2005

"What does this mean?" How Web-based consumer health information fails to support information seeking in the pursuit of informed consent for screening test decisions

Jacquelyn A. Burkell
Faculty of Information and Media Studies, Western University, jburkell@uwo.ca

D Grant Campbell
Western University

What are the chances? Evaluating risk and benefit information in consumer health materials

By Jacquelyn Burkell, PhD
jburkell@uwo.ca
Assistant Professor

Faculty of Information and Media Studies
Middlesex College
University of Western Ontario
London, Ontario
N6K 2C8 Canada

Much consumer health information addresses issues of disease risk or treatment risks and benefits, addressing questions such as ''How effective is this treatment?'' or ''What is the likelihood that this test will give a false positive result?'' Insofar as it addresses outcome likelihood, this information is essentially quantitative in nature, which is of critical importance, because quantitative information tends to be difficult to understand and therefore inaccessible to consumers. Information professionals typically examine reading level to determine the accessibility of consumer health information, but this measure does not adequately reflect the difficulty of quantitative information, including materials addressing issues of risk and benefit. As a result, different methods must be used to evaluate this type of consumer health material. There are no standard guidelines or assessment tools for this task, but research in cognitive psychology provides insight into the best ways to present risk and benefit information to promote understanding and minimize interpretation bias. This paper offers an interdisciplinary bridge that brings these results to the attention of information professionals, who can then use them to evaluate consumer health materials addressing risks and benefits.

INTRODUCTION

Information professionals working in the area of consumer health information are careful to select the best possible resources: those that are accurate, unbiased, and appropriate for the intended audience [1]. One important concern is accessibility: can readers understand and use the information? The most common measure of accessibility is reading level, and material written at a reading level of grade eight or lower is held to be appropriate for the general public [2]. This evaluation criterion is useful for consumer health material that is primarily textual in nature, but it is less appropriate for other types of information, including quantitative information and information organized in tables or figures rather than text [3].

Much consumer health information addresses questions of disease risk or treatment risks and benefits, providing information regarding questions such as ''What is the chance that I have West Nile virus?'' or ''Should I opt for surgery alone or surgery and radiation in the treatment of my cancer?'' Relevant information includes, for example, the proportion of people who have contracted West Nile virus or the survival rates for cancer patients treated with surgery com-

pared to surgery plus radiation. This information is essentially quantitative in nature in that it involves the concept of outcome likelihood (e.g., 1 in 500 people have a given infection, or the survival rate is 95%). As a result, reading level does not adequately reflect accessibility, and, to evaluate resources that communicate benefit and risk, other assessment criteria are required.

Research in cognitive psychology provides an excellent, if somewhat unexpected, source for these criteria. Cognitive psychology is the study of human information processing, and empirical research in the discipline examines the interaction of people with information. This body of research has been effectively mined to identify general principles for information presentation [4] and principles for the design of information graphics [5]. This paper extends this approach to the development of principles for the presentation of information regarding risks and benefits. Armed with these principles, information professionals can identify (and possibly design [6]) optimal presentations of risk and benefit information. Ultimately, the goal is to identify communications that, to borrow a phrase from Norman [7], ''make us smart,'' those that present risk and benefit information in a format that is *designed* to promote accurate and unbiased interpretation.

One does not have to look very far in consumer health information to find examples of risk and benefit communication, and the challenges for consumers in understanding and interpreting this information are immediately evident. Consider, for example, the following passage describing breast cancer risk factors:

Your chances of developing breast cancer increase as you get older. The disease rarely affects women under 30 years of age, while close to 80 percent of breast cancers occur in women over age 50. At age 40, you have a 1 in 217 chance of developing breast cancer. By age 85, your chance is 1 in 8. [8]

This information raises a broad range of questions, some of which are not answered by the information provided, and others of which involve complex calculations and reformulation of the data. Is a risk of 1 in 8 higher than a risk of 1 in 217? If so, how much higher? What does it mean to say the disease *rarely* affects women under 30? What is the risk of breast cancer in a woman 50 years old? Another example is the following quote from a Website providing information for teens about West Nile virus:

The good news is that, even in areas where mosquitoes are more likely to be carrying the virus, it's very unlikely that a person will become sick from a mosquito bite. Only 1% of the mosquitoes in a region affected by West Nile virus are actually infected with the virus. And less than 1% of the people who do become infected with West Nile virus become severely ill. [9]

Faced with this information, typical readers would have some degree of difficulty determining their risk of becoming severely ill with West Nile virus, which according to these statistics is 1 in 10,000 (or 0.0001 or 0.01%) if they are bitten by a mosquito in an affected region.

Information about outcome likelihood is particularly relevant to health care decisions. Those choosing between health care alternatives need to understand the likelihood of both the negative outcomes (risks) and the positive outcomes (benefits) associated with the available options to make informed choices between them. Thus, for example:

■ Informed decisions about screening tests (e.g., decisions about maternal serum screening) require at minimum an understanding of the baseline risk of having the condition, the probability of a false negative test result, and the probability of a false positive test result [10].

■ Women making decisions about hormone replacement therapy to treat menopausal symptoms must understand and weigh the reduced risk of osteoporosis, cardiovascular disease, colorectal cancer, and Alzheimer's disease against the increased risk of breast cancer, myocardial infarction, cerebrovascular disease, and thromboembolic disease [11].

