COMMUNICATING CONDITIONAL PROBABILITIES IN MEDICAL PRACTICE

Katharina Böcherer-Linder[1], Karin Binder[2], Theresa Büchter[3], Andreas Eichler[3], Stefan Krauss[4], Nicole Steib[4], and Markus Vogel[5]
[1] Albert-Ludwigs-University Freiburg, Germany
[2] Ludwig-Maximilians-University Munich, Germany
[3] University of Kassel, Germany
[4] University of Regensburg, Germany
[5] University of Education Heidelberg, Germany
boecherer-linder@math.uni-freiburg.de

Patients need to be informed correctly and comprehensibly about the implications of their medical test results. Reasoning in such situations, where, for example, a medical test result is used to make inferences on a particular disease, is called Bayesian reasoning. Prior research mostly concentrated on the ability to correctly calculate risks in Bayesian situations (so-called performance) and repeatedly demonstrated that performance is very low—even among medical experts. The need to also study communication within Bayesian situations has been brought forward. Here, we broaden the focus of Bayesian reasoning and present first insights into a study where medical students participated in a training course on the aspect of performance and show that this already improves the ability to judge doctor–patient communication within Bayesian situations.

INTRODUCTION

Informed consent and shared decision making are regarded as essential in the — interaction, which requires a transparent and understandable communication of risks to patients (Gigerenzer et al., 2007). For example, imagine the following situation: A pregnant 45-year-old woman has participated in a prenatal screening in order to detect possible Down syndrome in the unborn child. The woman tested positive, which may suggest that her unborn child has Down syndrome. The following statistics are available to the physician about 45-year-old pregnant women and about the test that was used in the prenatal screening:

- The probability is 3% that a 45-year-old woman's unborn child has Down syndrome (*prevalence*).
- If a 45-year-old pregnant woman's unborn child has Down syndrome, then the probability is 75% that this pregnant woman tests positive (*sensitivity*).
- If a 45-year-old pregnant woman's unborn child does *not* have Down syndrome, then the probability is 5% that this pregnant woman tests positive nevertheless (*false-positive rate*).

Naturally, the following questions arise:

- Question 1: What is the risk that the 45-year-old pregnant woman's unborn child actually has Down syndrome, given the positive test result (positive predictive value)?
- Question 2: How should the physician communicate this risk to the woman?

In the following, we will address these questions by referring to research about Bayesian reasoning. The structure of the described situation is typical for a Bayesian situation because binary information (e.g., positive vs. negative test result) is used to make inferences on a binary hypothesis (e.g., whether the unborn child has Down syndrome or not; Zhu & Gigerenzer, 2006). The ability to reason in such situations is called Bayesian reasoning.

Previous research in the field of Bayesian reasoning has often only studied the ability to calculate specific risks in Bayesian situations. Calculating the probability of the positive predictive value, as in question 1, can be done with Bayes' formula: $\frac{0.03 \cdot 0.75}{0.03 \cdot 0.75 + 0.97 \cdot 0.05} \approx 32\%$, resulting in the (unintuitively low) probability of only about 32%. However, it has repeatedly been pointed out, that research on Bayesian reasoning should also address further aspects, such as the communication of these risks as in question 2 (Gigerenzer et al., 2007; Navarrete et al., 2015) and the ability to evaluate effects of variations in the given information (Borovcnik, 2012). Therefore, in the project TrainBayes (http://www.bayesianreasoning.de/en/bayes_en.html), we define and study *three* aspects of Bayesian reasoning:

- *Performance*: The ability to calculate a conditional probability (e.g., the positive predictive value) in a Bayesian situation.
- *Covariation*: The ability to adequately evaluate effects of changes of the given parameters (e.g., prevalence, sensitivity, or false-positive rate) on the positive predictive value.
- *Communication*: The ability to appropriately communicate probabilistic information in a Bayesian situation in an expert-laymen setting (e.g., doctor–patient communication).

In this paper, we want to provide first insights into the results of a study concerning the aspect of communication. We consider performance in Bayesian reasoning fundamental for communication because it is necessary to correctly assess a risk in a Bayesian situation first, before being able to adequately communicate it. Therefore, while we want to focus on communication in this contribution, we first report on the fundamental insights of previous research on the aspect of performance. The study was recently completed, and therefore, data analysis is still ongoing. Consequently, we will primarily present qualitative insights but cannot provide final quantitative results at this stage.

