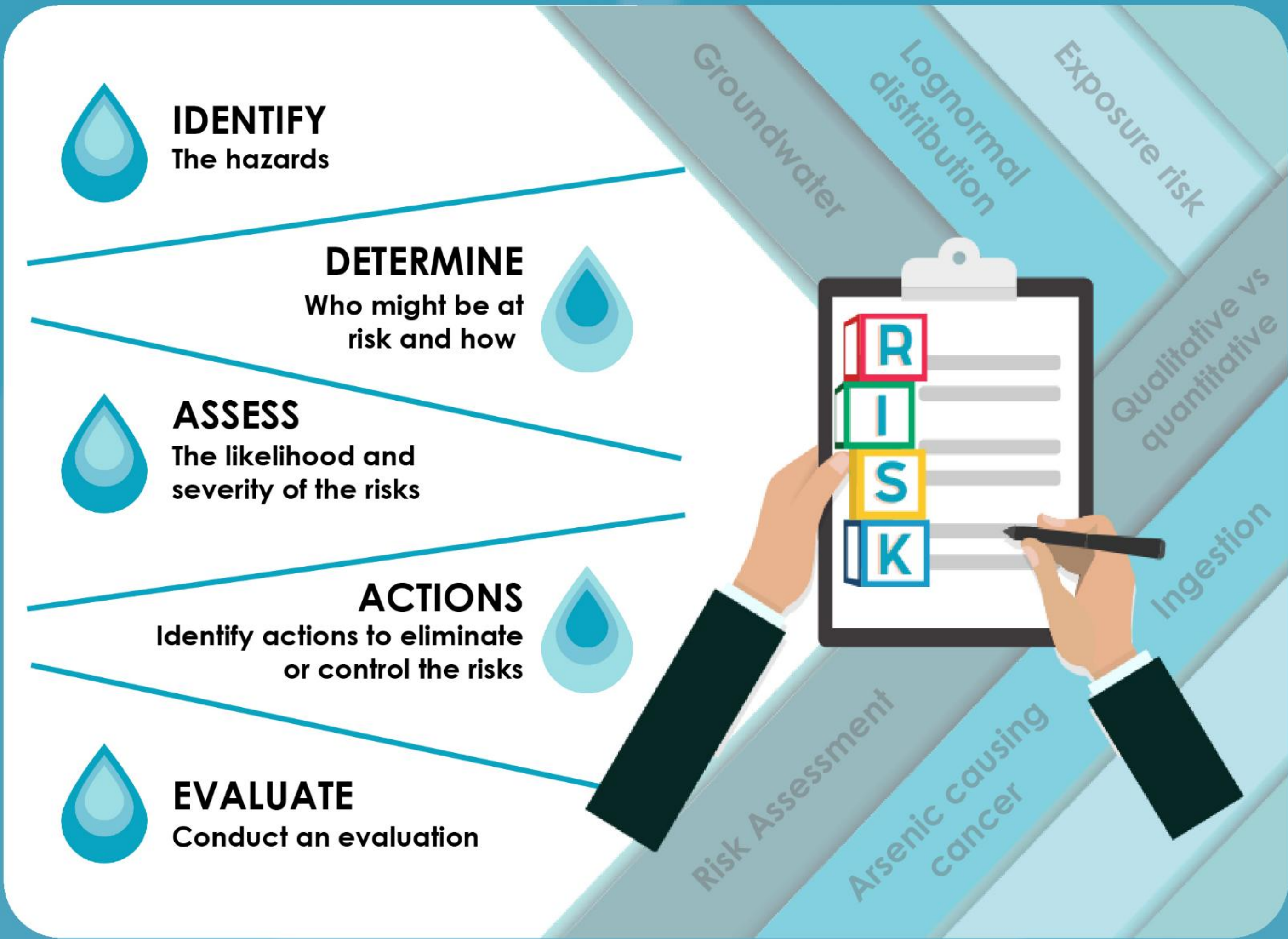


Groundwater Quality and Examples of Risk Interpretation Procedures

Edward McBean

THE GROUNDWATER PROJECT

Groundwater Quality and Examples of Risk Interpretation Procedures

The Groundwater Project

Edward McBean

Professor
School of Engineering
University of Guelph
Guelph, Ontario, Canada

Groundwater Quality and Examples of Risk Interpretation Procedures

The Groundwater Project
Guelph, Ontario, Canada

Published by The Groundwater Project, Guelph, Ontario, Canada, 2023.

McBean, Edward.
Groundwater Quality and Examples of Risk Interpretation Procedures / Author (Edward McBean) - Guelph, Ontario, Canada, 2023
44 pages
ISBN: 978-1-77470-032-7
DOI: https://doi.org/10.21083/978-1-77470-032-7↗.

Please consider signing up for the GW-Project mailing list to stay informed about new book releases, events, and ways to participate in the GW-Project. When you sign up for our email list it helps us build a global groundwater community. Sign up↗.

APA 7th edition Citation: McBean, E. (2023). *Groundwater Quality and Examples of Risk Interpretation Procedures*↗. The Groundwater Project.

THE GROUNDWATER PROJECT

Dedication

Dedicated to Matthew, Derek, and Melissa, as they are the future.

Table of Contents

The Groundwater Project Foreword

At the United Nations (UN) Water Summit held in December 2022, delegates agreed that statements from all major groundwater-related events will be unified in 2023 into one comprehensive groundwater message. This message was released at the UN 2023 Water Conference, a landmark event that brought attention at the highest international level to the importance of groundwater for the future of humanity and ecosystems. This message brought clarity to groundwater issues to advance understanding globally of the challenges faced and actions needed to resolve the world's groundwater problems. Groundwater education is key.

The 2023 World Water Day theme *Accelerating Change* is in sync with the goal of the Groundwater Project (GW-Project). The GW-Project is a registered Canadian charity founded in 2018 and committed to the advancement of groundwater education as a means to accelerate action related to our essential groundwater resources. To this end, we create and disseminate knowledge through a unique approach: the democratization of groundwater knowledge. We act on this principle through our website gw-project.org/↗, a global platform, based on the principle that

"Knowledge should be free, and the best knowledge should be free knowledge." Anonymous

The mission of the GW-Project is to promote groundwater learning across the globe. This is accomplished by providing accessible, engaging, and high-quality educational materials—free-of-charge online and in many languages—to all who want to learn about groundwater. In short, the GW-Project provides essential knowledge and tools needed to develop groundwater sustainably for the future of humanity and ecosystems. This is a new type of global educational endeavor made possible through the contributions of a dedicated international group of volunteer professionals from diverse disciplines. Academics, consultants, and retirees contribute by writing and/or reviewing books aimed at diverse levels of readers from children to high school, undergraduate, and graduate students, or professionals in the groundwater field. More than 1,000 dedicated volunteers from 127 countries and six continents are involved—and participation is growing.

Hundreds of books will be published online over the coming years, first in English and then in other languages. An important tenet of GW-Project books is a strong emphasis on visualization with clear illustrations to stimulate spatial and critical thinking. In future, the publications will also include videos and other dynamic learning tools. Revised editions of the books are published from time to time. Users are invited to propose revisions.

We thank you for being part of the GW-Project Community. We hope to hear from you about your experience with the project materials, and welcome ideas and volunteers!

