Words or numbers? The evaluation of probability expressions in general practice

BERNIE J. O'BRIEN

SUMMARY. A sample of 56 general practitioners were asked to rate, on a percentage scale, 23 words or phrases which denote frequency or likelihood. The hypothetical context of the exercise was that of communicating to patients the probability of a side-effect (headache) arising from an unspecified prescription medicine. Median phrase ratings ranged from 'never' at 0% to 'certain' at 95% with a 50% rating given to the phrase 'reasonable chance'. Despite relatively large variance in ratings between respondents, the median ratings of a number of phrases were similar, and some identical, to other studies from different medical professionals. Although the clinical context in which a given expression of probability is used may affect its meaning, the results are encouraging and suggest that phrases denoting likelihood might be systematically codified to enhance communication between doctor and patient. To move towards this objective more research is needed to evaluate how patients interpret expressions of probability, and the relative effectiveness of different modes of communicating likelihood.

Introduction

GIVEN the many uncertainties which surround the practice of medicine, a common feature of communication is the use of expressions such as 'likely' or 'probable'. In the consultation, for example, such expressions may be used to convey to patients the chances of symptoms persisting or of particular side effects from prescribed medicines developing. Similarly, communications between doctors on aspects of diagnosis and referral often contain expressions of uncertainty. Such phrases can also be found in prescribing guides such as the British national formulary in connection with possible adverse effects. The problem is that the same expression may convey different degrees of likelihood to different individuals. If ambiguity in communication is to be minimized it is important to know the extent to which there is consensus regarding the probability level each phrase conveys, or as one study asked, 'how probable is probable?'[1]

The question is amenable to empirical testing and some studies have been undertaken where doctors numerically rated (for example on a percentage probability scale) particular words or phrases which denote probability.[2-6] The findings of such studies indicate that although there exists a consistent rank ordering of particular phrases expressing likelihood, the variability in values assigned to phrases is large enough to warrant further study, using different respondent groups and rating contexts, in order to move towards a system of codification for such phrases.

The aim of this study was to analyse numerical ratings for expressions of probability from a sample of general practitioners in the context of communicating the likelihood of side-effects occurring with prescription medicines. With the exception of

B.J. O'Brien, BA, MSc, research fellow, Health Economics Research Group, Brunel University.

Mapes,[5] who studied a small number of expressions, this respondent group and context have received little attention.

Method

The respondents were all practising general practitioners and members of a general practice clinical research group regularly convened by a major pharmaceutical company. At a meeting of the group the purpose of the study was explained and questionnaire forms distributed for later completion and return by post. A total of 56 completed forms were received from 65 distributed (86% response rate). Returned forms were anonymous and no data were collected on respondent characteristics so it was not possible to test whether non-responders differed systematically from responders.

Respondents were asked to give a percentage probability rating to a list of 23 words or phrases which could be used in a hypothetical situation to convey to a patient the probability of headache occurring as a side effect from a drug they had prescribed. No drug name or type was specified. The phrases were presented in a random order.

As a second exercise, respondents were asked to consider the 'variability in meaning' or ambiguity for each expression of probability. They were required to rate each phrase on a simple three point scale of high, medium or low variability in meaning. The aim of this part of the study was to determine the extent to which variability in judgements of meaning for each phrase were concordant with the observed statistical variability between respondents for probability ratings.

The statistical methods used were mainly descriptive, reporting medians and means with interquartile ranges as measures of the spread of responses. The Spearman rank correlation coefficient was used to measure the association between the ambiguity ratings and the interquartile ranges for each phrase.

Results

The presentation and interpretation of the results are subject to two caveats. First, the responses of this small sample of general practitioners may not be representative of the population of general practitioners in the UK or elsewhere. However, there were no a priori reasons to suppose that there was any systematic sampling bias which invalidated the general findings of this study. The second caveat concerns whether all respondents understood the nature of the exercise. We suspected, as Bryant and Norman did,[3] that some respondents misunderstood the instructions. In particular, the forms of four respondents were not included for analysis because they gave implausible answers, rating 'certain' below 10% and 'never' higher than 90%. In the majority of cases, however, responses and the comments received back did not indicate difficulty or misunderstanding.