■ Men choosing among options for the treatment of localized prostate cancer want to know the likelihood of side effects associated with the treatment options before making their decision [12].

■ Participants in genetic counseling programs must understand the risks associated with treatment and the meaning of a positive test result to make informed decisions about genetic testing [13].

Not surprisingly, empirical research indicates that information about risks and benefits tends to be difficult to understand [14], at least in part because the interpretation of this type of information requires significant quantitative skill [15–17]. Quantitative literacy is quite limited in the general public: the International Literacy Survey [18] indicates that almost half of North Americans lack what are considered the minimum skills required to apply arithmetic operations to numbers embedded in printed materials. Fractions and proportions (exactly the type of quantitative information typically used to present risks and benefits) are the types of numerical information that prove most challenging for the average person [19]. Furthermore, even highly educated people have difficulty performing the quantitative operations that are commonly required in the interpretation of likelihood (e.g., converting from percentages to proportions and vice versa [20]), and experts fall prey to the same biases in interpretation that affect lay people [21]. Thus, the understanding of information regarding risks and benefits proves challenging for many, if not all, people.

Thus far, the news seems bad: communications about risks and benefits are ubiquitous in consumer health information, and people have trouble understanding and using these communications. So what can an information professional do? Training consumers of health information to make sense of medical data is one approach [22–24], consistent with the general principle of empowering consumers by supporting literacy initiatives [25]. Careful examination of the relevant research in cognitive psychology offers another, perhaps adjunct, method of addressing the issue. This research indicates that the format in which likelihood is presented—verbal, numeric, or visual—influences understanding. The research also identifies those other aspects of presentation that tend to produce biased interpretation of risk and benefit information. Based on these results, it is possible to identify the characteristics of ''good'' presentations of risks and benefits that maximize understanding and minimize bias.

Throughout this paper, one example will be used to illustrate the concepts being discussed. Imagine a forty-year-old woman, pregnant for the first time, coming to you for information about maternal serum screening. Her primary focus is screening for Down syndrome, and she wants to be sure to make an informed decision regarding whether to take the test. She is particularly concerned about the meaning of a positive test result, because she knows that a positive result (even a false positive) would cause her significant psychological distress, and because she understands that the tests commonly recommended to distinguish true positive from false positive results (amniocentesis and chorionic villus sampling) themselves carry a risk to the child. Much of her required information regards

outcome likelihood: her overall risk of having a child with Down syndrome (approximately 1%), the likelihood that a case of Down syndrome will be correctly identified by the test (termed *sensitivity,* maternal serum screening has a 90% sensitivity for Down syndrome, indicating that about 90% of cases will be correctly identified, while 10% while receive an incorrect negative test result), the likelihood that correct negative test result will be returned when the fetus does *not* have Down syndrome (termed *specificity,* maternal serum screening for Down syndrome has a specificity about 60%, indicating that about 60% of negative cases are correctly identified, while 40% of negative cases receive a false positive result), and the iatrogenic risk of amniocentesis (about 1%) and chorionic villus sampling (about 1%).

Verbal labels for likelihood

Likelihood is essentially a numerical concept; nonetheless, a wide variety of verbal terms are used to communicate the chance that an outcome will occur. One obvious advantage to the use of verbal labels for likelihood is that compared to numerical representations, verbal labels are generally viewed as easier to use and more natural, perhaps because they consist of common words that seem to be easily understood [26–28]. This apparent advantage, however, hides a serious drawback: inconsistent interpretation. On a positive note, verbal probability labels tend to be *ordered* consistently [29], so that people generally agree that some verbal labels imply lower likelihood (e.g., probabilities labelled as ''extremely low'' or ''low''), while others imply higher likelihood (e.g., probabilities labelled as ''high'' or ''very high''). There is, however, no consensus about the particular numerical figure that best represents a given verbal probability label [30, 31], and each verbal label tends to correspond to a wide range of numerical probabilities [32]. The numerical probabilities assigned to verbal probability labels differ across individuals (e.g., physicians and patients assign different numerical probabilities to the same verbal probability label [33]) and across context (e.g., with the outcome that is being considered [34, 35] or with the context in which the outcome occurs [36]). Thus, a ''low'' risk of complications may mean 10% to one person and 2% to another, or a ''high'' risk of death may be 1%, while a ''high'' risk of minor injury could imply 20%. Overall, the evidence suggests that while verbal labels for likelihood are viewed as easy to use, their interpretation is highly variable and dependent on the specific context.

When communicating likelihood, information providers tend to prefer to use verbal labels, especially when the exact probability of the outcome is unknown; information users, by contrast, usually prefer that likelihood be presented in numerical terms [37, 38]. Verbal labels are viewed as less precise than their numerical counterparts [39], which no doubt explains the different preferences for verbal versus numerical representations. Information providers choose verbal labels, because they are careful not to express more than they know, while users prefer numerical representations, because they want the most precise information they can possibly get. Both communicators and those receiving the communication agree that verbal labels are used to describe uncertain or vague probability estimates. Verbal labels, therefore, serve a dual purpose in communication: they indicate the general likelihood that an outcome will occur (e.g., low, medium, high), and they signal that there is some uncertainty about the exact level of probability.

