Do mental health professionals really understand probability? Implications for risk assessment and evidence-based practice

TIM M. GALE[1,2,3], CHRIS J. HAWLEY[1,2] & THANUSHA SIVAKUMARAN[1]

[1]*Department of Psychiatry, QEII Hospital, Welwyn Garden City,*
[2]*Department of Psychology, University of Hertfordshire, Hatfield, UK &*
[3]*Department of Computer Science, University of Hertfordshire, Hatfield, UK*

Abstract

Background: Many studies have highlighted poor understanding of risk and probability in NHS service users. By contrast, few studies have explored whether health professionals share similar conceptual difficulties. The current study is an attempt to address this issue within the field of mental health.

Aims: The principal aim of the study is to test knowledge of basic probability across a range of Mental Health Professionals (MHPs) and to identify those aspects which they found difficult.

Method: MHPs sample were compared to age-, gender-, and education-matched control group to establish whether MHPs were more skilled in dealing with probabilistic concepts. The assessment addressed several key aspects of probability that are important in risk assessment and evidence-based evaluations.

Results: Contrary to our predictions, the MHP group was no more accurate than the control group and many conceptual problems were highlighted, ranging from basic arithmetic difficulties to problems in evaluating relevant information. Systematic error was common in all areas tested. More generally, age, educational level, gender and professional group all strongly predicted response accuracy.

Conclusions: Our findings are discussed principally in relation to risk assessment practice. Our data suggest that all MHPs would benefit from receiving greater training in probability theory, since risk is central to so many concerns in mental health services.

Declaration of interest: None.

Keywords: probability, risk assessment, mental health professional, evidence-based practice.

Address for Correspondence: Dr. Tim M. Gale, Department of Psychiatry, QEII Hospital, Howlands, Welwyn Garden City, Hertfordshire AL7 4HQ, UK. Tel: + 44 (0)1707 365383; Fax: + 44 (0)1707 365168; E-mail: t.gale@herts.ac.uk

Introduction

In today's evidence-based healthcare culture, probability is an integral part of treatment-related decision-making. For example, therapeutic interventions are now commonly discussed in terms of odds ratios, relative risks or numbers needed to treat. Such statistics are intended to facilitate an informed treatment choice based on an empirically-derived chance of success or failure.

Over the last decade, concerns have been raised about whether NHS service users have sufficient grasp of probability theory to understand and use the information they are presented with in medical contexts. Unfortunately, the literature is rather pessimistic about this and many conceptual difficulties and cognitive biases have been identified in patient groups. For example, perception of risk is modulated by 'availability bias' (see Tversky & Kahneman, 1974), whereby there is a tendency to over-estimate the risk of an outcome that has received particular notoriety: so, because breast cancer receives more media attention than many other forms of cancer, most women greatly over-estimate their risk of developing it (see Bogardus *et al.*, 1999). Understanding of risk is also strongly influenced by the way in which information is framed (Edwards *et al.*, 2001; Edwards & Elwyn, 2001): thus a prospective patient is more likely to opt for a treatment if told that it is associated with a 99% survival rate (positive framing) rather than a 1% mortality rate (negative framing), even though the actual risks are identical. Patients also differ widely in their understanding of statistical information, whether presented numerically (e.g. Fuller *et al.*, 2001) or by verbal descriptors such as 'normal' (Marteau *et al.*, 2001), 'probable' or 'certain' (Biehl *et al.*, 2001). It is therefore unsurprising that recent recommendations are for more graphical decision aids to be used in the context of treatment evaluation by doctors and their patients (for a review see O'Connor *et al.*, 1999).

Conceptual difficulties with probability in a medical context almost certainly reflect a widespread lack of statistical knowledge in the general population. Indeed, one can find parallels for medical choice dilemmas in everyday life: the recent reluctance of UK parents to have their children vaccinated with the MMR jab because of an equivocal association with autism suggests that the average person does not easily conceptualize long odds (see also Kirkwood, 1994); similarly, this is reflected in the number of people who participate in the national lottery despite the near certainty that that they will not win the jackpot. In contexts such as these, Dillon & Gill (2001) have presented the case for a radical re-appraisal of the national science curriculum to promote greater understanding of risk and probability.

