TEACHING PROBABILITIES AND RISKS TO MEDICAL STUDENTS WITH A SPECIAL FOCUS ON COMMUNICATION WITH PATIENTS

Andreas Mayr, Tobias Fleckenstein, Thomas Kneib, and Bruno de Sousa
University Hospital Bonn, Germany
Georg-August University Göttingen, Germany
University of Coimbra, Portugal
amayr@uni-bonn.de

Medical doctors are frequently confronted with various tools for risk assessment, forecast models, and diagnostics based on conditional probabilities, but statistical literacy among medical students is often low. This leads to a potential loss of information when communicating with patients and poorer decision making. We argue that a special focus in teaching probabilities and risks to medical students not only should be on skills regarding the correct interpretation of the different measures from a clinical perspective but also should focus on how to communicate these risks via natural frequencies to patients. We identified some typical settings and small case-studies that illustrate potential solutions that could be used when teaching statistics to medical students, to ensure proper communication with patients.

BACKGROUND

Low statistical literacy among medical students or doctors is an often-discussed topic among statisticians working in medicine. There are numerous studies and articles that try to assess and quantify the problem, often via questionnaires and scores testing statistical numeracy, risk literacy, or knowledge of simple statistical measures and procedures (e.g., Cokely et al., 2012). Most authors in this context focus on the statistical skills needed by doctors to correctly interpret the results of clinical trials and medical research articles or to perform their own research (e.g., Altman & Bland, 1991). However, the correct interpretation and communication of probabilities and risks not only is a challenge in research but also is very relevant for routine medical care.

In clinical decision making and in everyday clinical practice, doctors are often confronted with uncertainty that is typically quantified via probabilities or risks. This starts with the diagnosis of a patient's particular illness, which is often based on diagnostic tests or simple decision rules that almost never can be 100% accurate. A particular challenge in this context is that the typical measures for performance of diagnostic tests (sensitivity, specificity) are in fact conditional probabilities that are conditioned on the true status of the patient (disease or no disease). In clinical practice, however, this true status of the patient is, of course, unknown and part of the inherent uncertainty associated with practice. Assessing and communicating the probability of a true outcome *after* a test result is known might be far more relevant in clinical practice; however, this leads to an additional challenge in which predictive values depend on the prevalence of the disease among those being tested (a priori or pre-test probability). Many medical doctors have severe problems interpreting the correct probability of a patient being diseased given a positive test result (Casscells et al., 1978; Hoffrage et al., 2000).

The appearance of side effects or adverse events associated with particular treatments, therapies, or surgeries are naturally unknown for a single patient beforehand. Typically, there are probability estimates available on the population level (e.g., incidence), and in some settings there might be also risk scores available for the individual patient (Moons et al., 2015). But still, the challenge remains how to communicate these risks so that patients can make informed decisions.

Another related important aspect is communication of risk factors that increases a person's liability for a certain negative clinical outcome. In clinical research, these are typically quantified as relative risks (compared to a patient without this risk factor). However, the correct interpretation and communication of these relative risk estimates often pose severe problems for doctors and patients as both the true reference class as well as the underlying absolute risk can be easily overlooked.

In addition to the specific challenges of the described settings, many authors argue that for the communication of risks and probabilities in general, natural frequencies are much easier to interpret and avoid common problems of ambiguities around reference classes or correct interpretation of percentages (Gigerenzer & Edwars 2003; Hoffrage et al., 2000). We will follow this frequency

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

approach and provide small case-studies for communication with patients, where we also focus on communicating frequencies instead of percentage values.

PROBABILITES AND RISK WITH HIGH RELEVANCE IN MEDICINE

Conditional Probabilities (Particularly in Diagnostic or Screening Tests)

In diagnostic tests, the main measures of accuracy are sensitivity and specificity, which reflect the conditional probabilities for a test outcome given the true status of the patient.

- Sensitivity: *P*(*test positive* | *disease*)
- Specificity: *P*(*test negative* | *no disease*)

In the development of a test based on a continuous marker (e.g., biomarker), the area under the receiver–operating characteristics curve (AUC) is often reported, particularly when different candidate markers are considered. The AUC is an overall measure that does not consider one single cutoff for binary decision making but summarizes the relationship between sensitivity and specificity over all possible cutoff values. Furthermore, it has to be stressed that the AUC is unfortunately sometimes confused by practitioners with the probability of a test giving a correct result. Some authors even argue that the AUC should be abandoned as measure of test performance (Wald & Bestwick, 2014).