THEORETICAL BACKGROUND

In previous research, which mainly focussed on the aspect of performance, it has been shown that laymen and (medical) experts alike struggle immensely to correctly evaluate risks in Bayesian situations (Eddy, 1982; McDowell & Jacobs, 2017). However, two strategies have been identified as being successful for improving performance in Bayesian situations: First, people are more likely to find a correct solution for a task as in question 1, if the statistical information is represented in form of so-called *natural frequencies* (Gigerenzer & Hoffrage, 1995). Thereby, the given information is related to a sample (e.g., 10,000 women, who are 45 years old and participate in the prenatal screening) and the probabilities are represented in pairs of frequencies:

- Prevalence: *300 out of 10,000* 45-year-old pregnant women's unborn children have Down syndrome.
- Sensitivity*: 225 out of the 300* 45-year-old pregnant women whose unborn children have Down syndrome test positive.
- False-positive rate: *485 out of the 9,700* 45-year-old pregnant women whose unborn children do not have Down syndrome test positive, nevertheless.

This representation of the statistical information facilitates the identification of the positive predictive value, i.e., 225 out of 710 45-year-old pregnant women who test positive (225 plus 485), actually have unborn children with Down syndrome, resulting in $\frac{225}{225+485} \approx 32\%$ as the probability for the positive predictive value.

The second successful strategy, which has been identified in the research on performance, refers to structuring the given information in a suitable visualization (e.g., Binder et al., 2021; Eichler et al., 2020). Adding this strategy to using natural frequencies increases performance in Bayesian situations even further (e.g., Binder et al., 2020). Comparisons between different visualizations showed that there are more and less helpful visualizations for supporting performance. The so-called double-tree and unit square are considered particularly promising (Binder et al., 2020; Böcherer-Linder & Eichler, 2019).

Transferring these results (which are known from research on the aspect of performance in Bayesian situations) to the aspect of communication means: Using natural frequencies should also be preferred to only using probabilities with regard to communicating the positive predictive value, because it increases the understanding of the patients. We consider this one promising strategy for communication in Bayesian situations and call this the *frequency-strategy*. Moreover, it has also been suggested that using adequate visualizations, which represent the risks in the given situation, increases the understanding of patients, which we call the *visualization-strategy*. From further research about communicating risks (which was not explicitly carried out in the context of Bayesian reasoning) it has additionally been shown that it is helpful to also interpret the concrete risk (e.g., 32%) within the context, for instance as a low, medium, or high risk (Ellermann et al., 2022). We call this the *interpretation-strategy*. Furthermore, previous research on doctor–patient communication has (among others) addressed the necessity to explain technical terms, which might otherwise often be misunderstood, such as the term "positive test result" (Auschra et al., 2020), which we call the *explanation-strategy*. Finally, research on how to communicate statistical information with visualization suggested avoiding framing bias by describing both positive and negative outcomes (Spiegelhalter et al., 2011). Transferring this idea to communication, both positive and negative outcomes should also be described, which we call the *balance-strategy*. Of course, doctor–patient communication entails many further important aspects such as creating a safe environment, showing empathy, inviting the patients to ask questions, etc. (Buckman, 2005). However, we focus here on describing the first aspects because it can also be transferred to expert–laymen communication in Bayesian situations outside the specific field of medicine, where Bayesian reasoning can be considered just as important (e.g., in law or economy).

In the few previous studies that have addressed communication in Bayesian situations, it has been shown that communication of medical experts in Bayesian situations is only very rarely adequate (Ellis & Brase, 2015; Prinz et al., 2015). For instance, it was demonstrated that while talking to a patient, about half of the consultants communicated an illusion of certainty for test parameters (e.g., sensitivity and false-positive rate) as well as for the positive predictive value (Prinz et al., 2015). Further critique about the quality of consultants' communication entailed the lack of using the frequency-strategy because the majority of consultants used probabilities or non-numerical values to describe the risk (Ellis & Brase, 2015). Therefore, the need arises to improve medical experts' Bayesian reasoning with regard to communication. As spelled out before, performance is a requirement for communication, and also the strategies that are supportive for performance may improve communication. Thus, we study the effect of a training that addresses the aspect of performance on the communication in Bayesian situations.

