.88)	2.16 (0.91–5.18)	0.76 (0.25–2.26)
School qualifications	387 (38.6)	0.84 (0.60–1.16)	1.49 (0.88–2.52)	0.98 (0.67–1.44)	0.94 (0.64–1.37)	0.91 (0.67–1.24)	1.13 (0.66–1.94)	1.20 (0.87–1.68)	0.70 (0.45–1.08)
University degree	565 (56.3)	c	c	c	c	c	c	c	c
Household illness ^a									
Yes—me	290 (28.9)	0.88 (0.61–1.26)	0.82 (0.46–1.47)	0.87 (0.58–1.30)	0.91 (0.59–1.38)	1.25 (0.89–1.77)	0.96 (0.51–1.80)	0.98 (0.68–1.41)	0.89 (0.55–1.43)
Yes—someone else	128 (12.9)	0.88 (0.51–1.30)	1.40 (0.66–2.99)	1.12 (0.63–1.99)	0.72 (0.40–1.30)	1.04 (0.67–1.62)	2.29 (1.13–4.68)	1.12 (0.67–1.88)	0.88 (0.47–1.66)
No	571 (56.9)	c	c	c	c	c	c	c	c
Experimental									
Side effect type ^a									
Mild	506 (50.4)	2.16 (1.57–2.96)	0.54 (0.32–0.92)	1.81 (1.26–2.58)	0.57 (0.40–0.83)	1.32 (0.99–1.77)	0.75 (0.44–1.28)	2.02 (1.47–2.76)	0.59 (0.39–0.88)
Severe	497 (49.6)	c	c	c	c	c	c	c	c
Psychological									
Optimism ^b	19.0 (6.0)	1.00 (0.97–1.04)	0.97 (0.92–1.03)	0.99 (0.95–1.03)	0.98 (0.94–1.02)	0.96 (0.93–0.99)	0.94 (0.89–0.997)	0.97 (0.94–1.01)	0.99 (0.95–1.04)
PS to medicines ^b	10.0 (6.0)	0.99 (0.96–1.04)	0.99 (0.93–1.05)	1.01 (0.96–1.05)	1.02 (0.98–1.07)	1.04 (1.00–1.07)	0.95 (0.89–1.01)	1.00 (0.96–1.04)	1.00 (0.95–1.05)
BMC ^b	12.0 (4.0)	0.99 (0.94–1.04)	0.99 (0.91–1.08)	1.02 (0.96–1.08)	1.03 (0.97–1.10)	1.04 (0.99–1.10)	0.99 (0.91–1.08)	0.99 (0.94–1.04)	0.99 (0.92–1.06)
BMC harm ^b	10.0 (4.0)	0.94 (0.89–0.997)	1.01 (0.92–1.10)	0.96 (0.90–1.02)	1.00 (0.94–1.07)	1.01 (0.95–1.06)	1.01 (0.92–1.11)	0.98 (0.93–1.04)	1.01 (0.94–1.09)

Table 2 continued

Variable	No. (%) or median (IQR)	Verbal risk descriptors							
		Very common, correct; adjusted odd ratio (95% CI)	Common, underestimated; adjusted odd ratio (95% CI)	Common, overestimated; adjusted odd ratio (95% CI)	Uncommon, underestimated; adjusted odd ratio (95% CI)	Uncommon, overestimated; adjusted odd ratio (95% CI)	Rare, underestimated; adjusted odd ratio (95% CI)	Rare, overestimated; adjusted odd ratio (95% CI)	Very rare, correct; adjusted odd ratio (95% CI)
Healthy anxiety ^b	2.0 (0.0)	1.03 (0.79–1.34)	0.99 (0.64–1.52)	1.10 (0.81–1.49)	0.90 (0.65–1.25)	1.28 (0.99–1.65)	1.16 (0.74–1.82)	1.07 (0.82–1.41)	0.99 (0.70–1.40)
Health illiteracy ^b	1.0 (1.0)	0.76 (0.65–0.89)	0.96 (0.76–1.22)	0.73 (0.61–0.87)	1.13 (0.94–1.37)	0.94 (0.80–1.10)	0.97 (0.74–1.27)	0.91 (0.77–1.07)	1.03 (0.83–1.27)
PIL reading frequency ^b	4.0 (2.0)	1.12 (0.98–1.28)	1.05 (0.85–1.31)	1.12 (0.96–1.30)	0.98 (0.84–1.14)	1.07 (0.94–1.22)	0.83 (0.66–1.03)	0.95 (0.83–1.10)	0.99 (0.83–1.18)

