RISK IN HEALTH: MORE INFORMATION AND MORE UNCERTAINTY

BOROVCNIK Manfred
Institute of Statistics, University of Klagenfurt, Austria
Manfred.Borovcnik@uni-klu.ac.au

KAPADIA Ramesh
Institute of Education, University of London, UK

ABSTRACT
Risk is a major factor in health, with a strong focus on minimising risk wherever possible. The mathematical starting point is probability. Reliable or relevant data is often missing or hard to get. Moreover, the results of studies are all too easily interpreted wrongly – even by medical experts. Usually it is seen as useful to have more information in making decisions. As we show below this is not always true; we will use the exemplar of breast cancer and screening as an illustration throughout this paper to explain circumstances where there is 'more information and more uncertainty' following Knight and type 2 errors. We identify the different stakeholders and parts of their internal criteria that form their 'rationality', which may well be idiosyncratic. The intention is to pave the way for a 'shared' decision which is best for individuals and for society simultaneously. Statistics educators will find an important field of research and teaching.

INTRODUCTION

The focus of this paper is on health and the fact that, not least with the increasing access to information via the Internet, there are advantages as well as dangers. We begin with some general comments before assessing the quality of information available. We analyse the different types of risk and we note the different interest groups involved and that the view of risk depends on one's role in the situation.

The key example we will use in this paper is mammography screening for breast cancer. Many (but not all) patients want to know about factors affecting their health. In the debate on informed consent, there is a growing tendency by doctors to explain *all* potential side effects of a suggested treatment. Shared decisions *and* responsibilities should assist those involved to be effective; it does not always work that way. Indeed it may lead to a decrease in accountability of professionals.

Well-informed patients can certainly improve the communication between medical doctors and themselves. For example, a well-informed patient can disclose and discuss wider issues relating to the health problem being discussed. But poor communication may sometimes increase the risk of incorrect treatment. More information may also reduce the mutual trust between doctor and patient: as Voltaire said, "medicine is what we do while nature takes its course" (the placebo effect), so trust is crucial. Hence there can be dangers in being too open, especially in discussing (Knightian) uncertainty, based on lack of knowledge rather than risk based on sound information. It is not always easy to separate the two aspects.

Sound information in health is usually based on statistical methods and the decisions rely on probability models. The methods involved are complex and mathematical concepts are rarely well understood, especially when conditional probabilities are involved, as shown by recent research by Huerta (2009). He has classified the complexity of such problems and shown that students at school and university are not able to apply ideas learnt in a new context.

Gigerenzer (2002) has developed methods, which are directly comprehensible. One example for his approach is his focus on natural frequencies, to replace the corresponding conditional probabilities.

However, Gigerenzer's approach is bound to the standard model, which presupposes *probabilities* of the diagnosing procedure to err in both cases of a woman with and without cancer of the breast and a probability (prevalence) of this cancer among women of such an age, despite strong arguments against such an averaging view (is the subgroup to which a patient is attributed, really relevant?). It also (wrongly) assumes that the mathematical result is equally relevant for all stakeholders, including the individual woman tested, the doctor who investigates her, the hospital

that has installed the equipment, and the health institutions that have promoted the screening scheme. Our analysis below shows that the stakeholders differ in their role so much that no common rationality is left as a unifying element. Furthermore, in their decisions people use ideas beyond what is usually perceived as rational. Notwithstanding that, pseudo-statistical figures or unreliable statistics are used in the discourse. Thus, to support people to understand the statistics is only one step in the right direction. A holistic view of the situation is required to improve the management of risk. In what follows a systems analysis approach is chosen to extend the perspective of a decision that finally has to be shared between unequal stakeholders.

THE QUALITY OF INFORMATION

People have developed a deeply causal image of life (see Kapadia, 2010). If anything is wrong we simply have to find a suitable remedy. The health sector, too, feels a strong pressure, as do doctors to find immediate cures. In health, emotions and desires also play an important role. There can be various differing scenarios of which two are noted below.