In the following we present first insights on the effect of participating in a training on Bayesian reasoning that focuses on the aspects of performance and covariation, on the ability to judge communication of medical experts in Bayesian situations. We study the ability to *judge* the communication of a medical expert in a Bayesian situation as a measurement of the aspect of communication, due to the following reasons. Identifying more and less appropriate strategies among doctor–patient communication is a precondition for successfully communicating test results on your own. Also, as spelled out above, successful doctor–patient communication entails more than just aspects that are closely linked to Bayesian reasoning. Thus, by testing how people judge different examples of doctor–patient communications, which systematically vary with regard to the strategies linked to Bayesian reasoning, we can more reliably test communication as an aspect of Bayesian reasoning.

METHOD

In the training study (see Figure 1), communication was examined in a pre-, post- and follow-up design. The follow-up test was conducted 8–10 weeks after the post-test. There was no separate training on the aspect of communication. However, there were four different training courses, each on the aspect of performance and covariation, which took place between pre-test and post-test and lasted about one hour. Moreover, there was a control group without training. In developing the training materials, previous research on helpful strategies was used as a key foundation. A detailed description of the training courses on the aspect of performance and covariation can be found in (Büchter et al., 2022). Our participants were 260 medical students. Participation in the study was voluntary; written informed consent was obtained from the participants; and they received payment for participation.

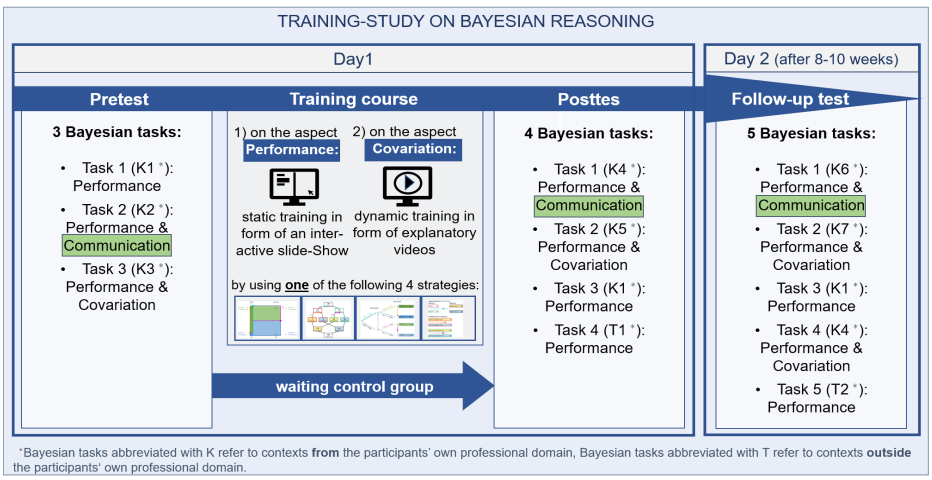

Figure 1. Design of the described training study on Bayesian reasoning

In the pre-test, post-test, and follow-up tests, communication was measured after a question on performance in the following way: The students watched six different videos in which a physician explains to a patient (for example, a 45-year-old pregnant woman from the example above) what the positive test result means. At the beginning, a cover story addressed important strategies specific to communication among physicians and patients and thus explained in advance that the physician empathetically communicates the positive test result to the patient and that the videos only differ with regards to interpreting the test result. The videos had to be rated on a scale from inappropriate to absolutely appropriate (see Figure 3). Following the rating of the videos on the scale, the students were asked to provide reasoning for their positioning of the videos on the scale. They were specifically asked to justify why a video was rated as the best video, especially in contrast to the second-best video.

As displayed in Figure 2, the videos were varied in the following way (compare supplementary material): In principle, there were always two videos communicating the correct value (positive predictive value: 32%) and four videos in which the physician makes a typical mistake and thus communicates the wrong probabilistic information (namely the sensitivity, 75%, or the difference between sensitivity and the false positive rate, 75%−5% = 70%) as the positive predictive value. We call two videos that address the same probabilistic information, twin videos, and explain in the following how these two videos differ regarding the strategies explained above. In one of the two videos the probability was also related to a sample, hence the frequency-strategy was applied. Moreover, in one of twin videos, the balance-strategy was used, which means that not only the positive predictive value but also its complement is discussed (i.e., the probability to be healthy despite the positive test result). In addition, one of the twin videos always explained what the positive result means (in the above example, it means that there is a suspicion of trisomy 21 in the unborn child) and thus applied the explanation-strategy. Finally, the videos also varied with regard to the interpretation-strategy, for example, interpreting the probability in the context as a “first hint” or a “hint that needs to be taken seriously.” We did not vary the visualization-strategy between the videos because this would have affected the rest of our study too much; using visualization was one of the central elements in the training courses.