Assuming then that conceptual barriers to probability and risk are commonplace, one might reasonably ask whether healthcare professionals *also* have difficulties with these issues. However, there is a dearth of research in this context and so the implicit assumption that health professionals are competent in dealing with concepts of risk may be ungrounded. With the advent of evidence-based healthcare, one might optimistically predict an increasing level of probabilistic knowledge amongst healthcare professionals. Routine clinical audit is now an obligatory part of any medical department's activity and so most professionals ought to have some experience

with numerical representation of health-care outcomes. Finally, and perhaps most importantly, many professionals, and especially those working within a mental health setting, are now expected to assess risk as part of normal clinical duty and this almost certainly entails some basic understanding of probability. In these respects then, it would be a reasonable prediction that mental health professionals should have a better than average understanding of probability and risk.

Turning specifically to risk assessment in mental health, this is an area which has been the focus of much recent attention following changes in service provision for people with enduring mental health problems and some highly publicised cases of homicide committed by psychiatric patients in the community. Indeed, legislation and guidelines in the mid-1990s in the UK reinforced the importance of thorough risk assessment and management (DoH, 1994; 1995; HMSO, 1995). Previous studies on the risk assessment of psychiatric patients has highlighted inter-professional variability in the levels of risk ascribed to specific cases (Montandon & Harding, 1984; Ryan, 1998). The gender of the assessor is also known to be a significant factor in determining whether ratings are more or less cautious (Ryan, 1998; Gale *et al.*, 2002). Moreover, in a previous study (Gale *et al.*, 2002) we demonstrated only moderate (and some-times worse) levels of inter-rater reliability for risk ratings of mentally ill patients, suggesting that mental health professionals (MHPs) may have difficulties both in conceptualizing *and* expressing risk information. Given that some probabilistic reasoning must underlie any given risk assessment (if only in

making a decision about the *relative* likelihood of an outcome), it would be pertinent to assess: (i) whether MHPs have conceptual difficulties with particular aspects of probability and risk and (ii) whether this might have implications for education and training, both in risk assessment and healthcare practice more generally.

Method

The Questionnaire

The questionnaire comprised 10 multiple-choice questions and addressed four broad aspects of simple probability that were considered relevant to evidence-based practice and risk assessment: (i) understanding the concept of randomness; (ii) interpretation and simple manipulation of probabilistic information; (iii) evaluation of outcomes using basic probability; and (iv) evaluation of whether sufficient information exists to make probabilistic inferences. Although the questionnaire was principally aimed at healthcare professionals, the use of clinical scenarios was minimized because: (i) a control group was also asked to participate and so we chose scenarios that would be familiar to all respondents; (ii) it was important for recruitment and ethical reasons that participants did not perceive this to be a vocationally-related test (it was not). The 10 questions are presented in Table 1.

The questions varied in difficulty to reduce the likelihood of floor or ceiling effects. However, no questions were considered to require specialist knowledge beyond that which might reasonably be expected of somebody who had studied O' level or GCSE maths. The questionnaire was initially presented to a group of individuals with experience in

Table 1: The 10 multiple-choice probability questions. These were presented in a standard multiple-choice questionnaire format with appropriate spacing and uniformity. A copy of the original questionnaire is available from the corresponding author on request and a description of the rationale for each question can be found in the appendix

1. An oil-rig worker is described as having a 2 in a thousand chance of being struck by lightning. This can also be expressed as: (a) 0.5% chance; (b) 0.2% chance; (c) 0.02% chance; (d) 0.05% chance.
2. A coin is tossed six times and you record whether it lands as a head or tail each time. Which of the following do you think is the most likely sequence you would see? (a) H T H T H T; (b) H H H H H H; (c) T H H T T H; (d) They are all equally likely.
3. Two dice are thrown simultaneously. What is the probability that both will show the same number when they land? (a) 1 in 6; (b) 1 in 36; (c) 1 in 12; (d) None of these.
4. A cloth bag contains 60 red sweets and 40 blue sweets. If you remove three sweets, one at a time, without looking at their colour, which is the most likely order of colours that you will pull out? (a) Red, Red, then Blue; (b) Red, Blue, then Blue; (c) Red, Red, then Red; (d) It is not possible to say which is most likely.
5. The six winning numbers on last week's National Lottery were: 9, 11, 22, 38, 44, 45. Which of the following sequences is most likely to occur this week? (a) 41, 42, 43, 44, 45, 46; (b) 9, 11, 22, 38, 44, 45; (c) 2, 17, 21, 29, 33, 48; (d) They are all equally likely.
6. Which of the following statements is most likely to be true? (a) Next week's National Lottery jackpot winner will have won the jackpot once before; (b) Next week's National Lottery jackpot winner will have a twin brother or sister; (c) Next week's National Lottery jackpot winner will not be able to drive; (d) It is not possible to say which is most likely.
7. Four playing cards are chosen at random from a shuffled pack. Which of the following scenarios is the most likely outcome? (a) The four chosen cards are all of the same suit; (b) The 4 chosen cards are all of the same value; (c) The four chosen cards are all pictures (i.e. J, Q, K or A); (d) It is not possible to say.
8. Two dice are thrown simultaneously. Which of the following is not true? (a) The probability of throwing two *sixes* is 1 in 36; (b) The probability of throwing *two even numbers* is 1 in 6; (c) The probability of throwing *one 5 and one 6* is 1 in 18; (d) The probability of getting a combined score of *3 or less* is 1 in 12.
9. Which of the following is least likely to be true? (a) An asteroid will hit the earth in 1000 years time; (b) The conservatives will be back in power in 1000 years time; (c) No-one will die from cancer in 1000 years time; (d) It is impossible to say which of these is least likely.
10. The probability of unwanted side-effects after taking a particular vaccine is 1 in 25. A group of 500 people are given the vaccine. Which one of the following statements do you think is true? (a) It is possible that no-one will experience side effects; (b) The expected number of people who will experience side effects is 40; (c) It is impossible that more than 50 people will experience side effects; (d) At least one person in the group will definitely experience side effects.

