TEACHING STATISTICAL MODELS FOR RISK ASSESSMENT

Mehdi Razzaghi
Bloomsburg University, USA
mrazzagh@bloomu.edu

Risk analysis is used across many disciplines. Biological scientists use risk analysis to determine the chance of an adverse effect occurring. Engineers use it to assess the risk of structural failures. People in business and industry perform risk analysis to estimate profit and loss in decision making. Although there are many definitions of risk, the most accurate definition is based on probabilistic principles. To estimate risk quantitatively, a statistical model is used to describe the process and the associated risk is thereupon calculated. The underlying model and derivation of risk can lead to some non-trivial mathematical expressions. Here, we discuss challenges in teaching probabilistic risk assessment to researchers who are non-statisticians and non-mathematicians. Specific reference will be made to toxicological risk assessment.

INTRODUCTION

The concept of risk is a phenomenon that has been used in many diverse settings. Although risk has varying definitions in different disciplines, the common theme is that risk refers to an adverse situation that involves the possibility of loss or harm. In fact, the Webster Dictionary (www.merriam-webster.com) defines risk as "possibility of loss or injury" whereas the British Dictionary (www.dictionary.com) defines it as "the possibility of incurring misfortune or loss; hazard." Risk analysis refers to the process of identifying and assessing risk factors with the goal of controlling and managing those risk factors. Biological scientists use risk analysis to determine the chance of an adverse effect occurring such as incidence of a disease or an environmental mishap. Engineers use it to assess the risk of structural failures and damages. In business and finance, risk analysis is performed to estimate profit and loss in decision making. Risk can most accurately be defined based on probabilistic arguments, and thus reliable risk analysis must rely on probabilistic models. Generally, in the scientific quantitative risk assessment procedure, an appropriate statistical model is used to describe the underlying process, and risk is calculated as the probability of an adverse event. Here, we discuss challenges in teaching probabilistic risk assessment to researchers and academicians who are not statisticians or mathematicians. For demonstration purposes, our focus of attention will be specifically on risk assessment processes in toxicology, where statistical models have long been used for the purpose of quantitative risk assessment. Since the early 1930's when Bliss (1934) used probit regression for dose-response modeling, to the 1980's when a myriad of publications on application of various models were published (see e.g., McCullagh & Nelder, 1989), to the more modern era, statistical models have been instrumental in toxicological risk assessment. For a more thorough discussion, we refer to Razzaghi (2020). In the next section, we provide the mathematical definition of risk and describe the process of risk assessment in toxicological experiments. To alleviate the challenges of quantitative risk assessment in toxicology, the United States Environmental Protection Agency (EPA, 2017) developed the Benchmark Dose software (BMDS) to assist practitioners in estimating risk. We discuss the properties of this software and provide an example for illustration.

RISK AND RISK ASSESSMENT

To give a mathematical framework to the process of risk assessment, it is first necessary to provide a mathematical definition of risk. Because of the diversity of risk applications, unfortunately there is no universal agreement and there is not a unique mathematical definition of risk. For example, in business and finance, risk is usually attributed to the possibility of monetary loss and the measure of risk is sometimes expressed as the standard deviation of the portfolio return or the loss distribution. In fact, the use of variance and standard deviation as a measure of risk in business and investment has a long history and dates to Houston (1964). The idea was more formally expressed in mathematical terms by Hickman and Zahn (1966). In biological experiments, specifically in case-control studies, the relative risk, defined as the ratio of the probabilities of an outcome in the exposed group to the probability of an outcome in the unexposed group, is used as a measure of risk. It is worth noting that

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

often the risk ratio is confused with the odds ratio. In fact, using clinical data from obstetrics and gynecology, Holcomb et al. (2001) warn that using the odds ratio can differ substantially from the risk ratio, and using the odds ratio as a measure of risk is a misrepresentation of research results because the two measures are not identical. In engineering, risk is sometimes measured as the probability of failure of a structure (Rockafellar & Royset, 2015).