The GW-Project Board of Directors

January 2023

Foreword

Risks are about events or circumstances that cause problems. Therefore, risk identification starts with attention to the source of the risk followed by awareness of the risk of exposure to humans or the environment and whether those exposure risks warrant attention, and if so, how to manage the risk. The issue of risk exposure and management is challenging because there may be many sources of risk, each with multiple pathways for exposure and options for management, often with limited data for characterization. While we typically seek to avoid risk, there are risks associated with everything a person does, resulting in the need to manage risks to the degree feasible. One example of risk management is to determine whether the risk is sufficiently small that it is comparable to the risk posed by other sources that we accept being exposed to on a regular basis; and if so, decide to accept this new risk because risk cannot be completely eliminated.

Dr. McBean has been involved in risk assessment and management decisions for more than fifty years. He is an expert in mathematical modeling of migration of contaminated groundwater and associated phenomena that influence contaminant levels, and has extensive knowledge of statistical interpretation of data that underlies the materials discussed in this book. The challenge is that the issues are and cannot be fully described in one book. The evolution of information and approaches related to assessing groundwater-related risks have been transformational in recent years. This book offers preliminary insight into this fascinating area of research and engineering practice, with the goal of providing guidance to those interested in pursuing further knowledge of this subject.

John Cherry, The Groundwater Project Leader

Guelph, Ontario, Canada, June 2023.

Acknowledgments

I deeply appreciate the thorough and useful reviews of and contributions to this book.

- Dr. Eileen Poeter, Professor Emeritus, Colorado School of Mines, United States;
- Dr. John Cherry, Adjunct Professor, University of Guelph; Professor Emeritus, University of Waterloo, Canada;
- Connie Bryson, Editor, The Groundwater Project, Canada;
- Bill Rixey, Associate Professor, Cullen College of Engineering, Department of Civil and Environmental Engineering, University of Houston, Texas, USA;
- Alexander Sun, Senior Research Scientist, Bureau of Economic Geology, The University of Texas at Austin, Texas, USA; and
- Matthys Dippenaar; Associate Professor, Engineering Geology and Hydrogeology, University of Pretoria.
- Everton de Oliveira, President of Hidroplan, Brazil, Director-President of the Instituto Água Sustentável (Sustainable Water Institute), Brazil.

I appreciate the contributions of Amanda Sills and the formatting team of The Groundwater Project for their oversight and formatting of this book. I thank Eileen Poeter (Colorado School of Mines, Golden, Colorado, USA) for reviewing, editing, and producing this book.

1 Introduction

Risk assessment is a broad issue, applicable to a vast array of concerns; hence groundwater is but one of many areas in which the concepts of risk assessment are utilized. In general terms, risk assessment begins with varying types and complexity of quantitative and qualitative analysis; then develops appropriate management strategies for problems associated with risks. Risk assessment is designed to offer an opportunity to better understand a system by adding structure and completeness to a problem evaluation. It generally embodies the heuristic approach of empirical learning that will provide a *best knowledge* estimate of the relative importance of risks.

Risk assessment is defined as a systematic process of identifying hazards and evaluating associated risks, then implementing reasonable control measures to remove or reduce the risks.

To better understand the concepts of risk assessment, it is important to establish that risk assessment is not used in isolation rather it is part of the broader context known as risk analysis. This broader term of risk analysis includes risk assessment, risk management, and risk communication, each as an element of the overall procedure. Specifically, each of these three components are contained within the overall subject of risk analysis:

1) <u>Risk assessment</u> may be qualitative or quantitative, although only quantitative risk assessment will be considered in this book. The procedure involves characterization and estimation of potential adverse impacts associated with a hazard. A groundwater-related example is evaluating consumption of water with entrained physical, chemical, and/or microbial agents. The first part of the risk assessment is to identify the hazard(s); the next part is to determine who might be harmed (who are the receptors exposed to the hazard) as well as the extent of exposure—e.g., duration of exposure and the concentration of the chemical; and finally, evaluation of the risk(s) (e.g., how dangerous is the consumption of the chemical). Hence, the aim of the first component involves evaluation of the hazards including assessment of the impacts of removing the hazard or minimizing the level of its risk by adding control measures such as installation of a drinking water treatment system.
2) <u>Risk management</u> is the second part of the process. It involves weighing different possible management alternatives and selecting appropriate actions, based on the findings of the risk assessment. An array of management alternatives may need to be explored including societal values, engineering, economics, legal, and political issues. The management of the risk includes providing guidance regarding how to manage the risk.

3) <u>Risk communication</u> – The third component involves communication of risks and may have many levels including providing information to managers, public officials, and the public. Complicating factors in this phase may include public perception differing from scientific fact and difficulty associated with exchanging scientific information.

To review, risk assessment has three components: source of the risk, transmission/migration from the source to a receptor, and impact to receptor(s) as illustrated in Figure 1. Absence of any of the three components (source, transmission, or receptor) means there is no risk because there is no way that a receptor can be exposed. There are many risk transmission pathways between the source and the receptor (e.g., via groundwater ingestion; via volatilization into the vadose zone resulting in the chemical being drawn into the household environment through a forced air furnace from the soil profile). However, in the event that there is no complete pathway for exposure of the receptor, then there is no risk (at least to that receptor).

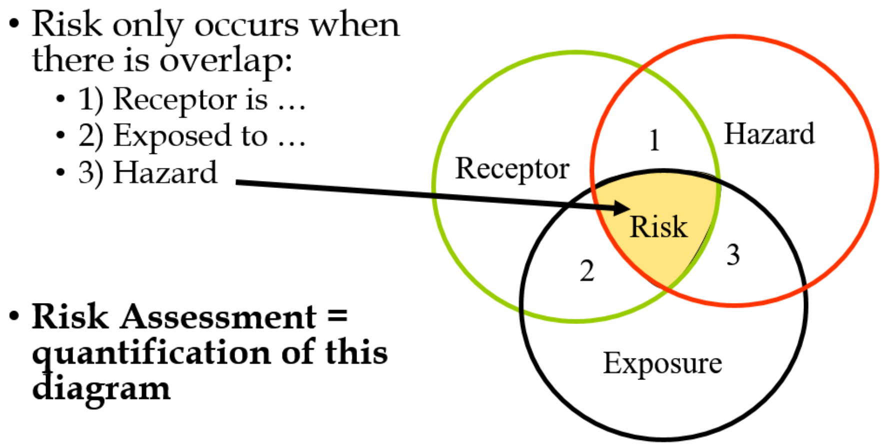


Figure 1 - The three components of risk.

Environmental risk is based on environmental data. There may be shortcomings associated with: characterization of the source; assessment of transmission pathways; and the reception (duration of exposure time due to activities of the receptor). As a result, data interpretation is always necessary as part of a risk assessment associated with groundwater issues. For environmental risk, it is necessary to assess potential severity and probability of occurrence. There are fundamental concepts involved in understanding environmental data that need to be understood so the risk assessment is based on defensible data and defensible interpretation of that data.

Risk is a measure of the combination of probability and severity of adverse consequences of exposure to potential receptors due to a system failure; it may simply be represented by the measure of an event (e.g., for what duration has the water been consumed?). Risk represents the assessed loss potential, often estimated by the mathematical expectation of the consequences of an adverse event occurring.