statistics to ensure that there were no errors or ambiguities. It was then pilot tested in a sample of 50 non-healthcare professionals (most of whom were IT workers) and the accuracy data were normally distributed around a mean of 50% correct. Chance performance (i.e. if a respondent guessed every answer) would be 25%. The rationale for each question, along with an explanation for each correct answer, is presented in the appendix.

Participants

One-hundred-and-fifty Mental Health Professionals (MHPs) were recruited on a voluntary basis from various NHS mental health units. Anyone who, as part of their job, undertook some form of risk assessment was included. The sample included: consultants (both psychiatrists and psychologists), junior doctors, CPNs, nurses, social workers, occupational therapists and managers (a breakdown of MHPs by general profession is displayed in Table 2). The mean age of the group was 41 ($\pm$ 10, range 21–67 years) and 58 (39%) were male. Forty-two (28%) were educated up to 18 years and 108 (72%) were educated to degree or postgraduate level.

One-hundred-and-fifty non-healthcare professionals were matched to the MHPs on mean age, education and gender-mix. This group included, amongst others, IT workers, office staff, civil servants, teachers, solicitors, employment consultants, skilled manual workers and retired people or housewives with professional backgrounds. Scientists and academics were excluded, as was anyone who worked in any kind of healthcare setting. These participants were recruited on a voluntary basis by friends and colleagues of the investigative team. The mean age of the control group was 41 ($\pm$ 13, range 19–75) years and 58 (39%) of the sample were male. Fifty-nine (39%) per cent had been educated up to age 18 and 91 (61%) had a higher qualification. The control group did not differ significantly from the MHP group on any of these demographic variables: age ($F < 1$, $p = 0.9$); gender mix ($X^2 = 0.01$, $p = 0.9$); educational level ($X^2 = 3.6$, $p > 0.05$). Although the difference in educational level did not reach significance between MHPs and controls, there was a trend towards a higher level of education in the MHP group. This was taken into account in the

Table 2: Breakdown of MHPs by profession, gender, age and educational level

| Profession	*n*	Male	Female	Age (mean)	Higher Education
Consultant	16	10	6	48	16 (100%)
Junior Doctor	32	15	17	34	32 (100%)
Social Worker	23	7	16	45	17 (74%)
Occupational Therapist	5	0	5	37	5 (100%)
CPN	19	7	12	42	9 (47%)
Nurse	37	14	23	42	21 (57%)
Manager & Other	18	5	13	44	6 (33%)

between-group comparisons (see results).

Procedure

Participants were asked to complete the questionnaire in their own time. Participation was voluntary and anonymity codes protected the identity of any respondents. All respondents were asked to indicate how long it had taken them to answer the questions. No answers were given out to any participants until all data had been collected. Ethical permission was given for the study by the East and North Herts Local Research Ethics Committee.