The median and mean percentage rating of the 52 general practitioners for the 23 expressions are presented in Table 1. As expected, the extreme markers were 'never' and 'certain' with median probability ratings of 0% and 95% respectively. The middle ground of the probability distribution appeared to be denoted by phrases such as 'reasonable chance' (50%), with a 25% rating for 'possible' and 75% rating for 'probable'.

The impact of qualifying expressions with adverbs appeared to be consistent, with expressions such as 'chance' moving upwards through the rating scale in the following order: 'small

Table 1. Probability ratings of 23 phrases by 52 general practitioners.

| | Probability rating of phrase (%) | | | | |
	Median	Mean	Inter-quartile limits	Inter-quartile range	Mean ambiguity rating[a]
Never	0[b]	6	0–0.5	0.5	1.1
Almost never	3	14	1–5	4	1.4
Very rare	5	11	2–7	5	1.2
Low probability	10	14	5–20	15	1.9
Low risk	10	15	5–18	13	1.9
Small chance	10	16	5–15	10	2.0
Unlikely	13	19	10–20	10	2.1
There is a chance	15	23	10–30	20	2.5
Sometimes	23	28	10–50	40	2.6
Possible	25	30	20–40	20	2.6
Perhaps	28	31	18–50	32	2.6
Could be	30	35	23–50	27	2.7
Moderate risk	40	39	30–50	20	2.5
Not certain	50	42	20–50	30	2.6
Reasonable chance	50	49	33–60	27	2.6
Significant chance	60	49	23–70	47	2.3
Reasonable to assume	70	61	50–80	30	2.6
Likely	70	69	60–80	20	2.2
Probable	75	70	60–80	20	2.1
Most likely	80	72	67–86	19	2.0
Expected	80	75	70–90	20	1.7
Almost certain	90	86	90–95	5	1.6
Certain	95	84	90–100	10	1.4

[a] Categorical rating scale, 3 = high, 2 = medium, 1 = low. [b] Rounded down from 1×10^{-6}; this phrase was zero rated by 73% of respondents.

chance' (10%), 'there is a chance' (15%), 'reasonable chance' (50%), 'significant chance' (60%). The result of qualifying 'certain' with 'almost' was to reduce the median percentage rating from 95% to 90% and this 5% difference was similar to that observed in the study by Kong and colleagues[6] where 'certain' was rated at 99% and 'almost certain' at 94%.

Another result consistent with previous findings[6,7] was that phrases and their negatives were not complementary. For example, 'likely' was rated at 70% but 'unlikely' at 13% giving a total of 83% rather than the 100% that would be expected if there were symmetry in the use of the phrases. The same asymmetry was found by Kong and colleagues[6] with the median rating for 'likely' being 67% and 'unlikely' being 11%.

One measure of the degree of inter-respondent consensus on expression rating is the interquartile range. As might be expected there was less variability at the extremes of the probability scale than in the middle. The interquartile ranges for expressions in the middle of the probability distribution were particularly large; the ratings for 'significant chance', for example, ranged by 47% from 23 to 70% and 'sometimes' ranged by 40% from 10 to 50%.

One interpretation of the patterns of variation in expression ratings is that particular words or phrases (particularly in the mid-range of probability) are more ambiguous in meaning than others. Respondents' ratings on the ambiguity of expressions were broadly similar to observed statistical variation with the lowest mean rating (1.1) for 'never' and the highest mean rating (2.7) for 'could be'. The relationship between ambiguity ratings and the interquartile range of probability ratings is illustrated in Figure 1, indicating a strong positive correlation between the two ($r = 0.86$, Spearman rank correlation).

Discussion

In terms of moving towards a broad consensus of probability denoted by particular expressions, the results of this study, when compared with other studies, are encouraging. Table 2 presents results from two similar studies for six common probability expressions. The concordance between studies is strong, with the rank order of expressions being almost identical. It is interesting to note that at the extremes of the probability scale, although results from two studies indicate 'never' is 0%, the opposite (that is 100%) is not true of 'certain' which is rated between 95% and 99%. One interpretation of this result is that when rating 'certainty', doctors may think cautiously in terms of statistical confidence limits ($P<0.05$, $P<0.01$), but this does not explain why the same reasoning should not be applied to the other extreme of the distribution.