1. There are severe symptoms requiring action. The need is to heal directly such as with a broken leg, or relief of severe pain. The expectation is that the treatment helps immediately. Effectiveness of the treatment is relatively easy to judge using related measurements linked to the symptoms in the short or long term.
2. There are indications that dangers await in the future. Actions are taken which either safeguard the patient or can be fear-mongering. An evaluation of effectiveness is hypothetical as the possible disease (such as cancer) may be benign.

What information is relevant for judging the situation? To get valid information has always been difficult; the change with the Internet is an explosion in the availability of information but no new ways of judging validity (Miller, 2004). More important, good information is often hard to understand and may involve mathematical concepts and statements on basis of models or scenarios, which are bound to (hidden) restrictions. The validity of information can be judged by its source and the role of the stakeholder in the 'game', or the specific health situation. In any case, information has to be interpreted and judged for quality. Where does the data come from? Does the information depend on restricted models? Is the information merely opinions or anecdotal? Ways to judge information are multi-faceted and subjective (even inside science) and may depend on minor issues like whether an opinion leader, a prominent actor, an attractive woman, or an 'expert' transmitted the information. With regard to the role of the stakeholder there are huge differences in their criteria:

- Doctors: A doctor has to follow the state of the art and avoid risk. If (s)he decides for a scheme deviating from treatment lege artis, (s)he is liable for adverse consequences.
- Experts: A scientist seeking funds has to make the case important, perhaps exaggerating potential negative consequences in the future. Such a stakeholder bears no risks.
- The pharmacology industry: Such enterprises have an economic interest in promoting specific treatments or schemes of diagnosis. Within certain limits, they have a liability and may be sued for compensation (Khamsi, 2005), but they also need to make a profit.
- The media and politicians: Media are interested in sensations and hot air. Their speculations on future (adverse) developments bear no risk or liability. Politicians have to be seen to be good at making decisions, and bear the risk of their personal career, which might end immediately if their decisions judged (even incorrectly) by the media to be wrong.
- The patient: (S)he wants to improve his (her) health and has to bear short and long-term risks and consequences and may not base decisions only on rationality.

An example of a conflicting view is measles and how the risk of infection is reduced by vaccination. There is a clash between an individual and a societal perspective. Information from studies might induce an individual to avoid the jab (because of its potential side effects); on the other side, for society – once the information from the studies is judged to be (at least slightly) in favour of the vaccination – there is a strong interest that all individuals take the vaccination to prevent an epidemic. Moreover, an individual's decision against vaccination is only successful if enough others have already taken the vaccination ('free-rider effect', see May & Silverman, 2005). Thus different criteria form the basis of rationality, for the research community in

medicine, for the health sector (hospitals and doctors), for the pharmaceutical industry, and, by no means least, for the patients; all have differing expectations, risks and responsibilities.

TYPES OF RISK

By risk we understand a global view on a situation with inherent uncertainty about the (future) outcomes, which are related to impact (cost, damage, or benefit). Risk is used heterogeneously – there is an overlap with everyday language, which blurs the issues. Some use 'risk' to designate the probability inherent to one adverse outcome without regarding its impact and the other outcomes; others designate by risk only the adverse outcome, a further group uses risk to denote the adverse outcome including its probability (but do not refer to the whole situation).

Knight (1921) tried to establish a distinction between risk and uncertainty in the economic sphere. According to him uncertainty cannot be quantified, while risk is measurable. Tversky & Kahneman (1979) noted in their work on Prospect Theory that risk is sometimes used when there can also be a benefit involved. For much human endeavour has some element of risk, yet is undertaken because of the likelihood of success and benefits. Adams & Thompson (2002) discriminate three types of risk: "directly perceptible, perceived with the help of science, and virtual – risks about which scientists disagree or confess ignorance." They also relate the judgment of such risks to perceptual filters, which can be "individualist, egalitarian, fatalist, and hierarchist". Dealing with directly perceivable risks is mainly intuitive; it is not easy to make such risks the topic of an explicit or shared approach as people have already internalized their perception and decisions.

Expected value based on information about probabilities is used to select from several decisions, which are 'at stake'. A decision between several choices of action might involve an individual or be 'shared' such as by a patient and a doctor who must decide the next steps. A decision may involve stakeholders who never 'meet' such as a health institution implementing mammography screening and an individual.