The result is that risk may be expressed as the product of probability of occurrence and severity of consequence as shown in Equation (1).

$$\text{Risk} = p\ S \qquad (1)$$

where:

p = probability of occurrence

S = consequence or severity of occurrence

In fact, the level of risk is dependent on the degree of hazard as well as on the number of safeguards or preventative measures against adverse effects; consequently, risk can also be defined by the following conceptual relationship:

Risk = hazard/preventive measures

or

Risk = f(hazard, exposure, safeguards)

where preventive measures or safeguards are considered to be a function of exposure, or to be inversely proportional to the degree of exposure.

Risk assessment and management associated with groundwater quality involve the use of factual information to define the potential health effects arising from exposures of humans and the environment. For example, this frequently involves determination of the extent to which a chemical migrates from the location of a spill, enters the groundwater, and reaches a drinking water well. Risk management associated with groundwater involves weighing policy and remediation alternatives; integrating the results of risk assessment with engineering data along with social, economic, and political concerns; and reaching a decision about management of the risk.

A major objective of risk assessment is to help develop risk management decisions that are systematic, comprehensive, accountable, and self-aware. It has long been recognized that nothing is either wholly safe nor dangerous per se, but that the objective involved and the manner and conditions of use determine the degree of hazard or safety.

This book discusses methods to estimate and manage risks. Issues of risk communication are beyond its foci. This book is focused on risk associated with groundwater, however the principles are broad and can be applied to a wide array of circumstances.

2 The Case for Risk Assessment Leading to Risk Management

2.1 Understanding Risk

The concept of risk is broad because it applies to a vast spectrum of issues. As applied to groundwater, an important dimension of risk is the possibility of experiencing harm from a hazard such as unacceptable groundwater quality that would make water unusable or dangerous for drinking. Carrying this further, risk assessment may be construed as, for example, evaluating the likelihood or frequency that groundwater quality will be unacceptable (e.g., exceed the water quality standard for certain uses and/or assessing seasonal variations of groundwater quality).

Since risks are about events that may cause problems (e.g., illness or death), normal practice is to start with attention to the source of the risk. This might involve completion of a *risk assessment* to evaluate the available scientific information including items such as a dose-response relationship and the extent of human exposure to a chemical where some indicator (e.g., the concentration of a chemical) suggests consumption of the groundwater for drinking represents a hazard.

However, there needs to be a distinction between *hazard* and *risk* as referred to in Section 1. Risk can be characterized by consequences or impacts multiplied by probability as shown in Equation (1). On the other hand, a hazard is the potential for harm or an adverse effect (e.g., a chemical that can cause illness if there is an exposure pathway to a person). It is important to understand this difference so that resources can be directed to actions based on whether there is just the existence of a hazard or whether there is truly a risk.

A *hazard* is something with the potential to create undesired adverse consequences (e.g., a chemical that is a carcinogen such as arsenic); exposure is the vulnerability to hazards (e.g., consumption of water that contains arsenic). If the arsenic concentration is sufficiently high, then death may be a consequence soon thereafter. However, if the arsenic concentration is relatively low but the consumption of the contaminated water occurs over a significant duration, then cancer may form and cause illness/death as a result of the lengthy exposure perhaps over many years. Hence, risk is the probability or likelihood of an adverse effect due to a hazardous situation. The assessment of potential hazards posed by, for example, chemicals in drinking water, must involve a critical evaluation of available scientific and technical information on the chemical, the vulnerabilities of potential receptors likely to be exposed, and the possible mode and duration of exposure.

The difference between *risk* and *uncertainty* also needs to be understood. As noted above, risk can be characterized by consequences or impacts multiplied by probability. On the other hand, uncertainty is a term that refers to lack of knowledge. There are several types of uncertainty with the primary types being epistemic and aleatory uncertainty.

Epistemic uncertainty is related to a single case that may occur or a single statement that may be true. Aleatory uncertainty is related to the probability of alternative outcomes in repeated experiments.

Some risk implications are only partially observable. For example, uncertainty might exist due to limited knowledge of exposure if there is not (yet) sufficient information about the magnitude of a future exposure scenario in terms of groundwater quality. For example, per- and polyfluoroalkyl substances (PFAS) are emerging contaminants that we do not know much about yet. We do not have sufficient information to understand the degree to which they represent a health hazard, nor are exposures always well understood (Dourson & Gadagbui, 2021).

Both risk and uncertainty are highly relevant to the subject of risk assessment and both concepts must be incorporated as part of a risk assessment. For example, if the concentration of a chemical in the groundwater is known, and its hazard is sufficiently understood, there may be sufficient information available to characterize the risk of illness arising from the consumption of the groundwater as drinking water. However, there may also be uncertainty in the measured concentrations and/or the described hazards, which would lead to ramifications regarding the impacts to human health of particular chemicals in the groundwater quality associated with a particular chemical. Understanding both risk and uncertainty are important because any the approximately 5,000 new chemicals introduced to the marketplace each year may become an emerging contaminant given that the adverse impacts associated with some of the contaminants are not known until later (McBean, 2019).[1]

As a demonstration that the presence of a substance in groundwater and health issues are separate but subtle, McBean (2019a) presents three chapters that discuss risk assessment and eight chapters that address characterization of uncertainty (i.e., how many data values are available and hence the uncertainty in the assignment of parameters such as concentrations).

Another important element of risk assessment is that society is moving towards more involvement in many decisions. For example, if it is proposed that a sewer pipe be installed to carry leachate from a solid waste landfill to a wastewater treatment plant, the

[1] PCBs (Polychlorinated Biphenyls) are a classic example. The issues of emerging contaminants have become well known. The first emerging contaminants to be acknowledged were PCBs. Initially, PCBs were heralded as a miracle chemical because they never lost their lubrication characteristics. They were later identified as causing widespread health issues. After evidence evolved over decades, it was learned that they are an extremely dangerous chemical. PCBs are no longer allowed to be used in a vast majority of possible applications (Barbalace, 2022).

community is likely to be involved in evaluating whether there is risk to an individual who frequently consumes water from a nearby well if the pipe leaks.

Typically, input from the public (people living near where a sewer pipe will be placed, and themselves relying on groundwater wells for their water supply) are frequently focused on rejecting the potential of having possible risks imposed upon them. People may not want to take on that risk and will therefore reject any proposal that elevates their potential risk. Popular acronyms for this kind of thinking are NIMBY for "not in my back yard" or NOPE for "not on planet Earth."

Hence, an appreciation of public attitudes regarding risk is essential. The result is a need to associate the risk with the chance of its occurrence. This requires quantification of its probability of occurrence. It requires that the procedure for risk assessment must be logical and transparent, to facilitate dialogue with the public about how the risk is calculated. In practice, it is typically required that the risk of chemical exposure to homeowners be minimal and reasonable. This might be as simple as comparing the risk of the probability of an unfamiliar event (e.g., leakage from the sewer pipe) to the risk of an event with which the public has some familiarity (e.g., getting cancer from cigarette smoking or having a car accident).

2.2 Risk Assessment

In short, there is a multi-stage process by which the elements of risk are calculated. First, the risks are identified and this is followed by assessment of whether their management is both necessary and feasible (i.e., Is the risk sufficiently severe to warrant efforts to control or manage the risk?). This procedure needs to consider the probability of occurrence. Individual risks are considered to be the frequency at which a given individual could potentially sustain a given level of adverse consequence from the realization or occurrence of specific hazards.