The second encouraging result from this study is the degree of correlation between observed and anticipated (by respondents) variability in numerical rating. Previous studies have observed
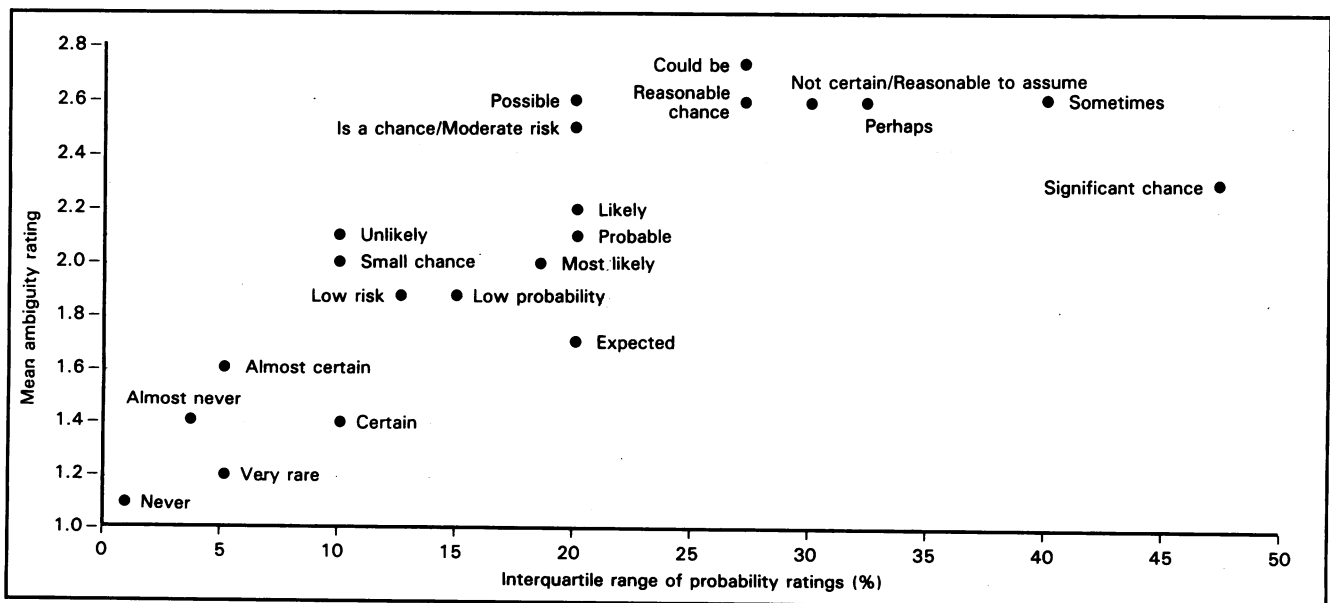

Figure 1. *Relationship between observed and predicted variability in meaning.*

Table 2. Probability ratings of six expressions from three studies.

	Probability ratings (%)		
	O'Brien (median)	Kong et al[6] (median)	Bryant and Norman[3] (mean)
Certain	95	99	95
Probable	75	70	77
Likely	70	70	73
Possible	25	20	47
Unlikely	13	11	20
Never	0	0	33[a]

[a] Bryant and Norman note that this value may be anomalous owing to misinterpretation of instructions by respondents.

high inter-respondent differences without relating this finding to awareness of the ambiguity inherent in each expression. Phrases like 'significant chance' varied widely in probability rating and respondents also rated this phrase as 'medium' to 'high' in terms of ambiguity of meaning. This raises the question whether the general practitioners' awareness of large variability in meaning influences his or her usage of such phrases, either in terms of the frequency of usage or in the use of numerical qualification or further explanation of the probability being conveyed.

A further confounding influence on interpretation and between-study comparison is the context in which the expression is to be used. The present context of the likelihood of side-effects using a prescription drug is clearly only one of many that can be used. Kong and colleagues,[6] for example, use the frequency of a particular symptom in a disease as the probabilistic magnitude to be expressed. The influence of context on such ratings was explored by Mapes[5] in an experiment where two groups of general practitioners rated the expression 'rare' in two scenarios: first the likelihood of side effects occurring with beta-blockers and secondly with antihistamines. In the first scenario 'rare' was rated by 59.4% of the group at 'less than 1 per 1000' although with the antihistamine group this rating for 'rare' was only given by 20.7% of the sample. One explanation for this difference is that prior knowledge and/or experience of the frequency and magnitude of side effects influences the evaluation of the probability expression. As Kong and colleagues[6] noted, however, although absolute ratings differ between contexts, the rank order of ratings on expressions within contexts is largely consistent.