If you found a resource for your client that described the likelihoods in verbal terms, it might read as follows:

Your overall likelihood of having a child with Down syndrome is *high.* If your baby actually has Down syndrome, it is *quite certain* that the test results will detect the problem; nonetheless, there is a *small possibility* that the problem will not be detected by the test. If your baby does not have Down syndrome, it is *somewhat likely* that the test result will be negative; it is, however, *possible* that the test will be positive even if the baby does not have Down syndrome. Amniocentesis or chorionic villus sampling may be recommended as further tests in the event of a positive test result; each of these procedures carries a *high* risk of spontaneous abortion.

It is important to note that, in this passage, the ''high'' risks of Down syndrome, amniocentesis, and chorionic villus sampling correspond to approximately 1%. The verbal label of ''high'' risk is chosen based on Calman's standardized verbal scale for risk [40]. Calman developed his scale to communicate low-probability risks associated with unlikely events such as being struck by lightning or contracting a rare disease. This scale, therefore, is not appropriate to communicate the sensitivity (90%) and specificity (60%) of the test. In fact, Calman's scale does not even have labels for probabilities in the range required. The verbal labels used to describe sensitivity and specificity were chosen on the basis of a study of the interpretation of standard verbal risk terms [41], which suggests, for example, that in *general* use the term ''somewhat likely'' corresponds to a chance of approximately 60%. Thus, the passage indicates that it is *somewhat likely* that the test result will be negative if your baby does *not* have Down syndrome.

This highlights one of the difficulties with verbal labels: the fact that interpretation changes with context. In general use, a 10% chance that an outcome would occur would be termed a ''small possibility'' [42] or a ''very low chance'' [43], but, when verbal labels are used to describe the likelihood of an uncommon adverse (usually medical) event, it has been suggested that risks of 1 in 100 (much lower than a 10% chance) should be termed ''high'' [44]. This leads to the counterintuitive situation where, in this passage, the ''high'' risk of Down syndrome is actually ten times *less* than the ''small possibility'' of a false negative result. There is no empirical evidence on whether people are able to accurately interpret multiple verbal labels for likelihood in a context where the outcome is changing, but simple perusal of the passage above

suggests that interpretation might pose a significant problem.

Verbal labels for likelihood are entirely appropriate for the communication of single probabilities that are vague or uncertain, that is, when the likelihood of an outcome is not precisely known [45, 46]. Thus, during the 2003 Severe Acute Respiratory Syndrome (SARS) crisis, it was appropriate to describe the risk of contracting this hitherto unknown disease on an airplane as ''low'' [47]. This type of use takes advantage of the positive qualities of verbal labels (ease of use and implicit communication of uncertainty) without incurring any of the costs of these labels incurred by their ''vague'' or indeterminate quality. When more than one likelihood is communicated for the purposes of combination or comparison (as in the Down syndrome example above), verbal labels are inappropriate because of the variability in interpretation. This is particularly true when the outcomes described range from very low-probability events (e.g., the possibility of a birth defect) to relatively high-probability events (e.g., the possibility of a positive test result).

Numerical representation of likelihood

One general conclusion arises from the research on verbal probability labels: if precise information about likelihood is available, the precision of a numerical representation is appropriate. Of course, the alternative also holds true: numerical representations of probability should not be used if probability is vague or uncertain. As Wallsten [48] argues, the use of numerical probability to represent vague or unknown likelihood results in an unwarranted assumption (on the part of the decision maker) about the precision of the probability estimate. This is particularly important because decision makers prefer options with precise probabilities over those where likelihood is vaguely specified [49] and, thus, tend to prefer options with likelihood described numerically over those where less precise verbal labels are used. Using a numerical representation, therefore, can bias the evaluation of an alternative based on the (possibly incorrect) assumption that the probability is precisely known. It is, of course, possible to indicate uncertainty in a numerical probability by specifying a range instead of a single value (e.g., between 10% and 40%) or by applying an adjective such as ''approximately'' to a numerical probability estimate (e.g., approximately 20%). However, little research has been done on the implications of these strategies for the interpretation of risk communications, and it remains a question whether either or both of these methods appropriately counteract the implied precision of the numerical representation. For vague or uncertain probabilities, therefore, numerical representations should be avoided, and, for probabilities that can be precisely specified, numerical representations are preferred.

Numerical representations of likelihood come in a wide variety of forms. The most common of these are single-event probability (e.g., 0.05), percent (e.g., 5%), frequency (e.g., 5 in 100), and absolute frequency (e.g.,

600). The first three of these representations incorporate information about the likelihood of both occurrence and nonoccurrence (because the likelihood of nonoccurrence is the inverse of each, 0.95, 95%, or 95 in 100 respectively). The last representation indicates only the number of times the outcome occurs (or is expected to occur) and does not offer any information about nonoccurrences. In a direct comparison of these formats, Brase [50] found that frequencies (e.g., 5 in 100, which he terms ''simple frequencies'') are perceived as clearest and easiest to understand, followed by percent format (e.g., 5%). Single-event probabilities (e.g., 0.05) are perceived as the most difficult to understand. These data are consistent with studies of statistical reasoning, which indicate that frequency presentations facilitate understanding of data [51–53]. Thus, based both on perception and on actual performance, frequency presentations of likelihood information are better than other formats.