Development of a risk assessment and decisions regarding the need for management of the risk is frequently a work-in-progress as the profession continues to improve assessment procedures. However, as procedures improve, even more issues and uncertainties are identified. Regardless, it is important to understand the underlying principles and incorporate updates in understanding as they evolve over time when data continue to be collected and the knowledge base expands.

This book addresses risk in its most fundamental form. The approach is:

1. identify a hazard,
2. analyze the exposures associated with that hazard,
3. determine the risk, and
4. determine if the elimination or control of the risk is warranted (e.g., this might include evaluating whether the groundwater should be treated before being consumed).

As a consequence, risk associated with groundwater has many dimensions. For example, does a nearby landfill leach contaminants to the groundwater, does that contaminated groundwater move to the point where someone may be drinking the water, and is there a reasonable basis to determine if that risk will cause an impact to the health of that individual?

Risk is relevant to the public in a wide spectrum of dimensions. Examples of risks that an individual may encounter daily are indicated in Table 1. Each day, people weigh risks of different types of activities (recognizing they may not be aware of any degrees of elevated risk) and decide whether or not to participate. The thrill of adventure may also influence their decision as to whether to participate.

Table 1 - Examples of daily risks encountered by the public.

Examples of risks	Basis for risk
Turning on the light	Possible electrocution
Falling on stairs	Injury from falling
Caffeine in coffee/tea	Elevation of blood pressure
Sweetener in coffee/tea	Sugar-related disease exacerbated
Peanut butter	Exposure to aflatoxin (a mold) and risk of liver cancer where, particularly in the developing world, the storage of peanuts is a substantial issue
Riding a bicycle to work or for exercise	Increased potential for an accident
Drinking of water	Cancer resulting from the disinfection by-products arising from chlorination of water
Brick and cinder blocks	Exposure to radon and hence potential for development of cancer
X-rays for disease identification	Identifying disease may, by itself, cause cancer
Air travel	Exposure to ultraviolet radiation; potential for a plane crash or terrorist activity

The ramifications of the array of issues encountered daily indicate that in most respects, individuals are exposed to many risks. It should also be apparent that individuals cannot reduce all risks to zero, but we can—and often do—avoid some risks. Different approaches are essential and appropriate in relation to different circumstances, and different extents of data may be available. The strategies and concerns related to groundwater quality, as an example, must consider risks with respect to both humans and the environment (e.g., consider that the impacts to vegetation from salinity are, in many respects, more challenging than the impacts to humans; hence, the realization of the breadth of possible concerns associated with groundwater are profound).

Ultimately, risk assessments need to include both human health and the environment. However, risk assessment is complicated by a spectrum of dimensions such as the degrees to which a chemical bio-accumulates and biomagnifies, the importance of ecological modeling, and ecological dose-response methodologies for numerous species. Thus, the field of risk assessment and management is data intensive as it applies to groundwater concerns (as well other activities) and requires careful consideration of data variability and availability in both space and time.

Risk management procedures go far beyond risk assessment. They need to consider with multi-faceted aspects of a problem and they need to be responsive to the severity of conditions, the cost of gathering additional data, and the degree to which management is feasible.

2.3 The Need for Risk Quantification

Risks may be characterized as *statistically verifiable* or *statistically non-verifiable*.

- Statistically verifiable risks are risks associated with voluntary or involuntary activities that have been determined from substantial observation such as statistics related to potential for exposure to arsenic in groundwater or in food irrigated with arsenic-contaminated groundwater (Joseph et al., 2015a, 2015b). These types of risks (e.g., cancer which may develop from consuming arsenic-contaminated water) may be feasibly compared to one another since substantial data are available.
- Statistically non-verifiable risks might arise from involuntary activities that are based on very limited data sets and mathematical equations (Sharma et al., 2007). An example is using the interpolation from two chemicals for which there are data to infer the likelihood of health impacts of a third chemical. As another example, we may know that the risk of a nuclear energy generation incident killing a person is low, but because very few events have occurred, there may be significant challenges in establishing probabilities. This type of risk assessment might then require reliance upon qualitative risk assessments, as opposed to quantitative risk assessments as described in Chapter 3 of McBean (2019a).

Hence, while aspects of statistically verifiable and non-verifiable risks are similar in some respects, they are also very different. This translates to the need to consider both aspects, although a comparison of them is not always feasible.

One of the most challenging aspects related to risk assessment is that the majority of people neither understand nor quantify the risks they face on a day-to-day basis. Most people behave as if life is largely free of risk (or, at a minimum, the thrill of participation allows them to accept the risk of participating in a dangerous activity).

Risks will usually be expressed as the probability of effects associated with a particular activity (e.g., drinking groundwater with low levels of arsenic). Alternatively, there can be issues as to whether the groundwater quality is deteriorating or only appears to be deteriorating due to poor characterization of groundwater quality, often due to the availability of only a few measurements (McBean & Rovers, 1985, 1992).

A comment made by many is that the world seems like a hazardous place, but it may be that we are now more aware of risks due to increased societal focus and there is increasing awareness of the issues of risk. For example, if we look back at the world of a century ago, Wilson (1979) indicated life expectancies have increased substantially from 50 years to than more than 70 years. Therefore, the sum of all risks must be less than it was historically.

The assessment procedures for managing risk in some situations will be obvious. For example, if the concentration of a chemical in drinking water causes an unacceptable health risk, then either water treatment must occur, or an alternative water source must be identified. As a result, ways of managing risk are a natural outgrowth of risk assessment. It is important to understand the underlying principles because they are not subject to change and are applicable in a broad sense to risk assessment. Some are amenable to quantitative analyses whereas some must remain qualitative.

Statistical interpretation of environmental quality data has a major role to play in areas such as qualifying effects, assessing consequences, measuring risks, and interpreting evidence.

2.4 Sources of Variability in Groundwater Quality Data

The variable quality of groundwater data greatly influences which types of statistical analysis will be effective. The specifics of statistical analyses depend on the way the phenomena of interest is defined and sampled. In general, the ability of groundwater quality data to characterize the population from which samples are drawn depends on the following:

- how many samples are available,
- the degree and randomness by which the sample data were obtained,
- the degree of independence between the observations in the sample, and
- the strategy to be utilized in a risk assessment.

If extensive monitoring is required, data assembly may be expensive. Nevertheless, in many assessments of environmental phenomena, estimates of groundwater quality must be developed from sparse data records (i.e., records that are limited in both time and space). The result is that frequently a small amount of data is usable for specific applications.

These features of groundwater data sets lead to difficulty in their statistical interpretation for some situations. The analysis procedure utilized in risk assessment must be sensitive to small changes (e.g., early detection of a contaminated groundwater plume indicating that a plume of contamination is starting to arrive at point of groundwater withdrawal for a community) and yet a point of diminishing returns occurs when the cost of additional data is not warranted given the minimal new information that can be gleaned from it.

Consequently, for many occasions there must be careful consideration as to how statistical analyses for risk assessment should proceed. Many statistical analysis techniques that are valid/useful for chemicals with long records of exposure have little utility in situations where only a short record is available. A further complication is that many of the data sets are highly variable or *noisy* due to factors such as seasonal phenomena, for example.