Risk involves probability and impact, components which are both prone to subjective features. Crucially, the judgment of impact varies between different individuals and health providers. To live with an artificial hip is heavily dependent on a person's lifestyle; a doctor or an institution recommending an operation may simply envisage the possibility of a failure from an 'average' point of view. The other component of risk, its inherent probabilities for the various possibilities involves an equalizing argument over all patients, which has a meaning only from a non-personal approach (the doctor, the institution). More crucially, the calculation of such a probability is strongly influenced by the patient as well as by the role and the experience of the stakeholder; this may also differ for institutional representatives.

We return to the two different scenarios noted above. The first scenario involves a patient with severe symptoms like pain or a mobility constraint of the hip joint. The options are physical therapies with the aim of reducing pain or a replacement of the joint by operation with the prospect of full mobility after a rehabilitation phase. Statistical evidence from past operations might indicate the relative success of the second option. In the early phase of hip operations the risk of an operation was only accepted (by doctors and patients) in case of a definite damage of the patient's hip. The second scenario relates to virtual risks in health issues, which comprise situations where the individuals have no restrictions and feel in good health. However, the *future* might bear risks. The question is whether there are precautionary measurements to prevent them.

Problems with understanding the situation, the underlying concepts, and the interpretation of results, increase with the case of a risk with a very high impact combined with a negligibly small probability. It may well be that avoiding a very high impact event (such as cancer) is worth the extra stress caused by screening, even when this may lead to a 'false positive' result where the chance of cancer after a positive test is still quite low. The question for a health service is how to select the people and at what frequency they should be screened. These are complicated ideas and are also influenced by external considerations such as speculation by politicians and the media.

STATISTICAL INFORMATION IN HEALTH ISSUES

Various factors may increase the risk for later occurrence of a disease. For medicine, the search for causality is not straightforward and causal factors vary across the population. Relative to biochemistry processes, hypotheses might be established about interrelations. For smoking we now have some knowledge about high-level exposure to ionising radiation on the growth of tumours and could extrapolate this pattern to low doses and come to the conclusion that smoking establishes a (causal) risk factor.

General patterns are derived from observations ex post: those who suffer from a disease and those who do not have this disease are compared whether they have been exposed differently to potential risk factors in the past. People with lung cancer show higher rates of smoking than people who do not have this disease. This confirms that smoking may well be a risk factor but does not identify causal relations, which leaves the information in an ambiguous state and quite unconvincing to some (such as manufacturers of tobacco). Such analyses could be corroborated by well-designed experiments, but it is an ethical infringement to assign people to smoke and not to smoke, and the data would be available only after a long period of time. Researchers seek surrogate endpoints. This might be deformation of lung cells or tar concretion. It is obvious that some of the surrogate endpoints convey little information about the endpoint under scrutiny. In other studies the all too loose connections of surrogate endpoints to the true endpoints might be less obvious.

We return to consider the benefit of mammography screening. For women between ages 50 and 69, the benefit of screening is evaluated by Kalager et al (2010) as net 2.4 deaths per 100,000 person-years. They compare historical and current groups together. For the current group they show 21.2 deaths for the non screening group and 18.1 for the screening group, which amounts to a reduction of 3.1. As their individual remaining life time is unclear, and it remains unclear how much longer they would have survived otherwise (without breast cancer) the net gain in life time is left to "speculation". Referring to the older meta-analysis of Nyström et al (1996), Gigerenzer (2002) reports a mean life time gained by joining the screening programme of 12 days for women of the same age group (50-69). The number of deaths for 1,000 women in the screening programme for 10 years is given as 3, which is 1 less than for those who do not follow the screening programme.

Crucial questions for an informed decision are:

i. Does screening reduce the incidence of cancer of the mammae? No, early detection is different from prevention.

ii. Are all cancers of the mammae progressive? No, especially with younger women there are carcinoma in situ (sleeping cancer) which might have done no harm for a long period.

iii. Is early detection always an advantage? No, as if the carcinoma is weakly progressive, an advantage is questionable.

iv. Is the whole process of diagnosis doing harm to the patients? The exposure to ionizing radiation by the mammogram itself may cause cancer. The radiation may affect other aspects of health. It can not be excluded that the biopsy following a positive mammogram may change the character of a carcinoma in situ to become invasive. This question is not open to empirical investigation, however. Concerning the adverse impact of the screening programme on the quality of life, van der Steeg et al (2011) report high anxiety for at least one year for women with a false positive mammogram.