In toxicological experiments, risk is generally defined as the measure that an adverse effect will occur under defined conditions of exposure to a chemical. More formally, let $\pi(d)$ denote the risk function, i.e., the probability that an adverse effect is observed as the result of exposure to a concentration level d of a toxic substance. Then, there are commonly two mathematical definitions of the measure of risk. One is Additional Risk, which is the difference in the probability of response in the exposed and unexposed subjects. Thus, if $P(d)$ is the probability of response at the concentration d, then

$$\pi_A(d) = P(d) - P(0) \tag{1}$$

The other definition of risk that is often used is Extra Risk, also referred to as the Relative Risk, which is the ratio of the additional risk to the probability of observing an effect in non-exposed individuals. Thus, it is the proportional increase in the risk adjusted for the background.

$$\pi_R(d) = \frac{P(d)-P(0)}{1-P(0)} \tag{2}$$

Note that both measures are monotonic functions of d and that

$$\lim_{d\to\infty} \pi_R(d) = 1 \tag{3}$$

The process of risk assessment in toxicology begins by first defining a dose-response function, which is a mathematical expression of the quantitative relationship that exists between the probability of response and the exposure level. Often a sigmoid-shaped function is used to describe this relationship. For quantal responses, the cumulative distribution function is generally selected for this purpose. Common dose-response functions for quantal responses include the exponential and logistic distributions. For continuous responses, the mean response is expressed as a mathematical function. Common models are polynomial and power functions. For a more comprehensive list we refer to Razzaghi (2020). A principal goal of establishing the dose-response relationship is to determine an exposure level of the toxin that may be used as the so-called Point of Departure (POD) or the starting point for calculating an acceptable exposure level for the human population. It is defined as the point on the dose-response curve corresponding to a low, fixed, nominal effect level. The modern approach for determining the POD relies on the Benchmark-Dose (BMD) methodology. The methodology was first introduced by Crump (1984) as an alternative to the old-fashioned approach of using the No-Observed-Adverse-Effect-Level (NOAEL), defined as the highest experimental dosage level that produces no statistically or biologically significant adverse effect. The NOAEL approach has been criticized by many authors and shown to be ill-defined, unreliable, subjective, and dependent on dose spacing. (See, for example, Leisenring & Ryan, 1992 and Gaylor, 1994). After the dose-response relationship is established, the BMD is calculated as the dose that causes a fixed preset change (usually 5% or 10%) in response, called the benchmark response (BMR). The lower confidence bound (usually 95%) called the benchmark dose limit (BMDL) is then used as the POD. Although the BMD methodology was introduced primarily to replace the NOAEL approach for finding the POD in non-cancer risk assessment, because of its many interesting properties, it was soon suggested by several authors, including Gaylor et al. (1999), that the methodology can effectively be used in cancer risk assessment as well. For a review of the BMD methodology, see Haber et al. (2018).