An additional dimension of increasing challenge is that improvements in instrumentation allow measurement of lower concentrations. In these situations, data available prior to that time that was determined using an instrument that had a higher lowest detection limit can only be expressed as censored data (e.g., rather than having a specific value, it can only be noted as less than whatever the lowest detection limit was at the time of analysis—perhaps <10 mg/L). The result is that problems associated with statistical analyses of censored data sets are increasingly challenging. Further, a number of chemicals have maximum concentration levels (MCLs) to which humans and the environment can be exposed that are very close to the minimum concentration that can be detected by available instrumentation.

For the variety of reasons indicated in this section, there is no single approach to statistical analysis in risk assessment. Instead, a series of approaches is frequently needed, with each approach providing useful information that is appropriate to address a particular question. Statistical interpretation of environmental quality data has a major role to play in such areas as qualifying effects, assessing consequences, measuring risks, and interpreting the ramifications of various actions (Unwin et al., 1983; McBean & Rovers, 1992).

2.5 Independence of Successive Data Values

Time series analyses are pertinent to the problem of estimating trends and cycles (e.g., seasonal variations). There are differing degrees of independence between each measurement in a site investigation and this must be assessed during statistical analysis of the data. For example, when a groundwater monitoring well yields high chloride concentrations today and similar values tomorrow, and nearby monitoring wells also yield elevated chloride concentrations, the values are not necessarily independent of each other. Similarly, replicate sampling (e.g., the splitting of one sample into several samples) does create independent observations. These issues need to be considered in data analyses.

3 Examples of Risk Interpretation Procedures for Groundwater Quality

3.1 Background to Groundwater Quality Risk Assessments

The issue of a risk assessment is complicated by the, typically, sparse data sets resulting from the expense of data collection as well as the need to take action soon after contaminants are discovered while the contaminant moves slowly in the environment hindering measurement of the contaminant plume evolution. Issues with the time-evolution of data sets arise because it can take a long time for the risk (e.g., the chemical) to travel from its source to the receptor. For example, a chemical release must travel through the soil, reach the water table, then migrate to the location where a receptor is exposed to the chemical in their well water. Further, the collection of groundwater samples, the need for laboratory analyses of the samples, and interpretation of the information can represent significant expenditures. Consequently, diligence must be utilized when assessing the resulting data to ensure the interpretations are both comprehensive and defensible.

Some understanding of risk exposure is gained from enhanced statistical interpretation capabilities now available to interpret environmental quality data. Statistical analyses are not an interpretation of the facts. Rather, when properly used, statistical analyses make the facts easier to grasp so that other considerations can enter into decision related to risk assessment and management.

Issues of risk assessment and management in groundwater quality issues are broad. Many exposure pathways may exist, and many contaminants may need to be included. As a result, the description of water quality risk assessments cannot be encompassed within a few examples. Instead, the following section focusses on providing general guidance and references in the technical literature that provide specific details.

3.2 Scenarios of Exposure Concentrations Causing a Risk

As environmental data are assembled it is useful to characterize its probabilistic distribution. The distribution can be used to establish the probability that a particular value will be exceeded, which can be used to determine if drinking the water presents a serious exposure risk. For example, if groundwater quality data can be represented by a Gaussian (i.e., normal) probability distribution, then the mean and standard deviation can be used to assess exposure risk and estimate if an individual is likely to become ill as a result of consuming the groundwater.

Probability distributions that are commonly utilized in groundwater risk assessments include the normal distribution (Box 1⤵), lognormal distribution, as well as the Gumbel or Log-Pearson distribution where the data are skewed (i.e., more of the data are on one side of an otherwise bell-shaped curve. Once a determination is made that a particular distribution (e.g., the Gaussian distribution) reasonably describes the data, then exceedance of a specific concentration can be calculated and an estimate of exposure risk can be made. This type of risk assessment is relatively straightforward, as long as sufficient data are available to determine the probability of illness for the chemical of concern.

Extensive efforts are ongoing to determine the extent of illness arising from chemical exposure. Part of the difficulty in these determinations is the enormous array of chemicals that exist and the desirability to decrease the risks of illness to extremely low levels (i.e., *de minimus*, or negligible risk).

The following sections present examples to demonstrate how issues of uncertainty, risk assessment, and risk management need to be considered.

3.2.1 Example 1: Arsenic concentration exposure for people using untreated groundwater as their drinking water source

A small village draws water from a nearby well for their drinking water. Arsenic is a common groundwater contaminant, sometimes from natural causes and sometimes from an industry in the neighborhood. For example, as described in Farrow and McBean (2016) and McBean (2013), over millennia, erosion from the Himalayan mountains resulted in arsenic deposits in river deltas, one of which flows through Bangladesh and now causes widespread issues of arsenicosis throughout much of the region.

The arsenic concentration data from the well are assumed to be lognormally-distributed because the data can be approximated by a straight line when plotted on lognormal paper as confirmed by plotting the data later in this example.

The lognormal distribution is frequently utilized to assign probabilities of exceedance values of a chemical because concentrations less than zero are not possible, just as a negative value of a log is not possible. Values of a logarithmic distribution cannot be negative and are unbounded on the high end.

The individual values of arsenic concentration data are listed in Column 2 of Table 2. Column 4 of Table 2 lists the rank-ordered values of the arsenic concentrations. Column 6 lists the log-transformed (ln) arsenic concentrations of the individual values from successive analyses of the groundwater quality from monthly sampling.

Table 2 - Arsenic concentrations (McBean, 2019a)

Sample number	Arsenic concentration (µg/L)	Rank	Rank-ordered concentration data	Weibull plotting position	Log-transformed concentration (base 'e')
1	1.1	1	22.7	0.06	3.122
2	1.8	2	15.4	0.13	2.734
3	2.7	3	10.2	0.19	2.322
4	15.4	4	8.6	0.25	2.152
5	8.6	5	6.6	0.31	1.887
6	2.75	6	5.1	0.38	1.629
7	3.2	7	3.7	0.44	1.308
8	2.2	8	3.2	0.50	1.163
9	10.2	9	2.9	0.56	1.065
10	2.9	10	2.75	0.63	1.012
11	5.1	11	2.7	0.69	0.993
12	22.7	12	2.4	0.75	0.875
13	2.4	13	2.2	0.81	0.788
14	3.7	14	1.8	0.88	0.588
15	6.6	15	1.1	0.94	0.095

Notes: The Weibull plotting position is m/(n+1) where *m* is the rank and *n* the total number of samples.

The Weibull plotting formula, namely $m/(n + 1)$, is applied where m is the rank of the value and $n = 15$, the total number of samples.

The next step involves plotting the rank-ordered data on probability paper as illustrated in Figure 2. The data plot acceptably well as a straight-line (Figure 2) so the assumption of a log-normal distribution is reasonable. Using the 5 percent probability value along the horizontal axis and projecting up to the fitted line we find the natural log of the arsenic concentration that will be exceeded 5 percent of the time is 3.1. In arithmetic terms the exponential of 3.1 is 22.7 µg/L. Finally, being risk-averse this value is assumed to be the arsenic concentration 95 percent of the time. Thus, a concentration of 22.7 µg/L will be utilized in the risk assessment procedure to estimate whether the villagers drinking the untreated groundwater may develop cancer over time as a result of this arsenic exposure pathway.