Results

Group performance

The MHP and control group did not differ in the number of questions answered correctly, scoring 4.6/10 ($\pm$ 1.9, range 1–10) and 4.6/10 ($\pm$ 2, range 0–10) respectively. Fig. 1 displays the accuracy levels for both groups and there was no significant between-groups difference on any single item (by X^2, all $p > 0.05$).
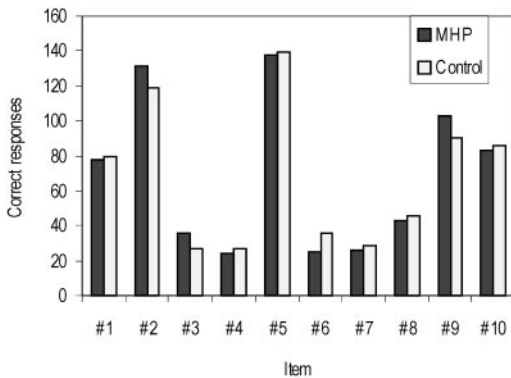
Figure 1:   Items analysis for MHPs and controls

Predictors of performance

Across all 300 participants, males did better than females (mean no. correct items = 5.2 vs. 4.2, $F = 16.8$, $p < 0.0001$) and, not surprisingly, those participants with some form of higher education were more accurate than those educated to 'A' level or below (means 5 vs. 3.7, $F = 33.8$, $p < 0.0001$). Age correlated negatively, though significantly, with total score ($r = -0.18$, $p = 0.002$). All three predictors were significant when entered into a multiple regression, together explaining 17% variance in total score (multiple $r = 0.41$, $p < 0.00001$).

As noted before, the MHPs and controls were matched across all of these predictor variables: nonetheless the MHP group did include a greater number of individuals with higher qualifications (108 vs. 91), though this difference did not reach significance. Given that education was such a powerful predictor of accuracy, the group comparisons, where re-run firstly for those participants educated to 'A' level or below, and secondly for those participants with a higher qualification. There were still no differences between MHPs and controls ($F < 1$, $p = 0.73$; $F < 1$, $p = 0.46$ respectively).

Within the MHP group *alone*, profession was also a strong predictor ($F = 3.86$, $p = 0.0007$, see Table 3) and this probably reflects the varied exposure to probability and statistics associated with different health professional training. Indeed, when entered into a multiple regression using *only* the MHP scores, age, educational level, gender and profession (as classified in Table 2) were all significant predictors, together explaining 21% variance in total score (multiple $r = 0.45$, $p < 0.00001$). The gender dif-

Table 3:   Inter-professional differences in response accuracy for MHPs

Profession	*n*	Mean no. items correct ($\pm$ SD)	Range (least-most accurate)
Consultant	16	6.25 ($\pm$ 2.5)	3 – 10
Junior Doctor	32	5.13 ($\pm$ 2.1)	1 – 9
Social Worker	23	4.60 ($\pm$ 1.6)	2 – 8
OT	5	4.60 ($\pm$ 0.9)	4 – 6
CPN	19	3.95 ($\pm$ 1.6)	2 – 8
Nurse	37	3.84 ($\pm$ 1.5)	1 – 8
Manager	11	4.64 ($\pm$ 1.9)	2 – 9
Other	7	3.86 ($\pm$ 1.3)	2 – 6

ference in the MHP group was almost identical to that found in the control group (M : F 5.2 vs. 4.2; M : F 5.2 vs. 4.3 respectively).

There was no correlation between the amount of time taken to complete the questionnaire and the level of response accuracy achieved by MHPs ($r = 0.05$, $F < 1$, $p = 0.6$). However, a small though significant relationship was observed for control participants ($r = 0.25$, $F = 7.3$, $p = 0.008$). This discrepancy is all the more interesting given that, on average, control participants took *less time* to complete the questionnaire (Mean 8.6 min $\pm$ 3.8, range 3 – 20) than MHPs (Mean 10.3 min $\pm$ 5.3, range 1 – 30) and this between-groups difference reached statistical significance ($F = 7.5$, $p = 0.007$). In short, MHPs did not seem to benefit from thinking longer about each question, suggesting that if they did not already know the answer they were unlikely to work it out.

Item analysis

The MHP responses for each item were investigated for evidence of systematic error. A systematic error would be indicated where the distribution of *erroneous responses* for an item was highly uneven (i.e. where a modal *incorrect* response could be clearly identified). So, for example, if 60 MHPs answered one question incorrectly, and 52 of these all opted for the same incorrect answer, the pattern of error would be systematic rather than random. All but one item elicited systematic errors (Table 4).

The individual items can be arranged into four groups based upon the performance accuracy of the MHP respondents: 1 Highly accurate ($> 75\%$ correct), 2 Moderately accurate (51 – 75% correct), 3 Inaccurate but still above chance (25 – 50% correct) and 4 Inaccurate below chance ($< 25\%$).