TEACHING QUANTITATIVE RISK ASSESSMENT

The process of toxicological risk assessment was developed by the U.S. Environmental Protection Agency through several reports (1991, 2012) to address the health risks to humans, animals, and other organisms because of exposure to a range of hazards, including chemicals, food additives, industrial waste, and so on. As described above, the process of risk calculation and quantitative risk assessment in toxicology involves several steps, from selection of a dose-response model to fitting the function to data and from calculating the risk to deriving a lower bound for the benchmark dose. Some of these steps can be quite mathematically challenging. The difficulty is that, in general, the practitioner and users of statistics in research are often not interested in the theory and the mathematical details of the risk assessment methodology. Therefore, the educator is faced with the challenge of teaching the material, trying to incorporate as much concept as possible while also trying to avoid the theory. The practitioner is interested in learning how to do the risk assessment and not so much in the theory behind the methodology. Thus, finding the right balance between theory and practice is the challenge that the statistician must resolve. Of course, the problem of teaching statistics to non-statisticians is not new and the challenge is present at all levels from elementary concepts to more advanced research topics. Several authors have addressed this issue and discussed various methods for overcoming these challenges. The American Statistical Association (ASA) has published several reports on priorities in statistics teaching. Most notably, the Guidelines for Assessment and Instruction in Statistics Education (GAISE) was first published by Aliaga et al. (2005) and later revised (GAISE College Report ASA Revision Committee, 2016). In the GAISE report, six recommendations are given for teaching statistics at all levels, focusing on what to teach and how to teach it. Mustafa (1996) argues that "to be more effective, it is essential that teaching objectives are clearly defined at the outset and issues of content and methodology are addressed accordingly" (abstract). Perhaps the most comprehensive research on challenges of teaching statistics to non-specialists is displayed in a more recent article by Bromage et al. (2021). In that article, the authors assert that "many of the key challenges stem from negative attitudes towards statistics coupled with poor motivation" (p. 46). Accordingly, the essential ingredients of the modern statistics course should concentrate on development of statistical literacy, and a focus on understanding of statistical concepts, rather than on the specifics of mathematical computations. The authors conclude that one of the most efficient ways to resolve the issue is to make use of the available technology. The emergence of high-level electronic calculators and development of a wide range of software programs have encouraged educators to utilize some of the many specially developed software programs to teach high level and sophisticated mathematical and statistical procedures, avoiding the deep theory and covering concepts at a rudimentary level. To name a few, we can mention commercially available software such as SAS, MATLAB, and SPSS or some of the open-source programs such as R/RStudio, JASP, and SOFA.

In risk assessment, many quantitative methods have been built into the software tools mentioned above to facilitate the process of risk assessment. Through the application of these tools, an educator can often find a balance between the necessary theoretical background and analysis and interpretation of results. For toxicological risk assessment, although there are some commercially developed software programs available (e.g., @RISK) that provide ample tools for probabilistic risk assessment purposes, the most comprehensive tool currently available is probably the BMDS software developed by the EPA. With the growth of popularity and importance of the BMD methodology in risk assessment, the EPA developed the benchmark dose software to assist practitioners in the process of model fitting and calculation of the reference dose (RfD), and hence determination of the POD. The software development has gone through many updates and the most current version, BMDS (EPA, 2017), contains several modifications and is used by thousands of users and risk assessors worldwide. The software is very versatile and available for download on the EPA web site (https://www.epa.gov/bmds/download-bmds). BMDS is a Microsoft Excel–based program that is highly user friendly and simple to use. The software can be used for cancer quantal data with single or multiple tumors as well as continuous outcomes. In addition, there is an option for nested data from developmental toxicity experiments. Several choices of dose-response models for each type of data are available with the choice of additional or extra risk as defined in (1) and (2) as the measures of risk. To alleviate the burden of choosing a single dose-response model, the software also provides the option of using the Bayesian Model Averaging (BMA) technique with user-provided weights. For more on BMA technique and its application in statistical modeling, see Fletcher (2018). The BMDS

output provides information about the model parameter estimates, goodness-of-fit, BMD, and BMDL for each model. For quantitative responses, options are available for choosing the adverse effect based on the number of standard deviations or the hybrid model. For more information about the software, refer to the EPA web site https://www.epa.gov/bmds. In the next section we provide a demonstrative example.

EXAMPLE

In this section, we use the results of a cancer bioassay study to demonstrate the application of the BMDS software.

Table 1. Incidence of Tetis in rats exposed to TCE

Dose (ppm)	Number Alive	Number Affected	Proportion Affected
0	121	6	0.05
100	116	16	0.138
300	116	30	0.259
600	122	31	0.254

Source: Maltoni et al. (1986)