The incidence of false positives is high. The incidence for breast cancer varies greatly with age so that with young women the rate of false positives is about 90% (for age 40-49) with the usual quality of diagnosing from mammograms (Gigerenzer, 2002). Elmore et al. (2005) analyze data of the Harvard Pilgrim Health Care (a health prevention organization that cares for 300,000 adults) with the result that in 10 ensuing mammograms the risk for at least one false positive diagnosis is close to 50%. For Germany, Koubenec (2000) argues that from 3 to 4 million screening mammographies, roughly 300,000 false positives lead to about 100,000 biopsies, showing the dimension of the problem.

A transparent balancing between benefits and potential harm is needed. The medical doctors and representatives of the health authorities focus on the benefits from prevention of deaths (the one) they claim to prevent in advocating mammography screening. The evaluation of

potential harm is very complicated and indirect. For the group of 40-49, the meta analysis by Gøtzsche & Nielsen (2011) did not find a positive screening effect. On the other hand, for older women, the false positives are lower but the gain in remaining life time is also smaller.

False-negative diagnoses occur in approximately 20% to 40% of women with breast cancer (Elmore et al., 2005). A further relevant point is the alternative methods of detecting cancer. Wainer (2011) gives a description of the huge economic costs of screening programmes. For the screening programme for prostate cancer recent evaluations (Sandblom et al., 2010) are pessimistic: "After 20 years of follow-up the rate of death from prostate cancer did not differ significantly between men in the screening group and those in the control group."

RISK MANAGEMENT IN HEALTH SYSTEMS

How can we develop schemes to arrive at a 'shared and mutually beneficial decision'? The focus has to be on a broader view of risk, which includes the following features:

- Risk is hard to estimate for different persons, and even harder to share between individuals and institutions.
- Risk management to improve the situation requires negotiation between patients, doctors, and the public health system.
- The perception of risks also differs between individuals and institutional stakeholders.
- Criteria are often unconscious and possibly blurred by psychological factors such as fear, or hope, or attitudes and bias.
- Perception is blurred by impact, idiosyncratic experience, indirect feedback, formal knowledge, or abstract and counter-intuitive relations.

Even medical experts lack a sound comprehension of the concepts. This results in an overestimation of the positive impact of treatment or diagnosing schemes. People have to be supported not only in the mathematical concepts involved but also in strategies of how to communicate about inherent risks.

If there is rationality the Western approach is that this should be unique, which differs from Eastern Philosophy where there can be several 'truths' which may even conflict. For applied mathematics, the vision of rationality is to strive for a unique and correct model. The aim is that better models are simply more refined, embedding the cruder model as a special case. However, it may well be that in this situation, two or more different (*equally valid*) models in parallel fit the needs of the stakeholders involved better.

The aims of different parties vary hugely: make profits; make a better living; increase one's resources; increase the resources of society; avoid liability; circumvent decisions; improve one's health status. To learn more about criteria and how people of a different position in the system act and how their actions may be reconciled, empirical investigations are needed:

- Structured interviews with doctors and members of the health system.
- In-depth observation of an interacting group with a health decision problem with doctors and patients (to represent their different roles).

An innovative approach is to find out the inherent criteria of the different stakeholders and how they can be reconciled to improve risk management for all. Risk management has to include all stakeholders – the approach has to be better than a meta-model with one final view: the final decision might mirror an 'informed decision' of the individual, which is respected and supported by the system, and which improves the system.

CONCLUSIONS

A decision problem (in health issues) is thought to comprise a unique situation where a model is fitted. The criteria for making the decision often appear to be presented as clear-cut and sharp. The only problem remaining seems to be to understand the procedure of information processing and the concepts involved like probabilities (especially very small probabilities), conditional probabilities (hypothetical to some conditioning statement, which might hold or not), or the Bayesian formula. To embed information in easily accessible forms helps to get 'correct' results. Yet, in practice, the task to interpret the results accordingly remains challenging.