1. Highly accurate

Two items (#2 and #5, which both related to prediction of random distributions) were correctly answered by the majority of respondents (87% and 92% respectively). As would be expected, given the thematic link between these items, there was a high level of response consistency: 83% respondents got both correct, 4% got both wrong and 13% got only one right. The slightly better performance on question 5 may reflect greater familiarity of chance within the context of lottery outcomes. Those people who did make errors on these items

Table 4: The distribution of MHP responses for each question: correct answers are indicated by bold typeface. The distribution of erroneous responses for each question was compared to chance distribution using χ^2. Significant results (at $p < 0.05$) indicate evidence for systematic error. The column 'NR' indicates the number of non-responses (i.e. failures to indicate any answer at all).

Item	A	B	C	D	NR	Systematic Error
No. 1	6 (4.1%)	**78 (53.1%)**	50 (34%)	13 (8.8%)	3	Yes
No. 2	10 (6.7%)	0 (0%)	9 (6%)	**131 (87.3%)**	0	Yes
No. 3	**36 (24.5%)**	60 (40.8%)	22 (15.0%)	29 (19.7%)	3	Yes
No. 4	19 (12.8%)	6 (4.0%)	**24 (16.1%)**	100 (67.1%)	1	Yes
No. 5	0 (0%)	1 (0.7%)	11 (7.3%)	**138 (92%)**	0	Yes
No. 6	1 (0.7%)	3 (2%)	**25 (16.9%)**	119 (80.4%)	2	Yes
No. 7	17 (11.4%)	3 (2%)	**26 (17.5%)**	103 (69.1%)	1	Yes
No. 8	23 (16.8%)	**43 (31.4%)**	29 (21.2%)	42 (30.6%)	13	Yes
No. 9	12 (8.2%)	14 (9.5%)	18 (12.2%)	**103 (70.1%)**	3	No
No. 10	**83 (56.1%)**	10 (6.7%)	8 (5.4%)	47 (31.8%)	2	Yes

tended to choose the answers that 'looked' most random (i.e. A or C on item #2, and C on item #5).

2. Moderately accurate

Three items (#1, #9 and #10) generated correct responses in more than half but less than three-quarters of the respondents (53%, 70% and 56% respectively). There was no particular thematic link between these items and response consistency was weak (34% were correct on all 3, 3% were incorrect on all 3 and 63% were correct on some). The majority of people who were incorrect for item No. 1, chose answer C (0.02% instead of 0.2%) suggesting an arithmetical difficulty in dealing with small numbers. This finding is important because many medical outcomes (e.g. risks of some side effects), have low probability of occurring and the implication here is that some health professionals may have difficulty conceptualising or expressing these odds (it is of particular concern that 34% junior doctors and 19% consultants were incorrect on this item). For item No. 9, the distribution of erroneous responses was almost random suggesting that those MHPs who did not know the answer responded with a pure guess. For item #10, however, most incorrect responders chose answer D suggesting a systematic conceptual error with regard to verbal descriptors of probability.

3. Inaccurate but above chance

Item #8 was answered incorrectly by approximately 69% of responders. Although the distribution of errors was more even than for most other items, it did deviate from chance ($p = 0.02$). In one respect, this was arguably the most difficult question because it necessitated multiple probability calculations before an answer could be reached. It is surprising then that other, apparently less difficult, questions were answered with *sub-chance* levels of accuracy (see 4 below).

4. Inaccurate below chance

Items No. 3, No. 4, No. 6, No. 7 all elicited levels of response accuracy worse than that predicted by chance (i.e. pure

guessing by all respondents would, in theory, give an overall accuracy level higher than was observed here), again suggesting widespread systematic error. For example, on item No. 3, the most popular response by far (at 43%) was answer B (1 in 36), indicating that many people thought this particular outcome was 6 times less likely than was actually the case. Moreover, for items No. 4, No. 6, and No. 7, the most popular response was answer D – 'it is not possible to say' (63, 77, 66% respectively) suggesting that some respondents equated being unable to answer the question with the question being impossible to answer. Indeed, for items #6 and #7, more people chose the 'it is impossible to say' option than for item #9 (which was the only question where it really was not possible to give a precise answer!). There was no strong thematic link between the items that elicited worse than chance performance suggesting that MHPs have difficulties with a whole range of different issues surrounding probability theory. Consistency for the sub-chance items was moderately high, simply because so many people performed poorly (only 3% were correct on all 4, 58% were incorrect on all 4 and 39% were incorrect on some).

Discussion

This was a very simple study which tested MHPs' ability to deal with some basic concepts in probability and arithmetic. Nonetheless, the data address some important issues.