When the frequency format is used to present information about likelihood, there is evidence that the interpretation is unduly influenced by the absolute number of occurrences reported. Overall, when larger numbers (higher frequency, larger reference group) are used in frequency presentations, events are seen as more likely [54]. Thus, death rates of 1,286 in 10,000 (probability of 0.1286) are incorrectly rated as more risky than rates of 24.14 in 100 (probability of 0.2414) [55], and subjects demonstrate an objectively irrational preference for a 9 out of 100 (probability of 0.09) chance of winning a small lottery over a 1 out of 10 (probability of 0.1) chance [56]. In the interpretation of these expressions of likelihood, it seems that the focus is first on the absolute number of occurrences, followed by an insufficient correction for the size of the reference or comparison group, consistent with the ''anchoring an adjustment'' cognitive bias identified by Kahneman and Tversky [57]. A general principle that arises in other contexts plays a role here: intuition tells us that larger numbers represent larger probabilities. This rule is entirely applicable for probabilities expressed as decimals or percent and holds for frequencies when they are expressed as counts over a group of standard size. It is, however, invalid when comparing frequencies occurring within references groups of different sizes. The rule seems to be applied by default or, at least, appears to have by default some influence on the subjective likelihood associated with a given explicit probability. Therefore, when likelihoods to be compared are expressed as frequency counts, they should be presented as occurrences counted over groups of a standard size, as opposed to a standard number of occurrences over groups of shifting size [58, 59]. Thus, comparisons between two likelihoods will be more accurate if they are presented as 5 out of 100 versus 25 out of 100, rather than the formally equivalent representation of 1 out of 20 versus 1 out of 4.

The advantage of frequency over probability representations is most pronounced in probabilistic reasoning tasks that prove difficult for lay people [60, 61] and experts [62, 63] alike. In the context of consumer health

information, the most common of these reasoning tasks is determining the predictive value of a symptom or screening test result. The positive predictive value (PPV) is the likelihood that a person actually has the condition given the presence of a symptom or a positive screening test result; the negative predictive value (NPV) is the likelihood that the person does not have the condition given the absence of the symptom or a negative test result. It is important that health care consumers understand the predictive value of symptoms and tests both for decision support and to help manage anxiety related to health and health care.

The predictive value of a symptom or test result is determined jointly by three factors: sensitivity (the probability that the test is positive or the symptom is present given that the person has the condition), specificity (the probability that the symptom is absent or the test result is negative given that the person does not have the condition), and the base rate of the condition (the proportion of people in the population who have the condition). When the relevant information is presented as either single-event probabilities (e.g., 0.05) or percents (e.g., 5%), the vast majority of experts and lay people strongly overestimate predictive value; when the same information is presented as frequencies, correct responding is much higher [64–66]. Using our example, the effect of format is immediately obvious. Here is the presentation of the relevant information in probability format:

The likelihood that a 40-year-old woman will have a child with Down syndrome is approximately 0.01. If your baby has Down syndrome, the likelihood that the test will detect the condition is 0.9, and the likelihood that the condition will not be detected by the test is 0.1. If your baby does not have Down syndrome, the likelihood that the test will be negative is 0.6, but there is a 0.4 likelihood that the test will be positive even if your baby does not have Down syndrome.

Compare this to the information presented in frequency format:

Of 1,000 pregnant women who are 40 years of age, 10 will have children with Down syndrome. If all 1,000 women were tested, 9 of the women with Down syndrome babies would test positive for the condition, and 1 would test negative. Of the 990 women whose babies do not have Down syndrome, 394 would test positive, and 596 would have test negative.

Given the first presentation, most people would guess that a positive test result would indicate a relatively high probability that the fetus has Down syndrome, on the order of 75%. However, when the information is presented as frequencies (as in the second example), it is immediately obvious that a positive test result carries *much* less diagnostic certainty: it is easy to see that a total of 403 positive test results are expected (9 true positive plus 394 false positive), and, of these, only 9 (slightly over 2%) are true positive results. Frequency format assists decision makers in making the correct interpretation; in contrast, presentation as probabilities or percents makes it difficult to determine predictive value.

It is important to note that, for the purposes of calculating predictive value, not all frequency representations are equal. To support this type of reasoning, the data must be presented in ''natural frequency'' format [67]. Natural frequencies are simply counts over a group of standard size; in the example above, all frequencies are expressed as incidents in the group of 1,000. A mathematically equivalent presentation of the same information could use different group sizes (e.g., 0.9 out of 100 test positive and have the condition; 1 of 1,000 tests negative and has the condition; 197 of 500 test positive and do not have the condition; 149 of 250 test negative and do not have the condition), but the representation no longer facilitates the correct interpretation. It becomes difficult to determine the predictive value with this presentation. Therefore, the two reasons to hold the size of the group standard when presenting frequencies are to facilitate comparisons (as discussed earlier) and to facilitate statistical reasoning.