Video number (compare Fig. 2)	Value communicated as positive predictive value	Frequency-strategy	Explanation-strategy	Balance-strategy
Video 5	positive predictive value (correct)	yes	yes	false and true positives
Video 2	positive predictive value (correct)	no	no	only true positives
Video 1	sensitivity (incorrect)	yes	no	only true positives
Video 4	sensitivity (incorrect)	no	yes	false and true positives
Video 6	difference between sensitivity and the false positive rate (incorrect)	yes	yes	false and true positives
Video 3	difference between sensitivity and the false positive rate (incorrect)	no	no	only true positives

Figure 2. Overview of the variation in the videos

RESULTS

Figure 3 shows how a student evaluated the videos in the pre-test (thus without any explanation of performance and covariation) and in the post-test (thus after having participated in the training on performance and covariation). In the pre-test (see the video boxes with black background in Figure 3), the student obviously did not realize that only in video 2 and video 5 the correct probability is communicated, because video 4, in which a wrong probability is communicated, was rated best. Furthermore, in the two videos that were rated highest, there is no reference of the probabilities to a sample, even though this is known to be easier to understand for a patient. As reasoning for video 4 as the best video, the student stated: “The probabilities are explained quite simply and clearly.” So, this reasoning is quite general and only refers to probabilities.

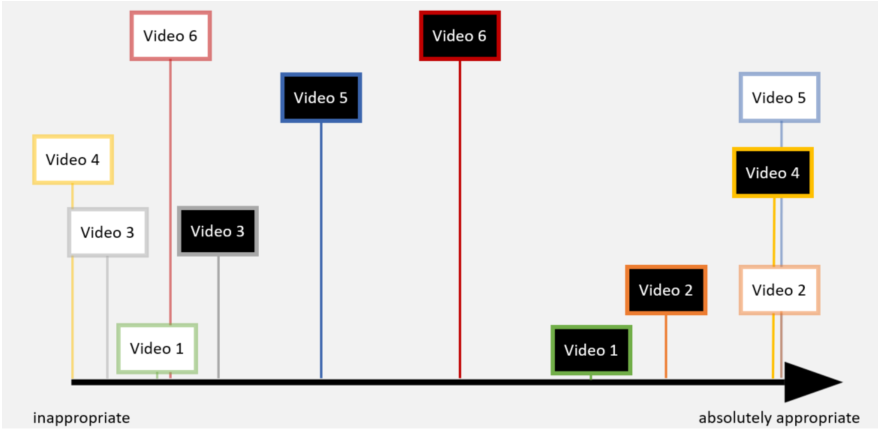

Figure 3. Arrangement of videos of a participant in the pretest (video boxes with black background) and posttest (video boxes with white background).

In the post-test (see the video boxes with white background in Figure 3), it becomes clear that now the videos with the correct probability can be clearly distinguished from those with an incorrect probability. In the reasoning, one had to decide for the best video. The student chose video 5 and then had to state why video 5 is better in differentiation to video 2. The following reasoning was given: "The positive predictive value is mentioned, which is the most important. Also, the comparison with the 100 positive women is certainly well understood by the patient." So, the correct probability was recognized and that facilitating strategy (i.e., translating the probabilistic information to frequencies) was also identified. Despite that quantitative data analysis is not finished yet, it can be already noted that our four trainings (on the aspect of performance and covariation) improve the judgment of communication (compared to the waiting control group). Moreover, this improvement due to training can still be observed in the follow-up test. However, it should also be noted that a considerable number of the medical students already identified the videos with incorrect probabilities in the pre-test.