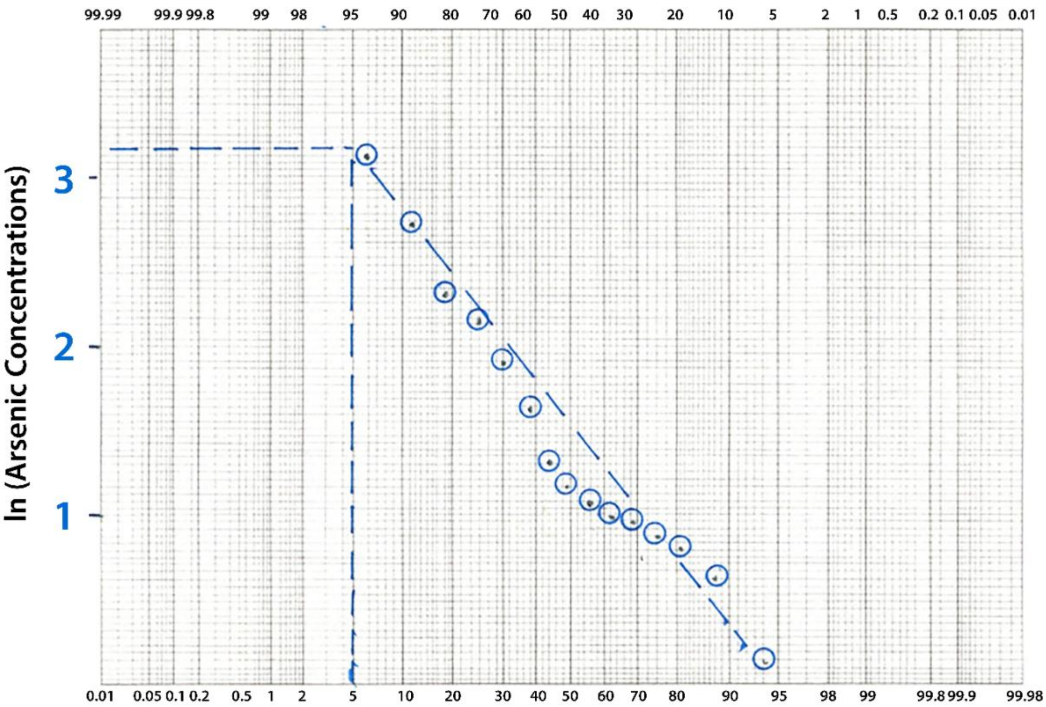


Figure 2 - Probability of exceedance of arsenic concentrations (ln = log transformed).

Additional insights into arsenic contamination are provided by Joseph and others (2015a, 2015b).

3.2.2 Example 2: Estimating intake (dose) of arsenic and probability of villagers developing cancer as a result of exposure to arsenic in the groundwater

Using the result of Section 3.2.1 that produced a concentration of 22.7 μg/L, we estimate the *intake rate* or *dose* of arsenic. Intake rate is typically expressed in terms of the mass of a chemical (in this case arsenic) that enters the body per unit of body weight (*BW*) per unit of time. The average daily intake is calculated using Equation (2).

$$\text{Intake (mg/kg}_{\text{BW}}\text{/day)} = \text{LADD} = \frac{(C)\ (IR)\ (EF)\ (ED)}{(BW)\ (AT)} \tag{2}$$

where:

$LADD$ = lifetime average daily dose

C = concentration in the water

IR = intake rate of contaminated media in L/day

EF = exposure frequency in number of days per year

ED = exposure duration in years

BW = body weight in kg

AT = averaging time in days

The time over which the intake is averaged (*AT*) varies depending upon how the exposure occurs. For this example, assume the *BW* of the adult is 70 kg, the drinking water ingestion is 2 L/day, and the exposure frequency is 365 days/year. The villager lives in the village throughout the year and is assumed to continue to do so for 70 years.

$$\text{LADD} = \frac{(22.7\ \mu\text{g/L})\ (2\ \text{L/day})\ (365\ \text{days/yr})\ (70\ \text{yrs})}{(70\ \text{kg}_{\text{BW}})(70\ \text{yrs})(365\ \text{days/yr})}$$

$$= \frac{0.649\ \mu\text{g}}{\text{kg}_{\text{BW}}\ \text{day}}\frac{1\ \text{mg}}{1000\ \mu\text{g}} = 0.000649\ \frac{\text{mg}}{\text{kg}_{\text{BW}}\ \text{day}}$$

Arsenic is a known carcinogen. Carcinogen exposure is assumed to be cumulative; each additional exposure is assumed to increase the likelihood of developing cancer, even if the exposures are separated by a period of years. No exposure threshold exists for carcinogens whose mode of action involves mutations; if the dose is greater than zero, risk is greater than zero.

To quantify the risk associated with exposure to carcinogens, the LADD is multiplied by the *cancer slope factor* (CSF) for the chemical, arsenic in this case, resulting in an estimate of the *incremental excess lifetime cancer risk* (IELCR) as shown in Equation (3).

$$\text{IELCR} = \text{LADD} \times \text{CSF} \tag{3}$$

The CSF (also called the *potency value*) is derived from an estimate of *unit risk* that describes the incremental risk per unit intake of a carcinogen based on epidemiological and toxicological data and modeling. The CSF is typically expressed in units of $(\text{mg/kg}_{\text{BW}}/\text{day})^{-1}$. For arsenic, the CSF is 1.5 $(\text{mg/kg}_{\text{BW}}/\text{day})^{-1}$. McBean (2019a) provides more information on this topic. Thus, the incremental life time risk is 9.7×10^{-4} as shown here.

$$\text{IELCR} = \frac{0.649 \times 10^{-3}\text{mg}}{\text{kg}_{\text{BW}}\text{day}} \times 1.5 \frac{1}{\left(\frac{\text{mg}}{\text{kg}_{\text{BW}}\text{day}}\right)}$$

$$\text{IELCR} = 9.7 \times 10^{-4}$$

Usually, the CSFs are based on upper bounds of a computer model to determine a risk level that is unlikely to be exceeded. However, this also means that the risk could be less, in fact, according to Subramanian et al (2006) the risk could even be zero. This value of upper bound risk, IELCR, of a villager consuming this water over a 70-year lifespan is 9.7×10^{-4}.

It is noted that *de minimus* risk, or negligible risk, is an upper bound interpreted as one-in-ten thousand to one-in-a-million, shown as 1×10^{-4} to 1×10^{-6}. While desirable to have “zero” risk, humans are continuously exposed to risks. The villagers may ride a bicycle to work in a factory and could be injured during both activities. The United States Environmental Protection Agency typically defines acceptable carcinogenic upper bound risk as 1×10^{-4} to 1×10^{-6}; risks greater than 1×10^{-4} are unacceptable and require the implementation of risk management measures. In this simple example, the upper bound risk of death from cancer is 9.7×10^{-4}, which is considered unacceptable and should be managed by an arsenic removal technology prior to use of the well as a drinking water source.

This example is only a single example of a risk assessment that might be carried out. Much more detailed risk analyses can be completed, with thousands of variations and assumptions involved.

3.2.3 Example 3: Scenario of a probabilistic environmental risk assessment

While the assessment in the previous section is relatively straightforward because the focus is on exceedance of some measure of risk associated with a specific chemical, there are greater challenges when many elements of risk are involved. The situation depicted in Figure 3 demonstrates such complications. Figure 3 shows a landfill containing refuse from a city. It is a sophisticated landfill with a cover and leachate collection system. The cover and collection system were designed to control infiltration through the surface cover material into the refuse where contaminants in the waste material are dissolved into the percolating water, creating leachate. The bottom, low-permeability liner is intended to intercept the migrating leachate and direct the leachate to the leachate collection system. The leachate collection system involves a series of perforated pipes that capture the majority of the leachate and transport it to a leachate treatment system prior to disposal back to the environment.