These data suggest that numerical representations of likelihood signal certainty about the chances that an outcome will occur and are appropriately used when likelihood is known. Frequency representations (e.g., 1 out of 50) are preferred over other formats, because they are easier to understand and they promote accurate statistical reasoning. When the goal is only to present likelihood, and no statistical reasoning is required, percent format (e.g., 2%) is also appropriate, because it is perceived as easy to understand. Single-event probabilities (expressed as a value between 0 and 1) present the greatest challenge to understanding and, thus, should be avoided. Interpretation of likelihood represented as frequency is subject to the bias that higher numbers (e.g., higher incident counts) are interpreted as representing greater probability, without appropriate correction for the size of the group in which the incidents are noted. Therefore, when multiple likelihoods are to be compared or combined, each should be expressed as the number of occurrences in a group of a standard size (e.g., 1 out of 100, 5 out of 100). This form of presentation (counts over a group of standard size) is optimal for supporting the most difficult probabilistic reasoning that consumers of health information are likely to encounter: determining the predictive value of tests or symptoms.

Visual representation of likelihood

Visual representation of likelihood has the obvious advantage that visual information is salient [68] and relatively easy to understand [69], suggesting that both comprehension and recall of information about likelihood could be improved with visual communication. The discussion of numerical representations indicated that frequency formats are preferred over probability formats for numerical representations of probability. Given this obvious advantage for frequency representations, this section will be limited to one type of visual representation: the representation of frequency in the form of pictographs.

Frequency representations of likelihood include (as discussed above for numerical formats) the number of

Figure 1
Pictograph representing a frequency of 3 out of 100

Figure 2
Pictograph showing partial figure. This pictograph represents a frequency of 9 in 100
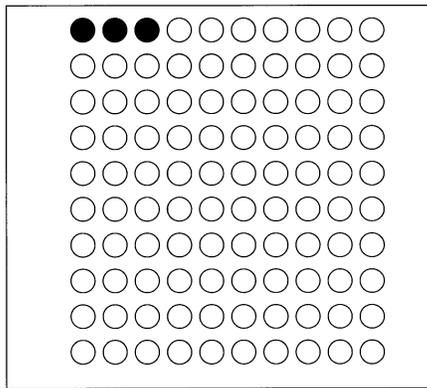

occurrences and the size of the group over which those occurrences are counted. In a pictograph, each member of the larger group is represented by a unique figure (e.g., a circle or an outline person), and the occurrences are shown by making a subset of the figures different in some obvious way. Thus, a frequency of 3 in 100 can be visually represented by 100 figures, three of which are visually distinct (Figure 1). This form of representation results in better understanding among both older and younger patients compared to verbal presentations as either frequency (e.g., 1 in 5) or fractional (e.g., 0.2 or 20%) probability [70, 71]. The only drawback to pictographs is that they require more space than equivalent numerical representations, particularly for very low likelihood events (e.g., 1 in 5,000), which require a large number of individual figures to represent likelihood.

There is some evidence that ''partial'' figures should be avoided in frequency pictographs. Thus, for example, to represent a frequency of 9 in 100, 10 figures could be presented, with 1 figure nine-tenths shaded (Figure 2). The evidence suggests that these partial figures are ''rounded up,'' so that the graphic would be interpreted as representing 1 in 10, not 9 in 100 [72]. The resulting interpretation would be an inflation of the actual likelihood.

As with numerical representations of frequency, the absolute number of distinct figures influences the perceived likelihood. Thus, for example, a frequency of 1 in 5 represented as one distinct figure among 5 will be seen as less likely than the same frequency represented as 20 distinct figures in 100 [73]. For frequencies that are to be compared, the lesson is both clear and familiar (from the discussion above regarding numerical frequencies): hold the size of the group constant (e.g., 11 out of 100 compared to 5 out of 100, not 11 out of 100 compared to 1 out of 20).

These results indicate that pictographs are a good way to present frequency information. See Figure 3 for a pictographic representation showing the hypothetical client the likelihood that she is carrying a child with Down syndrome. Information professionals
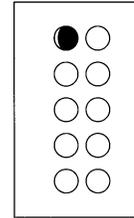
should, however, be aware, that in comparison to numerical representations, pictographic representations make risks more salient to decision makers. Frequency pictographs should follow the principle articulated for frequency representations in general: when multiple frequencies are presented, each should be shown as a number of incidents over a group of standard size. The size of the large group should be chosen so that frequencies can be represented without requiring partial figures, because these tend to be rounded up to whole numbers (e.g., 1.9 colored figures will be interpreted as 2).

CONCLUSION

Research in cognitive psychology, reviewed in this paper, leaves little doubt that the format of information about risk and benefit influences understanding and interpretation. Furthermore, based on this research, it is possible to identify optimal representations for this type of information. This paper offers an interdisciplinary bridge, so to speak, that brings these results to the attention of information professionals, who can then use them in the evaluation of consumer health

Figure 3
Likelihood of having a Down syndrome baby for a forty-year-old woman
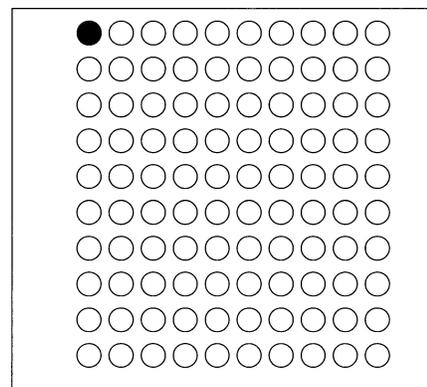
Each circle represents one forty-year-old woman carrying a baby. An empty circle indicates that the baby does not have Down syndrome. A filled circle indicates that the baby does have Down syndrome. This pictograph represents a 1 in 100 chance of having a Down syndrome baby.

resources that address risk or benefit. The general principles that emerge from this review of the literature are summarized below, providing succinct pointers for information professionals who want to identify those consumer health resources addressing risk and benefit that will most assist patrons in understanding this complex information.