CONCLUSION

Although Bayesian reasoning has been studied in the past mainly in relation to the aspect of "performance," we wanted to broaden this field and take a closer look at the aspect of communication. The present study aims to give first insights into how the aspect of communication is positively influenced by a training of the aspects on performance and covariation. Research on communication as an aspect of Bayesian reasoning should be pursued further in the future because it sheds light on a previously neglected aspect of Bayesian reasoning, which is an essential competence in dealing with statistical information and of particular importance in an expert–laymen setting as in doctor–patient communication. Furthermore, such training courses should find their way into learning scenarios in various university courses.

SUPPLEMENTARY MATERIAL

Supporting information can be downloaded at: https://osf.io/hzwsg/

FUNDING

The research was funded by the DFG—German Research Foundation (EIC773/4-1).

REFERENCES

Auschra, C., Möller, J., Berthod, O., Mazheika, Y., & Borusiak, P. (2020). Befundergebnisse verständlich vermitteln—Eine randomisiert-kontrollierte studie zur wortwahl in der ärzt*innen-patient*innen-Kommunikation [Communicating test results in a comprehensible manner: A randomized controlled trial of word usage in doctor-patient communication]. *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen*, *156-157*, 40–49. https://doi.org/10.1016/j.zefq.2020.07.007

Binder, K., Krauss, S., Schmidmaier, R., & Braun, L. T. (2021). Natural frequency trees improve diagnostic efficiency in Bayesian reasoning. *Advances in Health Sciences Education*, *26*(3), 847–863. https://doi.org/10.1007/s10459-020-10025-8

Binder, K., Krauss, S., & Wiesner, P. (2020). A new visualization for probabilistic situations containing two binary events: The frequency net. *Frontiers in Psychology*, *11*, Article 750. https://doi.org/10.3389/fpsyg.2020.00750

Böcherer-Linder, K., & Eichler, A. (2019). How to improve performance in Bayesian inference tasks: A comparison of five visualizations. *Frontiers in Psychology*, *10*, Article 267. https://doi.org/10.3389/fpsyg.2019.00267

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances De Investigación en Educación Matemática*, *1*(2), 5–27. https://doi.org/10.35763/aiem.v1i2.32

Büchter, T., Eichler, A., Steib, N., Binder, K., Böcherer-Linder, K., Krauss, S., & Vogel, M. (2022). How to train novices in Bayesian reasoning. *Mathematics*, *10*(9), Article 1558. https://doi.org/10.3390/math10091558

Buckman, R. A. (2005). Breaking bad news: The S-P-I-K-E-S strategy. *Community Oncology*, *2*(2), 138–142. https://doi.org/10.1016/s1548-5315(11)70867-1

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (1st ed., pp. 249–267). Cambridge University Press. https://philpapers.org/rec/EDDPRI

Eichler, A., Böcherer-Linder, K., & Vogel, M. (2020). Different visualizations cause different strategies when dealing with Bayesian situations. *Frontiers in Psychology*, *11*, Article 1897. https://doi.org/10.3389/fpsyg.2020.01897

Ellermann, C., McDowell, M., Schirren, C. O., Lindemann, A.-K., Koch, S., Lohmann, M., & Jenny, M. A. (2022). Identifying content to improve risk assessment communications within the Risk Profile: Literature reviews and focus groups with expert and non-expert stakeholders. *PLOS ONE*, *17*(4), Article e0266800. https://doi.org/10.1371/journal.pone.0266800

Ellis, K. M., & Brase, G. L. (2015). Communicating HIV results to low-risk individuals: Still hazy after all these years. *Current HIV Research*, *13*(5), 381–390. https://doi.org/10.2174/1570162x13666150511125629

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, *143*(12), 1273–1312. https://doi.org/10.1037/bul0000126

Navarrete, G., Correia, R., Sirota, M., Juanchich, M., & Huepe, D. (2015). Doctor, what does my positive test mean? From Bayesian textbook tasks to personalized risk communication. *Frontiers in Psychology*, *6*, Article 1327. https://doi.org/10.3389/fpsyg.2015.01327

Prinz, R., Feufel, M. A., Gigerenzer, G., & Wegwarth, O. (2015). What counselors tell low-risk clients about HIV test performance. *Current HIV Research*, *13*(5), 369–380. https://doi.org/10.2174/1570162x1366615051112520

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, *333*(6048), 1393–1400. https://doi.org/10.1126/science.1191181

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*(3), 287–308. https://doi.org/10.1016/j.cognition.2004.12.003