Overall accuracy across the questionnaire was 46% or, expressed another way, MHPs answered more questions incorrectly than correctly. Of course, percentage accuracy scores reflect an overall level of task difficulty, though

we did use questions that (i) were non-technical, (ii) did not probe advanced statistical concepts and (iii) were set at a level such that a person numerate to 'O' level or GCSE standard might reasonably expect to know the answers, and are not *absolute* measures of aptitude for solving these kinds of problems. Nonetheless, as *relative* measures, they do provide some useful indications. For example, MHPs performed identically to a matched control group, indicating that they do not have a marked deficiency in probability skills (although our initial prediction was that they should be more accurate). Similarly, percentage accuracy scores distinguished between (i) different mental health professions and (ii) genders. This may, of course, reflect prior training and educational experiences of these sub-groups: for example, that a robust gender effect was observed in both MHP and control groups suggests a common explanation and this is most likely that females are under-represented in groups of people with higher mathematical and scientific qualifications.

Analysis of the individual items in our questionnaire suggests that, aside from knowledge of random distributions which was generally quite good (e.g. No. 2, No. 5), MHPs had conceptual difficulties with all other aspects of probability tested here. These ranged from (i) simple manipulation of probabilistic information (e.g. No. 1) through (ii) appraisal of relative likelihood using information presented within the question (e.g. No. 4, No. 10) to (iii) appraisal of relative likelihood using information unavailable in the question (e.g. No. 6, No. 9). Few of the questions here actually required the respondent to compute actual probabilities for each possible

outcome: indeed, for several questions (e.g. No. 4, No. 7) it was possible to make a correct choice simply by counting set sizes (see appendix for explanations). However, it is noteworthy that such examples fell within the questions that were answered at a sub-chance level of accuracy: this suggests that respondents did not adopt or look for simple strategies but, rather, preferred the assumption that an answer could not be given. This was also true of the control group so, once again, the MHPs have not indicated abnormally poor conceptual skills in this respect.

However, one important difference between the control group and MHPs is the expectation that the latter should be skilled in handling concepts of risk. Indeed, the general premise of risk assessment in mental health is that the professional has access to data which (i) enables the chance of various outcomes to be considered and (ii) depending on the perceived likelihood and severity of the outcomes, can be used to support changes in care management. However, such a premise assumes the professional to be a rational actor, considering data in relation to a patient, and not making errors in the manipulation of that data. The current study demonstrates that MHPs are far from error-free in handling some fundamental concepts in probability. No matter how good the data input (e.g. patient characteristics, previous medication history, etc.) are, a clinical risk assessment process can have no meaningful output if the conceptual processes underlying decision making are flawed. On the one hand, it may be argued that simple probability dilemmas (e.g. coin-tossing experiments) do not characterize the type of scenarios that are considered by a MHP undertaking a risk assessment. On the other

hand, if an individual is unable to make relative (let alone absolute) predictions when provided with the relevant information for a *single-parameter* choice-dilemma, how can the same individual be expected to make accurate assessments in the real-world, where multiple parameters must be taken into account? A good example here is item No. 3 in our questionnaire: the number of MHPs who answered this question correctly was fewer than chance would predict, with the clear majority thinking that the outcome was six times less likely than it really was; here is a simple case of looking at two independent events (both of which have the same parameters) and calculating the odds of both events occurring. A clinical analogue of this is that a patient risk assessment may need to take into consideration many parameters, some of which are inter-dependent and some of which are independent: if a MHP does not know how to weight the data accordingly, this may lead to considerable over- or under-estimation of risk. This also leads on to a very important aspect of probability that we did not assess in our study: namely, *conditional probability*. Conditional probability is the likelihood of one event occurring *given* that another event has already occurred (i.e. where the two events are not independent) and this is perhaps more representative of the typical risk scenarios that a clinician may be faced with. However, the point here is that calculation of conditional probabilities (e.g. by using Bayes' Theorem) is considerably more complex than the simple probability calculations required in our questionnaire. Thus, we would predict that MHP understanding of conditional probability (and hence real-world risk assessment) will be even weaker than indicated by our questionnaire.

Some might argue that formal concepts of probability theory have little to do with real-world clinical risk assessment, and that the latter relies on intuitive and qualitative procedures. That intuition may serve some useful purpose in risk assessment is not something we would directly contest. However, risk is *by definition* about the statistical likelihood of potential outcomes, for example 1 in 10, 1 in 1000, and so on. We would argue that verbal descriptors which reflect an intuitive, qualitative assessment process (for example, 'high', 'moderate' or 'low-risk') simply blur the risk outcome and convey minimal information for directing future action. Moreover, evidence-based assessments of risk for given outcomes are often available from epidemiological data and so an individual assessment should increase, not decrease, the precision of any such prediction. For example, if we know the computed risk of suicide in a given population, this figure should arguably form the basis of any individual risk assessment within the same population. This is simply because a population mean will always be the best predictor under uncertainty. This is not to say that further individual risk assessment is unwarranted but, rather, that any additional information is considered in conjunction with pre-existing evidence. In short, to describe an individual as being at 'low' risk of suicide would be a meaningless statement.

