Health professionals’ and service users’ interpretation of screening test results: experimental study

Ros Bramwell, Helen West, Peter Salmon

Abstract

Objective To investigate the accuracy of interpretation of probabilistic screening information by different stakeholder groups and whether presentation as frequencies improves accuracy.

Design Between participants experimental design; participants responded to screening information embedded in a scenario.

Setting Regional maternity service and national conferences and training days.

Participants 45 pregnant women attending their first antenatal appointment in a regional maternity service; 40 companions accompanying the women to their appointments; 42 midwives; 41 obstetricians.

Participation rates were 56%, 48%, 89%, and 71% respectively.

Measures Participants estimated the probability that a positive screening test result meant that a baby actually had Down’s syndrome on the basis of all the relevant information, which was presented in a scenario. They were randomly assigned to scenarios that presented the information in percentage (n = 80) or frequency (n = 83) format. They also gave basic demographic information and rated their confidence in their estimate.

Results Most responses (86%) were incorrect. Obstetricians gave significantly more correct answers (although still only 49%) than either midwives (0%) or pregnant women (5%). Overall, the proportion of correct answers was higher for presentation as frequencies (24%) than for presentation as percentages (6%), but further analysis showed that this difference occurred only in responses from obstetricians. Many health professionals were confident in their incorrect responses.

Conclusions Most stakeholders in pregnancy screening draw incorrect inferences from probabilistic information, and health professionals need to be aware of the difficulties that both they and their patients have with such information. Moreover, they should be aware that different people make different mistakes and that ways of conveying information that help some people will not help others.
**Introduction**

Extensive psychological research has shown that most people, including health professionals, incorrectly interpret probabilistic information from screening tests. Laboratory research suggested extreme overestimation or underestimation of the probability that a positive screening result indicated that the relevant condition was present. Scenarios describing medical screening produced more overestimation than did ones describing screening of machine parts. The suggestion that evolution and experience equip people better to understand probabilistic information expressed as frequencies in a population, rather than as probabilities for an individual, led to the recommendation that medical practitioners should present screening information as frequencies, perhaps pictorially. We compared presentation as percentages and frequencies on the level of correct estimates and the types of error made in response to a scenario based on screening for Down's syndrome. We compared responses of obstetricians, midwives, pregnant women, and their companions at an antenatal appointment and their confidence in their ratings.

**Methods**

**Participants**

We recruited participants from four stakeholder groups: pregnant women, people accompanying them to antenatal appointments, midwives, and obstetricians. We randomised participants to one of four scenarios (two health screening; two machine parts screening).

Numbers who were approached and participated in each group for the study were as follows: 151 pregnant women approached, 82 (54%) responded, of whom 43 received the health scenario; 166 companions approached, 80 (48%) responded, of whom 40 received the health scenario; 92 midwives approached, 80 (48%) responded, of whom 41 received the health scenario; 166 companions approached, 80 (48%) responded, of whom 40 received the health scenario. Table 1 shows the number in each group who responded to a frequency or percentage scenario.

**Procedure**

We recruited pregnant women and the people accompanying them at a regional maternity service and health professionals at national training events or through a regional maternity service. Questionnaires presented a screening scenario (see box) that described a positive screening result presented in percentage format (n = 86) or frequency format (n = 83). We asked respondents to estimate the probability that a positive test result meant that the baby had Down's syndrome. They rated their confidence in their answer from 1 (not at all confident) to 6 (very confident).

**Analysis**

The correct response was 47.6% (see bmj.com). We categorised estimates from 45.0% to 50.0% as correct. We regarded all other responses as incorrect and categorised them into overestimates and underestimates.

**Results**

As expected, most (n = 142; 86%) responses were incorrect. Whereas the correct answer was 47.6%, most responses were close to 0% or 100%. The two most frequent answers were 1.0% (n = 32; 19% of the sample) and 90.0% (n = 46; 27% of the sample), and these were produced by all groups and in response to both presentations.

**Table 1** Percentage (number) of respondents from each stakeholder group who provided answers that were correct, overestimates, or underestimates

<table>
<thead>
<tr>
<th>Presentation as percentages</th>
<th>Pregnant women</th>
<th>Companions</th>
<th>Midwives</th>
<th>Obstetricians</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>9 (1)</td>
<td>15 (3)</td>
<td>0</td>
<td>5 (1)</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Overestimate</td>
<td>68 (10)</td>
<td>50 (10)</td>
<td>46 (10)</td>
<td>76 (16)</td>
<td>91 (21)</td>
</tr>
<tr>
<td>Underestimate</td>
<td>27 (4)</td>
<td>38 (8)</td>
<td>35 (7)</td>
<td>19 (4)</td>
<td>44 (29)</td>
</tr>
<tr>
<td>Total</td>
<td>100 (22)</td>
<td>100 (20)</td>
<td>100 (22)</td>
<td>100 (21)</td>
<td>100 (85)</td>
</tr>
</tbody>
</table>

**Table 2** Percentage (number) of respondents from each stakeholder group who provided answers that were correct, overestimates, or underestimates

**Between group differences**—The groups differed in the proportion of correct answers (table 1; \( \chi^2 = 20.9, \text{df} = 3, P < 0.001 \)). Obstetricians were more often correct than either pregnant women (\( P = 0.007 \)) or midwives (\( P < 0.001 \)), and companions were more often correct than midwives (\( P = 0.011 \)). The groups did not differ in the proportion of overestimates versus underestimates (\( \chi^2 = 6.4, \text{df} = 3, P = 0.093 \)).