When it comes to clinical practice, however, the main measures to judge uncertainty for a single patient cannot be sensitivity or specificity because they condition on the true status of the patient. This unknown true status was the reason to carry out diagnostics in the first place. Therefore, outside of medical research, the predictive values that condition on the outcome of the diagnostic test are far more relevant. Hence, it is necessary to raise awareness for communicating predictive values and not the more frequently available measures of a diagnostic test's performance (sensitivity, specificity).

- Positive predictive value: *P*(*disease* | *test positive*)
- Negative predictive value: *P*(*no disease* | *test negative*)

However, a remaining challenge in the use of predictive values is that they depend on the prevalence of the disease in the population being tested. This prevalence, also known as a priori or pretest probability, therefore depends on the current testing policies or recommendations followed by practitioners and might vary not only between populations and countries but also between other subgroups. The dependency of the positive predictive value on prevalence is often a major source of confusion for screening tests, e.g., for cancer. Due to the low prevalence of most cancer types among the general population (often there is high lifetime prevalence but only a very low point-prevalence), screening tests are most often associated with very low predictive values (e.g., mammography for breast-cancer screening, Prostate-Specific Antigen test for prostate cancer).

Probabilities for Side Effects or Adverse Events

In contrast to the conditional probabilities often necessary to address uncertainty with respect to the usage of diagnostic tests, the communication of risks for side effects or adverse events seems to be more straightforward. Typically, these are simple probabilities of a single event, e.g., the probability of suffering from a side-effect with the specific type of therapy.

A challenge arising in communicating these probabilities is that they typically do not specify the exact class of events to which they refer. Gigerenzer and Edwards (2003) report a psychiatrist prescribing Prozac (fluoxetine) to depressed patients, informing them about a "30% to 50% chance of developing a sexual problem." Later, the psychiatrists realized that some patients interpreted this information to mean having sexual problems in 30% to 50% of their sexual encounters. Although the correct reference class in this example refers to patients, some of these patients interpreted their own sexual encounters as the reference class. This problem can be avoided when using natural frequency statements instead of probabilities because these directly address the proper reference class (Gigerenzer & Edwards, 2003).

Relative Risks

To inform the public about potential risk factors (particularly lifestyle-related factors), public health agencies and the corresponding scientific literature typically refer to estimates for relative risks.

These estimates are often calculated from large cohort studies (e.g., via incidence ratios between groups with and without the risk factor) or are based on odds-ratios from case-control studies (comparing odds for being a case in the exposure group with the odds in the non-exposed group).

A typical challenge in the communication of relative risks is again the potentially misleading reference class that might often differ from the one the patient belongs to (Gigerenzer & Edwards, 2003). Another often discussed issue is the missing information about absolute risk. The effect of an exposure to a risk factor might seem enormous to a patient (e.g., four times more likely to develop a particular illness), but the patient might lose such fear when the underlying absolute risk (e.g., one in 10,000 without exposure and four in 10,000 with exposure) is communicated via frequency statements. These can be also accompanied with graphical representations such as icon arrays (see Figure 1, Galesic et al., 2009). On a population level, the attributable risk can be used because it combines the actual prevalence of the risk factor with the relative risk.

SMALL CASE STUDIES AND POTENTIAL COMMUNICATIONS WITH PATIENTS

Conditional Probabilities in Screening Tests

Case: A 65-year-old female patient underwent a mammography screening (sensitivity: 0.875, specificity 0.955, prevalence = 0.008) and got a positive result with an invitation for further exams (e.g., biopsy). The patient asks about her chances.

Potential Communication: Around eight out of 1,000 women have breast cancer; seven of those will get a positive result in the screening program on average. From the 992 patients without cancer, around 43 unfortunately get a false positive result, which will be later cleared by the biopsy. Therefore, around seven out of 50 patients who receive a positive mammography screening result actually have a tumor.

Comment: Instead of communicating conditional probabilities as percentages, natural frequencies are used. Information conditioning on the true status of the patient (sensitivity and specificity) is used via simple frequencies to derive the positive predictive value (around 14%) conditioning on the test result of the woman.