Although MHPs may not be currently trained to think about risk in terms of probability, this is not to say that real-world clinical risk assessment has nothing to do with probability. The fact that most risk assessments in mental health do not currently use a formal probabilistic language is not an argument for the validity of risk assessment as intuitive and qualitative but, rather, an observation of the inadequate empirical basis for risk assessment. One concrete recommendation arising from this paper concerns training. This would provide an opportunity to educate staff in the basics of probability theory, at least periodically. Whether it will be at all acceptable for healthcare professionals to 'go back to school' to learn some maths is an interesting question.

Acknowledgements

References

Biehl, M. & Halpern-Felsher, B.L. (2001). Adolescents' and adults' understanding of probability expressions. *Journal of Adolescent Health, 28,* 30–35.

Bogardus, S.T., Holmboe, E. & Jekel, J.F. (1999). Perils, pitfalls and possibilities in talking about medical risk. *JAMA, 281,* 1037–1041.

Department of Health (1995). *Building Bridges: a Guide to Arrangements for Inter-Agency Working for the Care and Protection of Severely Mentally Ill People*. London: DoH White Paper.

Department of Health (1994). *Guidance on the Discharge of Mentally Disordered People and their Continuing Care in the Community*. London: DoH White Paper.

Department of Transport (2001). *National Travel Survey*, DoT publication.

Dillon, J. & Gill, P. (2001). Risk, environment and health: aspects of policy and practice. *School Science Review*, *83*, 65–73.

Edwards, A.G.K. & Elwyn, G.J. (2001). Risks – listen and don't mislead: editorial. *British Journal of General Practice*, April 2001.

Edwards, A.G.K., Elwyn, G.J., Covey, J., Matthews, E. & Pill, R. (2001). Presenting risk information – a review of the effects of 'framing' and other manipulations on patient outcomes. *Journal of Health Communication*, *6*, 61–82.

Fuller, R., Dudley, N. & Blacktop, J. (2001). Risk communication and older people – understanding of probability and risk information by medical inpatients aged 75 years and older. *Age and Ageing*, *30*, 473–476.

Gale, T.M., Woodward, A., Hawley, C.J., Hayes, J., Sivakumaran, T. & Hansen, G. (2002). Risk assessment for people with mental health problems: a pilot study of reliability in working practice. *International Journal of Psychiatry in Clinical practice*, *6*, 73–81.

HMSO (1995). *Mental Health (Patients in the community) Act*. London, HMSO.

Kahneman, D. & Tversy, A. (1972). Subjective probability: a judgement of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kirkwood, A.S. (1994). Why do we worry when scientists say there is no risk? *Disaster Prevention and Management*, *3*, 15–22.

Marteau, T.M., Senior, V. & Sasieni, P. (2001). Women's understanding of a 'normal smear test result': experimental questionnaire based study. *British Medical Journal*, *322*, 526–528.

Montandon, C. & Harding, T. (1984). The reliability of dangerousness assessments: a decision making exercise. *British Journal of Psychiatry*, *144*, 149–155.

O'Connor, A.M., Rostom, A., Fiset, V., Tetroe, J., Entwistle, V., Llewellyn-Thomas, H., Holmes-Rovner, M., Barry, M. & Jones, J. (1999). Decision aids for patients facing health treatment or screening decisions: systematic review. *British Medical Journal*, *319*, 731–734.

Ryan, T. (1998). Perceived risks associated with mental illness: beyond homicide and suicide. *Social Science and Medicine*, *46*, 287–297.

Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science*, *185*, 1124–1131.

Appendix

The rationale behind each question and an explanation of the correct answers follow:

Question 1 simply assesses the ability of respondents to convert probabilistic information from one medium to another. This would have some parallels in the interpretation of epidemiological data. The correct answer is B, since 2 divided by 1000, and then multiplied by 100, is 0.2.

Question 2 tests knowledge of random distributions. It is well known that individuals ascribe too much weight to small samples, expecting them to produce stable patterns that are representative of the population of possible outcomes (e.g. Kahneman & Tversky, 1972). In these kinds of scenarios this may result in a bias towards belief that random events should *not* occur in sequences that appear 'orderly' even though this is in fact untrue. The correct answer is D since each of these exact outcomes is equally likely.