Because the outcome of cancer bioassay studies is dichotomous in nature, we use the option for dichotomous models. Table 1 gives the data from Maltoni et al. (1986) for incidence of the Tetis, Leydig cell tumor in rats exposed to trichloroethylene (TCE). The chemical, which is a nonflammable and colorless liquid, is used in adhesives and paint removers and is known to be carcinogenic to humans. Using Extra Risk as the measure of risk and 0.1 for the value of benchmark risk, the BMDS software was applied for risk assessment. Table 2 provides the summary risk assessment information when six different dose-response models are applied. More information about each of the models can be found in Razzaghi (2020). The BMDS software produces several interesting and useful results that may be utilized for further analysis along with a recommended model. However, here we give the values of BMD and BMDL for each model along with the *p*-value of the test of significance for each model and the value of the Akaike Information Criterion (AIC). This is an estimate of the predicted error and the relative quality of the model. The AIC is frequently used in model selection to choose from among several prospective models. In general, a model with the lowest AIC is preferred. From Table 2, other than the Hill model, the other five models are significant at the10% significance level. However, more importantly, judging by the AIC values, the Hill model has the lowest AIC. In fact, this is the model recommended by the BMDS software. For this particular model, Figure 1, also extracted from BMDS, depicts the estimated dose-response curve as well as the values of BMD and BMDL on the graph.

Table 2. Risk Assessment Summary

Model	BMD	BMDL	P-Value	AIC
Hill	101.29	39.33	0.936	417.74
Gamma	214.42	156.88	0.057	421.42
Log-Logistic	188.97	133.65	0.095	420.41
Multistage	214.42	156.88	0.057	421.42
Weibull	214.42	156.88	0.057	421.42
Logistic	338.50	275.87	0.007	425.87

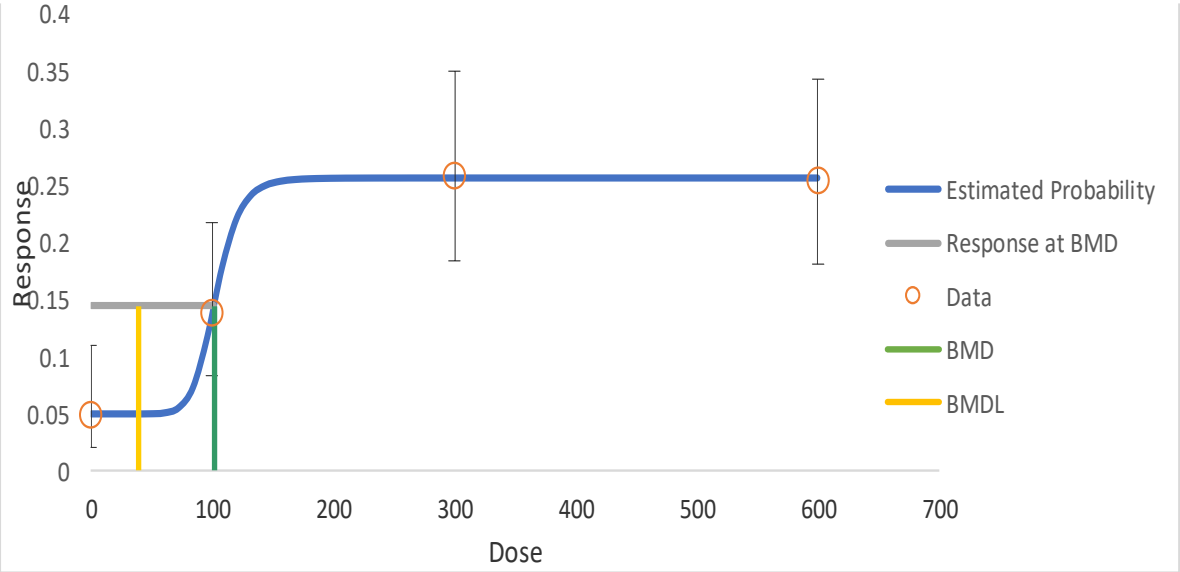

Figure 1. Frequentist dichotomous Hill model with BMR of 10% extra risk for the BMD and 0.95 lower confidence limit for the BMDL

CONCLUSION

Teaching probabilistic risk and risk analysis to non-statisticians is a challenging task because of the mathematical complexity of the underlying methodology. Utilizing technology and an appropriate software program can, to a large extent, alleviate the problem. Here, we have shown that in toxicology, the process of risk assessment involves some theoretical results that can deter the practitioner. By using publicly available and user-friendly BMDS software, a thorough analysis may be derived in a very simple manner. The software is compatible with a variety of data types and a wide range of models.