The discussion of probability phrases being influenced by the clinical context in which they are used leads on to the consideration of patients' understanding of probability expressions. This is an important area for further research. While the general practitioner (and other medical professionals) may broadly agree that 'rare' used in one clinical context differs from its use in a different context, there is no obvious basis for the patient to discriminate between the different probabilities that the phrase is intended to convey in different situations.

Regarding methods which may enhance doctor–patient risk communication, the use of visual aids may be a useful way of getting the risk message across. Consider an example; an anxious mother is deliberating about whooping cough vaccination for her child because she is aware that it carries some risk of neurological damage. If we take the incidence of permanent neurological damage as being one child in every 100 000 receiving the full course of three injections,[8] then a visual aid to help communicate this frequency is a piece of graph paper with 100 000 squares and with one square shaded. Furthermore, to help place the risk in perspective, on the same graph paper could be illustrated the population background incidence of

neurological damage. Another example of clinical risk information communication which might benefit from such visual aids is the counselling of women facing prenatal diagnosis such as amniocentesis. In this situation the pregnant woman must weigh-up the risks of spontaneous abortion during the test versus the risk of giving birth to a baby with Downs syndrome if she declines the test. An important input into the decision making process is the format of the probability information from the clinician. A number of studies have reported the ways in which the perception of risks can be biased simply by the framing of the choice and the risk terminology used.[9-11]

The findings of this small study suggest more evaluative research is needed into doctor–patient risk communication, particularly in those areas of elective therapy where the patient is required to decide between alternative treatment strategies. The choice of communication method is not as simple as words or numbers because the two modes can be complementary. Arguments that 'verbal specifications of frequency have no place in medicine'[12] can be contrasted with survey results which indicate that numbers and statistics hold very little meaning for the average member of the public.[13] But to involve the consumer of health care more in decision making it is necessary to develop effective methods for communicating the risks and benefits of alternative options. Although the doctrine of informed consent to treatment is a well established principle of medical practice, the methods of informing patients about probabilities and risks is in need of further investigation.

References

1. Beyth-Marom R. How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting* 1982; **1:** 257-269.
2. Toogood JH. What do we mean by 'usually'? *Lancet* 1980; **1:** 1094.
3. Bryant GD, Norman GR. Expressions of probability: words and numbers. *N Engl J Med* 1980; **302:** 411.
4. Kenney RM. Between never and always. *N Engl J Med* 1981; **305:** 1097-1098.
5. Mapes REA. Verbal and numerical estimates of probability in therapeutic contexts. *Soc Sci Med* 1979; **13A:** 277-282.
6. Kong A, Barnett GO, Mostellor F, Yeutz C. How medical professionals evaluate expressions of probability. *N Engl J Med* 1986; **315:** 740-744.
7. Lichtenstein S, Newman JR. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science* 1967; **9:** 563-564.
8. National Childhood Encephalopathy Study. In: *Whooping cough: reports from the Committee on Safety of Medicines and the Joint Committee on Vaccination and Immunization.* London: HMSO, 1981.
9. McNeil BJ, Weichselbaum R, Sox HC, et al. On the elicitation of preferences for alternative therapies. *N Engl J Med* 1982; **306:** 1259-1262.
10. Eraker SA, Sox HC. Assessment of patients' preferences for therapeutic outcomes. *Med Decis Making* 1981; **1:** 29-39.
11. O'Brien BJ. *What are my chances doctor? A review of clinical risks.* London: Office of Health Economics, 1986.
12. Nakao MA, Axelrod S. Numbers are better than words: verbal specifications of frequency have no place in medicine. *Am J Med* 1983; **74:** 1061-1065.
13. Prescott-Clarke P, Mostyn B. Public attitudes towards the acceptability of risks. In: Prescott-Clarke P. *Public attitudes towards industrial work-related and other risks.* London: Social and Community Planning Research, 1982.

Address for correspondence

Mr B.J. O'Brien, Research Fellow in Health Economics, Health Economics Research Group, Brunel University, Uxbridge, Middlesex UB8 3PH.