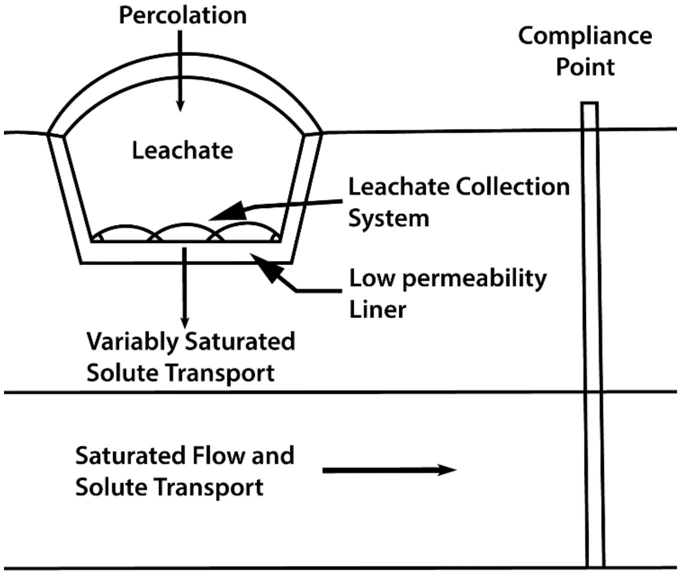


Figure 3 - Schematic of modeling components of exposure pathways impacting the water quality at the point of compliance (McBean, 2019a).

The depiction of the landfill and its components illustrates that constructing a high integrity landfill involves substantial expense—however, there are degrees to which expenditures to control the migration of leachate can be decreased. Of interest are these questions: If people are using the water as a water supply, what are the concentrations to which these people may be exposed? Is this level of exposure a risk to the people relying upon this extracted groundwater? Additional expenses might involve preparation of a better landfill liner to ensure greater capture of the migration leachate and, hence, decrease the concentration of contaminants in the water at the compliance point.

In this scenario, the issue is one of determining the exposure risk associated with a landfill as related to the drinking water quality at a downgradient well (the compliance point). Each element of the landfill design, as well as the migration pathways of the leachate to the groundwater well, must be evaluated. Examples of measures that are relevant in this type of evaluation include:

- *The volume of leachate generated from the landfill*. Considerations include the magnitude of rainfall, evaporation, and infiltration at the landfill surface and, ultimately, how much liquid might then drain into the refuse and how much leachate would enter into the leachate containment system.
- *The degree to which seepage of leachate creates a hydraulic mound on landfill liner*. This depends on the characteristics of the leachate collection system and the liner (Murray et al., 1995). Issues that influence the degree of mounding on the low permeability liner (Figure 4) are the quality of the drainage sand that allows the passage of leachate horizontally to the drainage tiles, the spacing of the drainage tiles, and the permeability of the liner. Further, the permeability of the liner system is influenced by how carefully it was installed. If the liner is a *flexible membrane liner* (e.g., a high-density polyethylene liner), the integrity of seals where rolls of material overlap influence the ability of the leachate to migrate through holes or gaps along the seams. Comprehensive quality assurance of the sealing at time of placement influences both the cost of liner placement and the potential for excursion of leachate through the membrane liner.

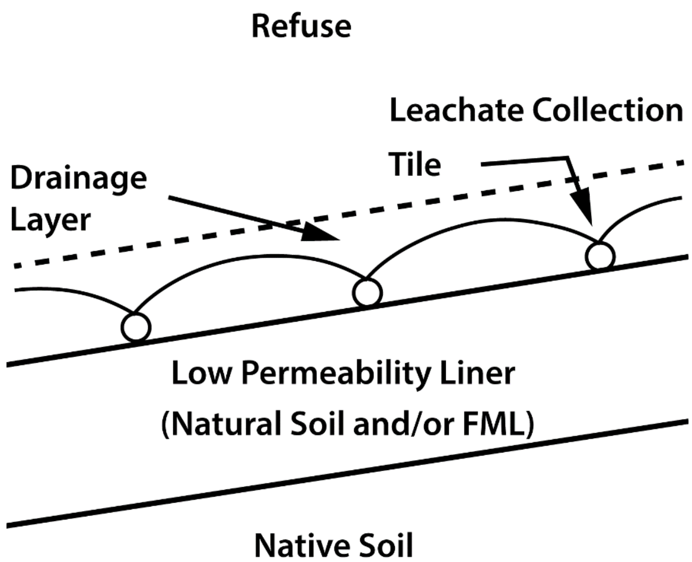


Figure 4 - Leachate collection and liner system (McBean, 2019a).

- *How the leachate is collected and removed*. To allow collection and removal of the leachate, an array of leachate collection tiles are placed at specified intervals to drain the collected leachate over to the edge of the landfill, followed by removal of the leachate and subsequent treatment. Closer spacing of the leachate collection tiles would reduce the mounding over the flexible membrane liner, decreasing the hydraulic head and hence decreasing the volume of leachate migrating through the low permeability liner. However, closer spacing of the leachate collection tiles increases the cost of constructing the landfill. This is another example of the trade-offs between the risk of migration of leachate and the cost to construct the landfill.
- *Additional protections*. For additional protection against excursion of leachate through the low permeability liner at the bottom of the landfill, a low permeability clay layer can be placed beneath the flexible membrane liner. Again, the placement of both a flexible membrane liner and a clay soil layer below the liner would reduce risk, but at increased cost.

Expenditures to decrease the risk of leachate migration to the groundwater and subsequently to the groundwater well can be characterized. It is a matter of trade-offs between the expense and the risk. This complicated scenario of exposure risk is also more challenging in terms of calculations for each of the dimensions of the decision variables (e.g., increasing/decreasing the quality of the clay layer to ensure that it is placed to minimize leakage involves additional costs). At what point does a designer decide the risk is sufficiently low that additional expenditure to decrease the risk is not warranted?

A methodology to identify the trade-offs between the cost and exposure risk associated with a landfill collection and liner system is summarized in Murray and others (1996). The results are useful in demonstrating the point at which additional levels of sophistication in the design of a leachate/liner system do not produce a significant reduction of exposure risk. This was demonstrated through use of a Monte Carlo computer modeling effort. Each of the risk variables was assigned in accord with the probability distribution pertinent to the parameter, as briefly summarized in Table 3.

Table 3 - List of parameters used in to model risk for migration of leachate to the point of exposure at a groundwater well (McBean, 2019a).