REFERENCES

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE): College report.* American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/2005gaisecollege_full.pdf

Bliss, C. I. (1934). The method of probits. *Science*, *79*(2037), 38–39. https://doi.org/10.1126/science.79.2037.38

Bromage, A., Pierce, S., Reader, T., & Compton, L. (2021). Teaching statistics to non-specialists: Challenges and strategies for success. *Journal of Further and Higher Education, 46*(1), 46–61. https://doi.org/10.1080/0309877X.2021.1879744

Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundamental of Applied Toxicology, 4*(5), 854–871. https://doi.org/10.1016/0272-0590(84)90107-6

Fletcher, D. (2018). *Model averaging*. Springer.

GAISE College Report ASA Revision Committee. (2016). *Guidelines for assessment and instruction in statistics education (GAISE) college report 2016.* American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf

Gaylor, D. (1994). Dose response modeling. In C. A. Kimmel & J. Buelke-Sam (Eds.), *Developmental toxicology* (2nd ed., pp. 363–375). Raven Press.

Gaylor, D. W., Kodell, R. L., Chen, J. J., & Krewski, D. (1999). A unified approach to risk assessment for cancer and noncancer endpoints based on benchmark doses and uncertainty/safety factors. *Regulatory Toxicology and Pharmacology*, *29*(2), 151–157. https://doi.org/10.1006/rtph.1998.1279

Haber, L. T., Dourson, M. L., Allen, B. C., Herzberg, R. C., Parker, A., Vincent, M. J., Maier, A., & Boobis, A. R. (2018). Benchmark dose (BMD) modeling: Current practice, issues, and

challenges. *Critical Review in Toxicology*, *48*(5), 387–415. https://doi.org/10.1080/10408444.2018.1430121

Hickman, J. C., & Zahn, D. A. (1966). Some mathematical views of risk. *The Journal of Risk and Insurance*, *33*(3), 437–446. https://doi.org/10.2307/250988

Holcomb, W. L., Jr., Chaiworapongsa, T., Luke, D. A., & Burgdorf, K. D. (2001). An odd measure of risk: Use and misuse of odds ratio. *Obstetrics & Gynecology*, *98*(4), 685–688. https://doi.org/10.1016/s0029-7844(01)01488-0

Houston, D. B. (1964). Risk, uncertainty, and sampling. *The Journal of Risk and Insurance*, *31*(4), 511–538. https://doi.org/10.2307/250806

Leisenring, W. & Ryan, L. (1992). Statistical properties of the NOAEL. *Regulatory Toxicology and Pharmacology*, *15*(Part 2), 161–171. https://doi.org/10.1016/0273-2300(92)90047-D

Maltoni, C., Lefemine, G., & Cotti, G. (1986). Experimental research on trichloroethylene. carcinogenesis. In C. Maltoni & M. A. Mehlman (Eds.), *Archives of research on industrial carcinogenesis* (Vol. 5, pp. 1–393). Princeton Scientific Publishing.

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). Chapman and Hall.

Mustafa, R. Y. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistical Education*, *4*(1). https://doi.org/10.1080/10691898.1996.11910504

Razzaghi, M. (2020). *Statistical models in toxicology*. Chapman and Hall/CRC Press.

Rockafellar, R. T., & Royset, J. O. (2015). Risk measures in engineering design under uncertainty. In T. Haukaas (Ed.), *Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP)*. University of British Columbia. https://doi.org/10.14288/1.0076159

United States Environmental Protection Agency. (1991). *Guidelines for developmental toxicity risk assessment.* (Report No. EPA/600/FR-91/001). https://www.epa.gov/sites/default/files/2014-11/documents/dev_tox.pdf

United States Environmental Protection Agency. (2017). *Benchmark dose software* (Version 3.0) [Computer software]. https://www.epa.gov/bmds

United States Environmental Protection Agency Risk Assessment Forum. (2012). *Benchmark dose technical guidance.* (Report No. EPA/100/R-12/001). Retrieved August 8, 2022, from https://www.epa.gov/sites/default/files/2015-01/documents/benchmark_dose_guidance.pdf