Probabilities for Side Effects

Case: A patient recently went into anaphylactic shock after being stung by a bee. The next step would be to propose a hypo sensitization (allergen immunotherapy); however, the patient needs to be informed about the risk of systematic reactions (estimated incidence around 10%).

Figure 1. Icon array to communicate the risk of side effects (www.iconarray.com)

Potential Communication: Ten out of 100 patients undergoing hypo sensitization show side-effects in the form of systematic reactions (see also Figure 1).

Comment: By avoiding percentages and ambiguity about the reference class, the corresponding frequency statement (one out of every 10 patients undergoing this particular treatment) helps the patient to assess the risks of the proposed treatment.

Relative Risks

Case: A female patient with multiple sclerosis asks about having children. She has read that the chance of getting multiple sclerosis increases by up to 3,000% for children of parents who also suffer from the illness.

Potential Communication: It is true that the risk for multiple sclerosis is 20–30 times higher for children where at least one parent suffers from multiple sclerosis compared to children with parents that are not affected (relative risk). This means that while in the general population around one out of 1000 children develop multiple sclerosis during their lifetime, this number increases to about 20 to 30 out of 1000 children when at least one parent is affected by the disease (absolute risk).

Comment: The high relative risk of 20–30 is communicated without relying on increased percentages, which are difficult to interpret. The relative risk is compared to the absolute risk of 2%–3%. By using natural frequencies (20–30 out of 1000 children with one affected parent) the reference class is automatically specified.

CONCLUSION

The overall low statistical literacy of medical students and doctors is a widely discussed and well reported issue (Casscells et al., 1978; Cokely et al., 2012; Hoffrage et al., 2000). In the era of digital medicine, larger amounts of available data, and genetic analyses, the communication of risks and the interpretation of probabilistic forecasts on a patient's health are getting even more essential not only for medical doctors, but also in classical medical routine, where it is fundamental that physicians interpret and communicate conditional probabilities or risks for single events.

We have identified three areas where problems in communicating risks and probabilities to patients are very common: conditional probabilities from diagnostic or screening tests, probabilities of side-effects or adverse events, and relative risks from the exposure to risk factors. We highlighted the challenges in their interpretation and provided three simple case-studies how the communication with patients can be improved using natural frequencies instead of percentages.

While the problem of lacking statistical skills can be tackled for doctors working in medical research by a strong biostatistical consulting with physicians and other healthcare professionals (LeBlanc et al., 2022), the same does not hold true for challenges arising in the field of routine medical care. There is typically no statistical consulting when it comes to interpreting diagnostic tests or how to communicate risks and probabilities to patients. It is hence even more important to incorporate these practical challenges in patient-doctor communications in the curricular statistical teaching of medical students.

REFERENCES

Altman, D. G., & Bland, J. M. (1991). Improving doctors' understanding of statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *154*(2), 223–267. https://doi.org/10.2307/2983040

Casscells, W., Schoenberger, A., & Grayboys T. B. (1978). Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine*, *299*(18), 999–1001. https://doi.org/10.1056/nejm197811022991808

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*(1), 25–47.

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology, 28*(2), 210–216. https://doi.org/10.1037/a0014474

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *BMJ*, *327*(7417), 741–744. https://doi.org/10.1136/bmj.327.7417.741

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*(5500), 2261–2262. https://doi.org/10.1126/science.290.5500.2261

LeBlanc, M., Rueegg, C. S., Bekiroğlu, N., Esterhuizen, T. M., Fagerland, M. W., Falk, R S., Frøslie, K. F., Graf, E., Heinze, G., Held, U., Holst, R., Lange, T., Mazumdar, M., Myrberg, I. H., Posch, M., Sergeant, J. C., Vach, W., Vance, E. A., Weedon-Fekjær, H., & Zucknick, M. (2022). Statistical advising: Professional development opportunities for the biostatistician. *Statistics in Medicine*, *41*(5), 847–859. https://doi.org/10.1002/sim.9290

Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., Vickers, A. J., Ransohoff, D. F., & Collins, G. S. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, *162*(1), W1 –W73. https://doi.org/10.7326/M14-0698

Wald, N. J., & Bestwick, J. P. (2014). Is the area under an ROC curve a valid measure of the performance of a screening or diagnostic test? *Journal of Medical Screening*, *21*(1), 51–56. https://doi.org/10.1177/0969141313517497