**Frequency presentation versus percentage presentation**—The proportion of correct answers was much higher to presentation as frequencies than to presentation as percentages (table 1; \( \chi^2 = 10.4, \text{df} = 1, P = 0.001 \)). However, this overall effect concealed an important between group difference. In pregnant women, companions, and midwives, the proportion of correct answers remained low whether presentation was as frequencies or percentages, with no difference between the two. In obstetricians, presentation as frequencies produced many more correct answers than did presentation as percentages (\( \chi^2 = 16.5, \text{df} = 1, P < 0.01 \)). Presentation also influenced the types of error made (\( \chi^2 = 6.7, \text{df} = 1, P = 0.010 \)). Percentages produced more overestimates, and frequencies produced more underestimates.

**Confidence in responses**—Table 2 shows that, even among respondents who answered incorrectly, many were confident in their responses. Whereas most of the pregnant women who gave incorrect answers scored in the lower range of the confidence rating, people who were in a position to advise them were more confident. In particular, obstetricians who gave incorrect...
What is already known on this topic

Most people, including health professionals, do not draw mathematically correct inferences from probabilistic screening information.

Some studies suggest that presentation as frequencies aids interpretation.

What this study adds

Presentation as frequencies does not help everyone: a simple change from percentages to frequencies increased correct responses in obstetricians but not in midwives or service users.

The change in presentation did change the type of errors that people made.

Many respondents were very confident about their incorrect answers.

Discussion

In this study, a simple change from presentation as percentages to presentation as frequencies improved the accuracy of interpretation of information about screening for Down's syndrome, but only obstetricians benefited and only 65% of these were correct. Almost all the potential users of the test—that is, pregnant women, their companions, and midwives—were still incorrect.

The benefits of presentation as frequencies therefore depended on the characteristics of the respondents. Most respondents judged that the genetic anomaly was almost certainly present or almost certainly absent.

Although presentation as frequencies did not increase accuracy overall, it did significantly change the balance of overestimates versus underestimates. Given that the errors are so extreme, a minor change in presentation can have a major impact on the interpretation of results of screening tests.

One of the main criticisms of previous probabilistic reasoning research is that it lacks ecological validity.4 However, the screening test was one in which each respondent group was potentially involved, and the elements of information in the scenario match those covered in the standard NHS leaflet prepared by the UK National Screening Committee.

Screening scenario

Version 1: percentages

The serum test screens pregnant women for babies with Down's syndrome. The test is a very good one, but not perfect. Roughly 1% of babies have Down's syndrome. If the baby has Down's syndrome, there is a 90% chance that the result will be positive. If the baby is unaffected, there is still a 1% chance that the result will be positive. A pregnant woman has been tested and the result is positive. What is the chance that her baby actually has Down's syndrome?

Version 2: frequencies

The serum test screens pregnant women for babies with Down's syndrome. The test is a very good one, but not perfect. Roughly 100 babies out of 10,000 have Down's syndrome. Of these 100 babies with Down's syndrome, 90 will have a positive test result. Of the remaining 9900 unaffected babies, 99 will still have a positive test result. How many pregnant women who have a positive result to the test actually have a baby with Down's syndrome?

Readers might be reassured by the finding that more obstetricians were correct, but midwives are the main source of information for pregnant women about this test, and nearly two thirds of obstetricians were incorrect. Many obstetricians and midwives were confident in their incorrect answers, indicating a disturbing lack of insight into their poor understanding of information directly relevant to their clinical practice.

Limitations of the study

The theoretical justification for using presentation as frequencies is that it facilitates a more "natural" style of reasoning. Some people have suggested that truly effective communication of probabilistic information will need to use decision aids such as visual presentations of risk.3 We chose to make the minimum changes necessary to convey information in a frequency format consistent with previous research. Nevertheless, this needed minor wording of the scenario, and different changes might have led to different results. Presentation of information in frequency format for some combinations of key screening test parameters would require the use of larger numbers than the 10,000 denominators used here.

Future research will need to examine whether users' difficulty in thinking about large numbers might counteract the benefits of presentation as frequencies. Finally, participation was voluntary, and clearly the study may be biased towards those who felt more comfortable with probabilistic data.

Implications for practice

Comparisons between stakeholders in screening highlight the importance of future research with user groups and non-medical professionals. Health professionals need to be aware that screening information presents difficulties to professionals and service users alike and that the erroneous conclusions being drawn by different groups may differ. The inability of the people actually using probabilistic screening information, both professionals and service users, to draw correct conclusions from it seriously challenges the usefulness of such screening in practice.

Contributors: See bmj.com.

Funding: Economic and Social Research Council, award reference RES-060-22-0352. The funders were not involved in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Competing interests: None declared.

Ethical approval: Liverpool (Adult) Research Ethics Committee 9/1/04, reference 02/05/063A.


doi 10.1136/bmj.38884.663.102.AE

286 BMJ VOLUME 333 5 AUGUST 2006 bmj.com