Question 3 tests whether individuals are able to work out the exact probability of two thrown dice showing the same

number. To do this, it is necessary to calculate the number of outcomes where this is true ($n = 6$) as a proportion of the total number of possible outcomes ($n = 36$). Thus, the correct answer is A. Another way of thinking about this problem is that it does not really matter what number the first die reads, but there is a one in six chance that the second will read the same number. Essentially, this problem can be solved by counting.

Question 4 tests whether the respondent can use the information given to predict the most likely result. Given that there are 20 more red than blue sweets, it is always more likely that a red sweet will be drawn on each of the three occasions. The probability of this happening can be calculated as 0.212 which is higher than any of the other possibilities given. Thus the correct answer is C. However, although it is possible to calculate exact probability of each outcome, this is not actually necessary because one only needs to take account of the relative counts of each colour to make a judgement.

Question 5 is similar to question 2 except that the range of possible outcomes is much greater. Folk psychology may suggest that the same lottery numbers will not come up two weeks running but in fact, last week's sequence is just as likely to come up as any other sequence. The correct answer is D.

Question 6 does not provide and direct statistical information but it is still possible to rank the options in order of *least* to *most* likely by considering the relative frequencies of each type of person in the population of possible lottery participants. There are many more people who cannot drive than people with twin siblings (The 1999/2001 UK travel survey estimates that 29% of the population aged over 17 do not hold any form of driving license [DoT, 2001] and national birth statistics indicate that only $1-2\%$ of births are multiple). The number of people who have won the lottery jackpot before would be the smallest group, being less than 0.01% of the eligible population. Thus, the most likely of the given options is C. Even though it is very unlikely that people would know the specific probabilities associated with each option, they ought to be able to work out which one is most likely. This question therefore assesses whether individuals can recognise that an assessment of likelihood can sometimes be made using *relative* rather than *absolute* probabilistic information and has parallels in risk assessment (where absolute risks may be unknown).

Question 7 tests whether an individual can use frequency information to decide which outcome is most likely. In a pack of cards, the number belonging to any one suit is 13, the number which have the same value is four and the number of picture cards (J, Q, K and A) is 16. Thus, it is more likely that four cards chosen at random will be picture cards (C) because this is the largest set. As with question 4, it is possible to prove this is true using the multiplication rule, but this is not necessary to answer the question correctly.

To answer question 8 correctly, one can either use multiplication and addition rules, or one could calculate the proportion of outcomes that would fulfil each event as a proportion of the total number of possible outcomes. The probability of throwing two sixes is one sixth squared (1 in 36), the probability of throwing two even numbers is one-half squared (one in four), the probability of throwing one 5 and one 6 is one-sixth

squared multiplied by two (one in 18) and the probability of a combined score of three or less is one-sixth squared multiplied by three (one in 12). Thus, the correct answer (i.e. the statement that is false) is B.

Question 9 presents three options that may or may not be true in 1000 years time. Although it is sometimes possible to rank events from the least to most likely (i.e. more or less probable) even when the absolute odds are not available (for example, see Question 6), this does not hold for question 9 because it is very unlikely that any current information or intuition will hold constant over such a long time period (e.g. consider the differences in politics, medicine and astronomy between 1000AD and 2000AD: can we be sure that the conservative party will even exist in 1000 years time? Do we know whether a small, and as yet, undetected asteroid may collide with the earth at the third millenium? Can we be sure that absolutely no-one, anywhere on the planet, will die of cancer?). Thus the correct answer is D. This question tests whether individuals are able to recognise when it is not possible to make an assessment because the necessary information is lacking.

Question 10 tests whether individuals can evaluate each possible outcome using the given information. If the risk of side effects is 1 in 25, the expected number of people experiencing these effects will be 20 (i.e. 500 multiplied by 1/25). Thus option B is incorrect. Although it may be *unlikely* that more than 50 people will experience side effects, it is certainly possible. Thus C can also be discounted as false. Although it is likely that at least one person in the group will experience side effects, this is not a certainty and so D is false also. Thus the correct answer is A, even though it seems rather unlikely that no-one will experience side effects. This question also tests the respondents' understanding of verbal probability descriptors (see Biehl & Halpern-Felsher, 2001).

QUERIES: to be answered by AUTHOR

AUTHOR: The following queries have arisen during the editing of your manuscript. Please answer the queries by marking the requisite corrections at the appropriate positions in the text.

QUERY NO.	QUERY DETAILS	QUERY ANSWERED
	No queries	