Adjusted odd ratios are presented for those who provided correct estimates for 'very common' and 'very rare' descriptors and those who over/underestimated 'common', 'uncommon' and 'rare' descriptors in comparison to those who gave correct estimates. For unadjusted odd ratios, see the electronic supplemental material (online resource 1). Variables without a reference category were treated as continuous, and therefore the odds ratios represent the result of a 1-point increase in these variables. Bold values indicate $p < 0.05$

BMQ Belief about Medicines Questionnaire, *CI* confidence interval, *IQR* interquartile range, *PIL* patient information leaflet, *PS* perceived sensitivity

^a Controlled for each other

^b Controlled for demographic and experimental variables

^c Reference category

comparing these different types of response options for estimating the numerical risk of verbal descriptors.

In addition many of the questions used in the survey were hypothetical, e.g. estimating risk of side effects to an imagined drug. Future research should replicate this study with patients given a newly prescribed medication to remove any limitations relating to the hypothetical scenario used in our survey. We excluded over 65 s because of concerns about how representative they are in online surveys. However, over 65 s are the heaviest medication consumers [44]; therefore, extension of our findings to this age group would be useful. Research should also examine whether use of numerical, rather than verbal, descriptors produce more realistic risk estimates among participants. As with verbal risk descriptors, different numerical formats are possible (e.g. reframing the risk in terms of the number/proportion of people who remain side effect free). Identifying the best way of presenting this information remains an important goal.

5 Conclusion

Members of the public commonly overestimate the risk associated with verbal risk descriptors. It may be difficult, if not impossible, to find the perfect verbal risk descriptors that are interpreted by the public in line with the different levels of side effect risk. It may be that PILs should abandon the use of verbal risk descriptors altogether. This will limit the opportunity people have to overestimate the likelihood of side effects, allowing patients to make informed decisions about their medication and reducing the occurrence of side effects brought on from negative expectations, e.g. due to the placebo effect.

Acknowledgements We would like to thank the study participants for their time in completing this survey, and Louise, Steph and team from Ipsos MORI for carrying out the data collection.

Compliance with Ethical Standards

Funding The research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emergency Preparedness and Response at King's College London in partnership with Public Health England (PHE), and by PHE. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

Conflicts of interest Rebecca Webster, John Weinman and James Rubin declare that they have no conflict of interest.

Ethical approval This study was approved by the Psychiatry, Nursing and Midwifery Research Ethics Committee at King's College London (reference HR-15/16-2104). All procedures performed in studies involving human participants were in accordance with the ethical standards of King's College London and with the 1964 Helsinki declaration and its later amendments or comparable ethical