Firstly, is the prior probability of suffering from a disease in the whole population appropriate for the person in question or not? In the case of mammography the conventional

model leads to new and unforeseen problems, not least due to the base rate, a fallacy already considered by Tversky & Kahneman (1979). Secondly, the input information is far from reliable as it includes (random) sampling variation. Results are different when applied to different single cases, with the severe concern as to whether an 'averaging' on all might be applicable to any of these single cases (is the prior probability chosen relevant also for *this* person?). Thirdly, the quality of information beyond sampling variation is at stake. Flaws of studies differ and lead to different – even contradictory – information (see Goldacre, 2005), other studies are relatively critical about the success claims of screening programmes (Kalager et al, 2010, Elmore et al, 2005).

More information in this case therefore almost inevitably leads to controversial information and the inherent problem of judging its quality. To assist the various stakeholders, a prerequisite is to enable them to understand the underlying concepts. A systems analytic approach will lead beyond and help to 'negotiate' the information available or required. A final decision will not really be rational if the role of people involved and their varied interests are not sufficiently appreciated. The process of empowering stakeholders in a shared decision requires more systematic and intensive study on the stakeholders' internal criteria, beyond the underlying mathematical concepts. It is for statistical educators to rise to this challenge.

REFERENCES

Adams, J., & Thompson, M. (2002). *Taking account of societal concerns about risk. Framing the problem*. London: Health and Safety Executive.

Elmore, J.G., Armstrong, K., Lehman, C.D., & Fletcher, S.W. (2005). Screening for Breast Cancer. *Journal of the American Medical Association* (JAMA) 293(10).1245-1256.

Gigerenzer, G. (2002). *Calculated Risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Goldacre, B. (2008). *Badscience*. London: Fourth Estate.

Gøtzsche, P.C., Nielsen, M. (2011). Screening for breast cancer with mammography – Review. *The Cochrane Library*, 4. doi: 10.1002/14651858.CD001877.pub4.

Huerta, P. (2009). On conditional probability problem solving research – structures and contexts. *International Electronic Journal of Mathematics Education* 4(3), 163-194.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, XLVII, 263-291.

Kapadia, R. (2010). Chance and necessity: the languages of probability and mathematics. In C. Reading (Ed.), *Proceedings of the Eight Intern. Conference on Teaching Statistics (ICOTS 8)*. Online: www.stat.auckland.ac.nz/~iase/publications.php?show=icots8.

Kalager, M., Zelen, M., Langmark, F., & Adami, H.O. (2010). Effect of Screening Mammography on Breast-Cancer Mortality in Norway. *New England J. of Medicine* 363,1203-10.

Khamsi, R. (2005). Painkiller verdict shows mistrust of Merck. *Nature* 436, 1070. Online: www.nature.com/nature/journal/v436/n7054/full/4361070a.html

Koubenec, H.J. (2000). Mammographie-Screening: Überschätzen wir den Nutzen? *Berliner Ärzte* 37(8). 11–16.

May, T., & Silverman, R.D. (2005). Free-riding, fairness and the rights of minority groups in exemption from mandatory childhood vaccination. *Human Vaccines* 1(1), 12-15.

Miller, E. (2004). The Great Wikipedia Authority Debate. *The Importance of ...* . Online: http://importance.corante.com/archives/2004/08/30/the_great_wikipedia_authority_debate.php

Sandblom, G., Varenhorst, E., Rosell, J., Löffman, O., & Carlsson, P. (2010). Randomised prostate cancer screening trial: 20 year follow-up. *British Medical Journal* 342. d1539. doi: 10.1136/bmj.d1539.

van der Steeg, A.F.W., Keyzer-Dekker, C.M.G., De Vries, J., & Roukema, J.A. (2011). Effect of abnormal screening mammogram on quality of life. *British Journ. of Surgery* 98, 542–543.

Wainer, H. (2011). How should we screen for breast cancer? Using evidence to make medical decisions. *Significance* 8(1), 28-30.