Recommendations for evaluating risk communications

1. Verbal labels signal a vague or uncertain probability. They should, therefore, be used only to describe probabilities that are unknown or vague. One example would be communication of the risk of a new virus in the blood supply, where is it known that there is some degree of risk, but the risk cannot be precisely specified.

2. The meaning of a verbal label changes with the outcome being described, particularly if the outcomes range from very low-probability events (e.g., being struck by lightning) to higher-probability events (e.g., the chances of a thunderstorm occurring). As a result, verbal labels should not be used to describe multiple likelihoods in a single communication. For example, for low-probability risks (e.g., miscarriage as a result of amniocentesis), a 1% chance is labeled high, but when verbal labels are used in a more general context (e.g., to describe the likelihood of a false positive test result), a 10% chance is considered a small possibility. The interpretation of these two labels in a single communication presents difficulty for the information consumer and, thus, increases the likelihood of miscommunication and misunderstanding.

3. Numerical representations of probability are preferable to verbal labels when likelihood can be precisely specified. Thus, for example, discussions of the likelihood of medication side effects should use numerical representations of likelihood, because the chance of these side effects occurring can be precisely specified on the basis of clinical trial results.

4. When numerical representations are used for precisely known likelihoods, frequency format (e.g., 5 times out of 100) is most preferred, followed by percent (e.g., 5%). Probability format (e.g., 0.05) should be avoided, as this representation proves most difficult for consumers to understand. Thus, when presenting a known risk such as the chance of a 40-year-old woman having a child with Down syndrome, it is best to present the information as ''the chances are 1 in 100 that you will have a baby with Down syndrome'' (frequency format), or ''there is a 1% chance that you will have a baby with Down syndrome'' (percent format), but not ''there is a 0.01 likelihood that you will have a baby with Down syndrome'' (probability format).

5. When multiple risks are presented using frequency format, the size of the comparison group should be held constant (e.g., 1 in 100, 5 in 100, and 20 in 100 rather than 1 in 100, 1 in 20, and 2 in 10). When the size of the comparison group is held constant, it is easier to compare and combine different likelihoods.

6. Pictographs showing frequency representations of likelihood tend to be the easiest format to understand. The only drawback is that they take up a large amount of space, particularly in comparison to numerical representations and, thus, may be inappropriate when there are many likelihoods to be communicated or when presenting very low probability events (because a frequency of, for example, 1 in 500 requires a pictograph of 500 figures, 1 of which is visually distinct).

7. In a pictograph, the overall number of figures presented should be chosen, so that occurrences in this group can be shown as a number of *whole* figures, because partial figures tend to be rounded up in interpretation of pictographs. Thus, it is better to show a frequency of 9 in 100 as 9 distinct figures in a field of 100, rather than 1 figure that is 9/10 colored in a field of 10 figures, because the second presentation will be viewed as representing a frequency of 1 in 10, rather than 9 in 100.

8. When multiple pictographs are used to present risks or benefits in a single communication, each pictograph should depict occurrences in a group of a standard size (see point 5, above, for numerical frequency presentations). Thus, to present the risks of miscarriage separately for amniocentesis and chorionic villus sampling, each risk should be depicted as 1 distinct figure in a field of 100.

When consumer health information regarding risk and benefit is evaluated according to these guidelines, information professionals can rest assured that they have identified the best resources for their clients: resources that capitalize on cognitive capabilities while compensating for limitations.

REFERENCES

1. BAKER LM, MANBECK V. Consumer health information for public librarians. Lanham, MD: Scarecrow Press, 2002.
2. IBID.
3. RUDD RE, MOEYKENS BA, COLTON TC. Health and literacy: a review of medical and public health literature. In: Comings J, Garner B, Smith C, eds. The annual review of adult learning and literacy. v.1. San Francisco, CA: Jossey-Bass, 1999:158–99.
4. VAIANA ME, MCGLYNN EA. What cognitive science tells us about the design of reports for consumers. Med Care Res Rev 2002 Mar;59(1):3–35.
5. LIPKUS IM, HOLLANDS JG. The visual communication of risk. J Natl Cancer Inst Monogr 1999;(25):149–63.
6. WILLIAMS MD, GISH KW, GIUSE NB, SATHE NA, CARRELL DL. The Patient Informatics Consult Service (PICS): an approach for a patient-centered service. Bull Med Libr Assoc 2001 Apr;89(2):185–93.
7. NORMAN D. Things that make us smart: defending human attributes in the age of the machine. Reading, MA: Addison Wesley, 1993.
8. MAYO CLINIC. Breast cancer risk factors. [Web document]. Phoenix, AZ: Mayo Clinic, 2003. [rev. 4 Jun 2003; cited 27 Jun 2003]. <http://www.mayoclinic.com/invoke.cfm?objectid=B105D3F6-BFE9-4D89-A5E53914589F7E86&section=4>.
9. NEMOURS FOUNDATION. Should I worry about West Nile virus? [Web document]. Jacksonville, FL: Nemours Foundation, 2002. [rev. Aug 2002; cited 27 Jun 2003]. <http://kidshealth.org/teen/infections/bacterial_viral/west_nile.html>.
10. GOYDER E, BARRATT A, IRWIG LM. Telling people about

screening programmes and screening test results: how can we do it better? J Med Screen 2000;7(3):123–6.