standards. However, we did not publish the protocol on a publicly accessible database as it is not industry standard for market research surveys to be registered in advance.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. The labelling of medicinal products for human use and on package leaflets. Brussels: European Commission Council Directive; 1992. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31992L0027&from=EN>.
2. Raynor DK, Silcock J, Knapp P, Edmondson H. How do patients use medicine information leaflets in the UK? *Int J Pharm Pract*. 2007;15(3):209–18. doi:10.1211/ijpp.15.3.0008.
3. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. *J Natl Cancer Inst*. 2011. doi:10.1093/jnci/djr318.
4. A guideline on the readability of the label and package leaflet of medicinal products for human use. Brussels: European Commission; 1998. <http://www.pharma-eu.com/pdfs/Guideline%20on%20Readability%20EMEA.pdf>.
5. Council for International Organizations of Medical Sciences Working Group 3. Guidelines for preparing core clinical-safety information on drugs: report of CIOMS Working Group III. Geneva: World Health Organisation; 1995. ISBN: 9290360623.
6. Knapp P, Gardner PH, Woolf E. Combined verbal and numerical expressions increase perceived risk of medicine side-effects: a randomized controlled trial of EMA recommendations. *Health Expect*. 2015. doi:10.1111/hex.12344.
7. Harris K, Dickinson R, Raynor DK, MacDonald J, Knapp P. Changes in side effect risk communication in patient information leaflets over the past decade: results of a survey. *Drug Saf*. 2015;38(8):721–7. doi:10.1007/s40264-015-0307-7.
8. Zipkin DA, Umscheid CA, Keating NL, Allen E, Aung K, Beyth R, et al. Evidence-based risk communication: a systematic review. *Ann Intern Med*. 2014;161(4):270–80. doi:10.7326/m14-0295.
9. Buchter RB, Fechtelpeter D, Knelangen M, Ehrlich M, Waltering A. Words or numbers? Communicating risk of adverse effects in written consumer health information: a systematic review and meta-analysis. *BMC Med Inf Decis Making*. 2014;14:76. doi:10.1186/1472-6947-14-76.
10. Berry DC, Knapp PR, Raynor T. Is 15 per cent very common? Informing people about the risks of medication side effects. *Int J Pharm Pract*. 2002;10(3):145–51. doi:10.1111/j.2042-7174.2002.tb00602.x.
11. Berry DC, Raynor DK, Knapp P. Communicating risk of medication side effects: an empirical evaluation of EU recommended terminology. *Psychol Health Med*. 2003;8(3):251–63. doi:10.1080/1354850031000135704.
12. Berry DC, Raynor DK, Knapp P, Bersellini E. Patients' understanding of risk associated with medication use. *Drug Saf*. 2003;26(1):1–11. doi:10.2165/00002018-200326010-00001.
13. Knapp P, Berry DC, Raynor DK. Testing two methods of presenting side effect risk information about common medicines. *Int J Pharm Pract*. 2001;9(S1):6. doi:10.1111/j.2042-7174.2001.tb01066.x.
14. Berry DC, Holden W, Bersellini E. Interpretation of recommended risk terms: differences between doctors and lay people. *Int J Pharm Pract*. 2004;12(3):117–24. doi:10.1211/0022357044120.
15. Ziegler A, Hadlak A, Mehlbeer S, König IR. Comprehension of the description of side effects in drug information leaflets: a survey of doctors, pharmacists and lawyers. *Deutsches Arzteblatt Int*. 2013;110(40):669–73. doi:10.3238/arztebl.2013.0669.
16. Guideline on the readability of the labelling and package leaflet of medicinal products for human use. Brussels: European Commission; 2009. http://ec.europa.eu/health/sites/health/files/files/eudralex/vol-2/c/2009_01_12_readability_guideline_final_en.pdf.
17. 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 Saf*. 2009;32(10):837–49. doi:10.2165/11316570-000000000-00000.
18. Knapp P, Gardner PH, Woolf E. Combined verbal and numerical expressions increase perceived risk of medicine side-effects: a randomized controlled trial of EMA recommendations. *Health Expect*. 2016;19(2):264–74. doi:10.1111/hex.12344.
19. Webster RK, Weinman J, Rubin GJ. A systematic review of factors that contribute to placebo effects. *Health Psychol Off J Div Health Psychol Am Psychol Assoc*. 2016;35(12):1334–55. doi:10.1037/hea0000416.
20. Faasse K, Petrie KJ. The placebo effect: patient expectations and medication side effects. *Postgrad Med J*. 2013;89(9):540–6. doi:10.1136/postgradmedj-2012-131730.
21. Geers AL, Helfer SG, Kosbab K, Weiland PE, Landry SJ. Reconsidering the role of personality in placebo effects: dispositional optimism, situational expectations, and the placebo response. *J Psychosom Res*. 2005;58(2):121–7. doi:10.1016/j.jpsychores.2004.08.011.
22. Internet users 2015. Office for National Statistics. 2015. <http://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2015>.
23. Coverage error in internet surveys who web-only surveys miss and how that affects results. Pew Research Center; 2015. <http://www.pewresearch.org/2015/09/22/coverage-error-in-internet-surveys/>.
24. National Readership Survey. 2015. <http://www.nrs.co.uk/2015>.
25. Schwartz LM, Woloshin S. The drug facts box: improving the communication of prescription drug information. *Proc Natl Acad Sci USA*. 2013;110(Suppl 3):14069–74. doi:10.1073/pnas.1214646110.
26. Schwartz LM, Woloshin S, Welch H. Using a drug facts box to communicate drug benefits and harms: two randomized trials. *Ann Intern Med*. 2009;150(8):516–27. doi:10.7326/0003-4819-150-8-200904210-00106.
27. Lichtenstein S, Newman JR. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychon Sci*. 1967;9(10):563–4. doi:10.3758/bf03327890.
28. Johnson EM. Numerical encoding of qualitative expressions of uncertainty. Alexandria: US Army Research Institute for the Behavioral and Social Sciences; 1973.
29. Beyth-Marom R. How probable is probable? A numerical translation of verbal probability expressions. *J Forecast*. 1982;1(3):257–69. doi:10.1002/for.3980010305.
30. Morris NS, MacLean CD, Chew LD, Littenberg B. The Single Item Literacy Screener: evaluation of a brief instrument to identify limited reading ability. *BMC Fam Pract*. 2006;7:21. doi:10.1186/1471-2296-7-21.
31. Salkovskis PM, Rimes KA, Warwick HM, Clark DM. The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. *Psychol Med*. 2002;32(5):843–53.

32. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a re-evaluation of the Life Orientation Test. *J Pers Soc Psychol.* 1994;67:1063–78.
33. Horne R, Weinman J, Hankins M. The Beliefs About Medicines Questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. *Psychol Health.* 1999;14(1):1–24. doi:[10.1080/08870449908407311](https://doi.org/10.1080/08870449908407311).
34. Horne R, Faasse K, Cooper V, Diefenbach MA, Leventhal H, Leventhal E, et al. The Perceived Sensitivity to Medicines (PSM) scale: an evaluation of validity and reliability. *Br J Health Psychol.* 2013;18(1):18–30. doi:[10.1111/j.2044-8287.2012.02071.x](https://doi.org/10.1111/j.2044-8287.2012.02071.x).
35. Woodward SC, Bereznicki BJ, Westbury JL, Bereznicki LR. The effect of knowledge and expectations on adherence to and persistence with antidepressants. *J Patient Prefer Adher.* 2016;10:761–8. doi:[10.2147/ppa.s99803](https://doi.org/10.2147/ppa.s99803).
36. Barsky AJ, Saintfort R, Rogers MP, Borus JF. Nonspecific medication side effects and the nocebo phenomenon. *JAMA.* 2002;287(5):622–7.
37. Savage I. Demographic influences on risk perceptions. *Risk Anal.* 1993;13(4):413–20. doi:[10.1111/j.1539-6924.1993.tb00741.x](https://doi.org/10.1111/j.1539-6924.1993.tb00741.x).
38. Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. *Cognit Psychol.* 1973;5(2):207–32. doi:[10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9).
39. Horne R, Graupner L, Frost S, Weinman J, Wright SM, Hankins M. Medicine in a multi-cultural society: the effect of cultural background on beliefs about medications. *Soc Sci Med.* 2004;59(6):1307–13. doi:[10.1016/j.socscimed.2004.01.009](https://doi.org/10.1016/j.socscimed.2004.01.009).
40. 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. *Br J Health Psychol.* 2009;14(3):579–94. doi:[10.1348/135910708X375344](https://doi.org/10.1348/135910708X375344).
41. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *Am Psychol.* 2004;59(2):93–104. doi:[10.1037/0003-066x.59.2.93](https://doi.org/10.1037/0003-066x.59.2.93).
42. Ramsey SR, Thompson KL, McKenzie M, Rosenbaum A. Psychological research in the internet age: the quality of web-based data. *Comput Hum Behav.* 2016;58:354–60. doi:[10.1016/j.chb.2015.12.049](https://doi.org/10.1016/j.chb.2015.12.049).
43. Woloshin S, Schwartz LM, Black WC, Welch HG. Women's perceptions of breast cancer risk. *Med Decis Making.* 1999;19(3):221–9. doi:[10.1177/0272989X9901900301](https://doi.org/10.1177/0272989X9901900301).
44. Health, social care and lifestyles. In: Craig R, Mindell J, editors. *Health Survey for England; 2013.* <http://content.digital.nhs.uk/catalogue/PUB16076/HSE2013-Ch1-Intro.pdf>.