11. RYMER J, WILSON R, BALLARD K. Making decisions about hormone replacement therapy. BMJ 2003 Feb 8;326(7384): 322–6.

12. STEGINGA SK, OCCHIPINTI S, GARDINER RA, YAXLEY J, HEATHCOTE P. Making decisions about treatment for localized prostate cancer. BJU International 2002 Feb;89(3):255–60.

13. CROYLE RT, LERMAN C. Risk communication in genetic testing for cancer susceptibility. J Natl Cancer Inst Monogr 1999;(25):59–66.

14. BOGARDUS ST, HOLMBOE E, JEKEL JF. Perils, pitfalls, and possibilities in talking about medical risk. JAMA 1999 Mar 17;281(11):1037–41.

15. BLACK WC, NEASE RG, TOSTESON AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995 May 17;87(10):720–31.

16. SCHWARTZ LM, WOLOSHIN S, BLACK WC, WELCH HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med 1997 Dec;127(11):966–72.

17. WOLOSHIN S, SCHWARTZ LM, MONCUR M, GABRIEL S, TOSTESON AN. Assessing values for health: numeracy matters. Med Decis Making 2001 Sep–Oct;21(5):382–90.

18. INTERNATIONAL ADULT LITERACY SURVEY. Literacy in the information age: final report of the International Adult Literacy Survey. Paris, France: Organisation for Economic Co-operation and Development, 2000.

19. IBID.

20. LIPKUS IM, SAMSA G, RIMER BK. General performance on a numeracy scale among highly educated samples. Med Decis Making 2001 Jan–Feb;21(1):37–44.

21. KAHNEMAN D, TVERSKY A. Judgment under uncertainty: heuristics and biases. Science 1974 Sep;185(4157):1124–31.

22. SEDLMEIER P. BasicBayes: a tutor system for simple Bayesian inference. Behav Res Methods, Instrum Comput 1997 Aug;29(3):328–36.

23. SEDLMEIER P, GIGERENZER G. Teaching Bayesian reasoning in less than two hours. J Exp Psychol Gen 2001 Sep; 130(3):380–400.

24. WOLOSHIN S, SCHWARTZ LM. How can we help people make sense of medical data? Eff Clin Pract 1999 Jul–Aug; 2(4):176–83.

25. SULLIVAN E. Consumer health: an online manual: health literacy. [Web document]. Houston TX: National Network of Libraries in Medicine: South Central Region, 2000. [rev. 26 Feb 2003; cited 27 Jun 2003]. <http://nnlm.gov/scr/conhlth/hlthlit.htm>.

26. BRUN W, TEIGEN KH. Verbal probabilities: ambiguous, context dependent or both? Organ Behav Hum Decis Process 1988 Jun;41(3):390–404.

27. KONG A, BARNETT GO, MOSTELLER F, YOUTZ C. How medical professionals evaluate expressions of probability. New Engl J Med 1986 Sep 18;315(12):740–4.

28. WALLSTEN TS, BUDESCU DV, ZWICK R, KEMP SM. Preferences and reasons for communicating probabilistic information in verbal or numerical terms. Bull Psychon Soc 1993 Mar;31(2):135–8.

29. CLARKE VA, RUFFIN CL, HILL DJ, BEAMEN AL. Ratings of orally presented verbal expressions of probability by a heterogeneous sample. J Appl Soc Psych 1992 Apr;22(8):638–56.

30. WALLSTEN T, BUDESCU DV, RAPPAPORT A, ZWICK R, FORSYTH B. Measuring the vague meanings of probability terms. J Exp Psychol Gen 1986 Dec;115(4):348–65.

31. THEIL M. The role of translations of verbal into numerical probability expressions in risk management: a meta-analysis. J Risk Research 2002 Apr;5(2):177–86.

32. MAZUR DJ, MERZ JF. How age, outcome severity, and scale influence general medicine clinic patients' interpretations of verbal probability terms. J Gen Intern Med 1994 May;9(5):268–71.

33. OHNISHI M, FUKUI T, MATSUI K, HIRA K, SHINOZUKA M, EZAKI H, OTAKI K, KUROKAWA W, IMURA H, KOYAMA H, SHIMBO T. Interpretation and preference for probability expressions among Japanese patients and physicians. Fam Pract 2002 Feb;19(1):7–11.

34. WALLSTEN TS, FILLENBAUM S, COX JA. Base rate effects on the interpretations of probability and frequency expressions. J Mem Lang 1986 Oct;25(5):571–87.

35. WOLOSHIN KK, RUFFIN MT, GORENFLO DW. Patients' interpretation of qualitative probability statements. Arch Fam Med 1994 Nov;3(11):961–6.

36. OHNISHI, op. cit.

37. OLSON MJ, BUDESCU DV. Patterns of preference for numerical and verbal probabilities. J Behav Decis Making 1997 Jun;10(2):117–31.

38. WALLSTEN, 1993, op. cit.

39. WALLSTEN, 1993, op. cit.

40. CALMAN KC. Cancer: science and society and the communication of risk. BMJ 1996 Sep 28;313(7060):799–802.

41. TAVANA M, KENNEDY DT, MOHEBBI B. An applied study using the analytic hierarchy process to translate common verbal phrases to numerical probabilities. J Behav Decis Making 1997 Jun;10(2):133–50.

42. IBID.

43. BIEHL M, HALPERN-FELSHER B. Adolescents' and adults' understanding of probability expressions. J Adolesc Health 2001 Jan;28(1):30–5.

44. CALMAN, op. cit.

45. EREV I, COHEN BL. Verbal versus numerical probabilities: efficiency, biases, and the preference paradox. Organ Behav Hum Decis Process 1990 Feb;45(1):1–18.

46. WALLSTEN T. The costs and benefits of vague information. In: Hogarth RM, ed. Insights in decision making: a tribute to Hillel J. Einhorn. Chicago, IL: University of Chicago Press, 1990:28–43.

47. MSNBC NEWS. SARS risk on planes appears low. [Web document]. MSN, 2003. [rev. Jun 2003; cited 7 Oct 2003]. <http://www.msnbc.com/news/922633.asp?cp1=1>.

48. WALLSTEN, 1990, op. cit.

49. ELLSBERG D. Risk, ambiguity, and the Savage axioms. Q J Econ 1961 Nov;75(4):643–69.

50. BRASE GL. Which statistical formats facilitate what decisions? the perception and influence of different statistical information formats. J Behav Decis Making 2002 Dec;15(5): 381–401.

51. BRASE GL, COSMIDES L, TOOBY J. Individuation, counting, and statistical inference: the role of frequency and whole object representations in judgment under uncertainty. J Exp Psychol Gen 1998 Mar;127(1):1–19.

52. COSMIDES L, TOOBY J. Are humans good intuitive statisticians after all? rethinking some conclusions from the literature on judgment under uncertainty. Cognition 1996 Jan; 58(1):187–276.

53. GIGERENZER G. Ecological intelligence: an adaptation for frequencies. In: Cummins DD, Allen C, eds. The evolution of mind. New York, NY: Oxford University Press, 1998:9–29.

54. DENES-RAJ V, EPSTEIN S, COLE J. The generality of the ratio-bias phenomenon. Pers and Soc Psychol Bull 1995 Oct; 21(10):1083–92.

55. YAMAGISHI I. When a 12.86% mortality is more danger-

ous than 24.14%: implications for risk communication. Appl Cogn Psychol 1997 Dec;11(6):495–506.

56. DENES-RAJ V, EPSTEIN S. Conflict between intuitive and rational processing: when people behave against their better judgment. J Pers Soc Psychol 1994 May;66(5):819–29.

57. KAHNEMAN, op. cit.

58. GRIMES DA, SNIVELY GR. Patients' understanding of medical risks: implications for genetic counseling. Obstet Gynecol 1999 Jun;93(6):910–4.

59. WOLOSHIN S, SCHWARTZ LM, BYRAM S, FISCHOFF B, WELCH HG. A new scale for assessing perceptions of chance: a validation study. Med Decis Making 2000 Jul–Sep;20(3): 298–307.

60. COSMIDES, op. cit.

61. GIGERENZER G, HOFFRAGE U. How to improve Bayesian reasoning without instruction: frequency formats. Psychol Rev 1995 Oct;102(4):684–704.

62. AARON E, SPIVEY-KNOWLTON M. Frequency vs. probability formats: framing the three doors problem. Proceedings of the Twentieth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates, 1998: 13–8.

63. HOFFRAGE U, GIGERENZER G. Using natural frequencies to improve diagnostic inferences. Acad Med 1998 May;73(5): 538–40.

64. GIGERENZER, 1998, op. cit.

65. HAMM RM, SMITH SL. The accuracy of patients' judgments of disease probability and test sensitivity and specificity. J Fam Pract 1998 Jul;47(1):44–52.

66. HOFFRAGE, 1998, op. cit.

67. GIGERENZER G. Adaptive thinking: rationality in the real world. New York, NY: Oxford University Press, 2000:62–8.

68. JARVENPAA SL. Graphic displays in decision making—the visual salience effect. J Behav Decis Making 1990 Oct–Dec;3(4):247–262.

69. WOLOSHIN, 2001, op. cit.

70. FULLER R, DUDLEY N, BLACKTOP J. Risk communication and older people—understanding of probability and risk information by medical inpatients aged 75 years and older. Age Ageing 2001 Nov;30(6):473–6.

71. FULLER R, DUDLEY N, BLACKTOP J. How informed is consent? understanding of pictorial and verbal probability information by medical inpatients. Postgrad Med J 2002 Sep; 78(923):543–4.

72. SCHAPIRA MM, NATTINGER AB, MCHORNEY CA. Frequency or probability? a qualitative study of risk communication formats used in health care. Med Decis Making. 2001 Nov–Dec;21(6):459–67.

73. RUDSKI J, VOLKSDORF J. Pictorial versus textual information and the ratio-bias effect. Percept Mot Skills 2002 Oct; 95(2):547–54.