Examples of risks	Brief characterization of uncertainty
With landfill and leachate/liner components	
Percolation rate into the refuse	Geometric mean of 0.070 m/yr and a $\log_{10}$ standard deviation of 0.110
Drainage layer hydraulic conductivity within leachate/liner system	Geometric mean of 1 x 10^{-2} cm/s and a log10 standard deviation of 0.333
Natural soil hydraulic conductivity	A geometric mean of 7.68 x 10^{-9} cm/s and $\log_{10}$ standard deviation of 0.2646 for high-quality clay
Hole frequency in the flexible membrane liner	A geometric hole frequency with mean of 125 holes/ha for a site with poor quality assurance and quality control
Initial leachate solute (chloride) concentration	A geometric mean of 3,500 mg/L and a $\log_{10}$ standard deviation of 0.052
Fraction of municipal solid waste comprised of soluble chloride mass	A geometric mean of 0.139 percent and a $\log_{10}$ standard deviation of 0.052
Municipal solid waste dry density	A geometric mean of 327 kg/m3 and a log10 standard deviation of 0.056
With variably saturated zone modeling	
Longitudinal dispersivity	A geometric mean of 1.0 m with a $\log_{10}$ standard deviation of 0.023
Apparent molecular diffusion coefficient	A geometric mean of 0.40 and a $\log_{10}$ standard deviation of 0.100
Average degree of saturation	A geometric mean of 0.40 and a $\log_{10}$ standard deviation of 0.100
With saturated zone modeling	
Groundwater recharge	A geometric mean of 0.110 m/yr and a $\log_{10}$ standard deviation of 0.087
Hydraulic conductivity of aquifer	A geometric mean of 273 m/yr with values ranging from 189 to 471 m/yr
Porosity	A geometric mean of 0.30 and a $\log_{10}$ standard deviation of 0.059
Longitudinal dispersivity	A geometric mean of 0.40 m and a $\log_{10}$ standard deviation of 0.133

The risk/cost trade-offs for one of the examples are depicted in Figure 5; on the vertical axis is the compliance point concentration which is an indicator of the exposure risk for those individuals who are using the water from the compliance point as their drinking water. On the horizontal axis, is an indication of the cost of the liner system. As more money is expended (i.e., points further along the horizontal axis) the quality of water at the compliance point improves). This figure shows how additional expenses in liner construction improve quality of the water. The question becomes, what risk level should be used to decide the expense to be incurred?

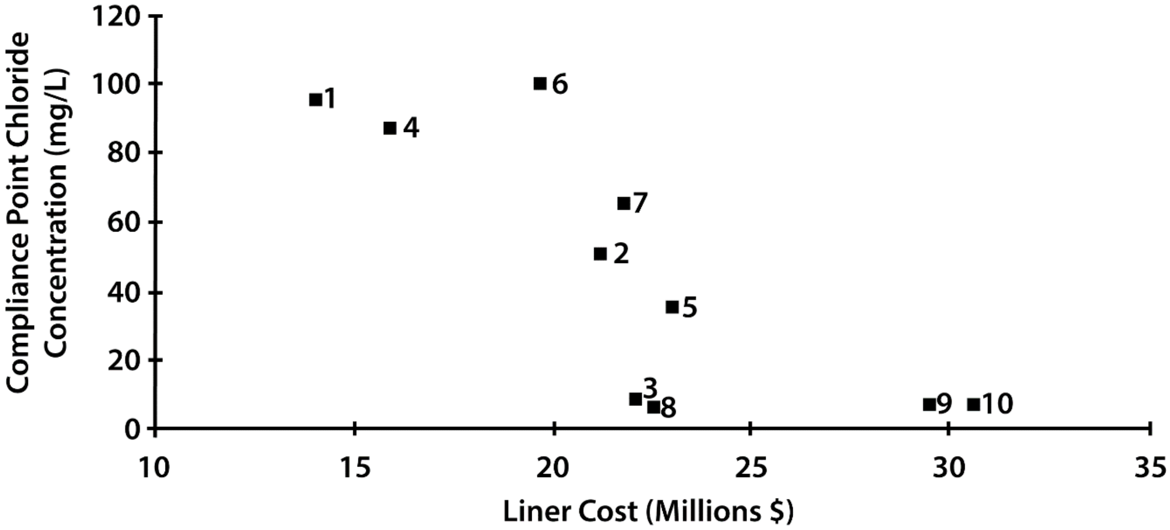


Figure 5 - Expected values of the compliance point chloride concentration versus liner cost expense for different modeling scenarios indicated by numbers next to the data point. The scenarios involve various drainage blanket materials, various clay types and qualities, various leachate tile spacings, and different quality assurance/quality control undertakings.

Figure 5 reveals that expense associated with the ninth and tenth scenarios of design value combinations for each of the components in the risk calculations do not provide significant improvement over the preceding eight scenarios. At some point, a decision is needed as to the expense sufficient to reduce the risk to an acceptable level.

This example illustrates how to undertake an assessment of cost versus risk for a landfill design. Assessing and modeling particular design scenarios, provides the opportunity to determine the level of diminishing return, where additional expense does not provide lower risk, and where the reduction of risk may not warrant the additional expense.

4 Wrap Up

Assessment of risk associated with drinking groundwater involves the use of factual information to define the potential for exposure of individuals consuming the water and the resulting health effects. Management of the risk associated with groundwater usage includes the process of weighing policy and remediation alternatives. It involves integrating the results of risk assessment with engineering data and social, economic, and political concerns to reach a decision involving the management of the risk.

The array of possible contaminants is enormous (e.g., about 5,000 new chemicals are introduced to society each year). The understanding of the fate and transport characteristics of chemicals is improving but continues to be a challenge, due to the opportunities for contaminants to reach humans via different pathways and to be transformed by physical, chemical, and biological processes along the path.

Future issues of adequacy of both data quantity and quality, and the overall issues of water security are poised to become some of the most important concerns of the twenty-first century. As a result, groundwater risk assessment is multi-dimensional. Further, opportunities exist to improve evaluation of risk applied to groundwater quality and quantity. This book presents a few examples of how risk assessments are undertaken.

The intent of this book is to emphasize that the key to risk assessment for groundwater must include the elements in Figure 1. Consideration must be given to the hazard, exposure, and receptor. There needs to be a linkage between these dimensions for a risk assessment procedure. When the hazard and receptor are linked by exposure pathways, there is a risk assessment issue. Groundwater risk assessment is a large and evolving field. The descriptions here offer a few indications of risk assessment procedures.

Questions of uncertainty are not unique to matters of risk assessment and management. This book describes the general character of risk and relies on some specific examples to demonstrate how issues of uncertainty, risk assessment, and risk management need to be considered. In the broader view, the field of risk assessment and risk management is large and multi-faceted. McBean (2019a) describes more details of both risk assessment and risk management procedures as they relate to groundwater. In essence the problem in determining risk is insufficient data. If the future leads to a plethora of data on site conditions and exposure impacts then risk could be readily assessed, so perhaps for today's society there is only uncertainty as opposed to a risk.

5 Exercises

Exercise 1

The background measurements for a groundwater monitoring well are 0.8, 3.1, 1.7, 0.6, 1.1, 2.8, 1.8, and 0.9 for magnesium. Calculate the 95 percent limits on the mean of the background concentrations using these data. If you are not familiar with calculating mean, standard deviation, and confidence limits for a set of sample data, search the Internet to find the formulas. The table of the student's t-distribution that is needed to find the values for calculating 95 percent confidence limits is provided in Box2↴. If you have difficulty determining the procedure, the solution provided for this exercise demonstrates it.

Click for Solution to Exercise 1↴

Exercise 2

As long as the concentration of chemical XYZ in groundwater samples is in general less than 5.2 mg/L, we can assume that the potential for contamination in the groundwater is acceptable (i.e., not sufficient to cause illness in people who consume the water). We have found sufficient data on the chemical analyses of XYZ in groundwater to have a mean of 12 mg/L, a standard deviation of 9, and know that the data can be described by a Gaussian distribution.

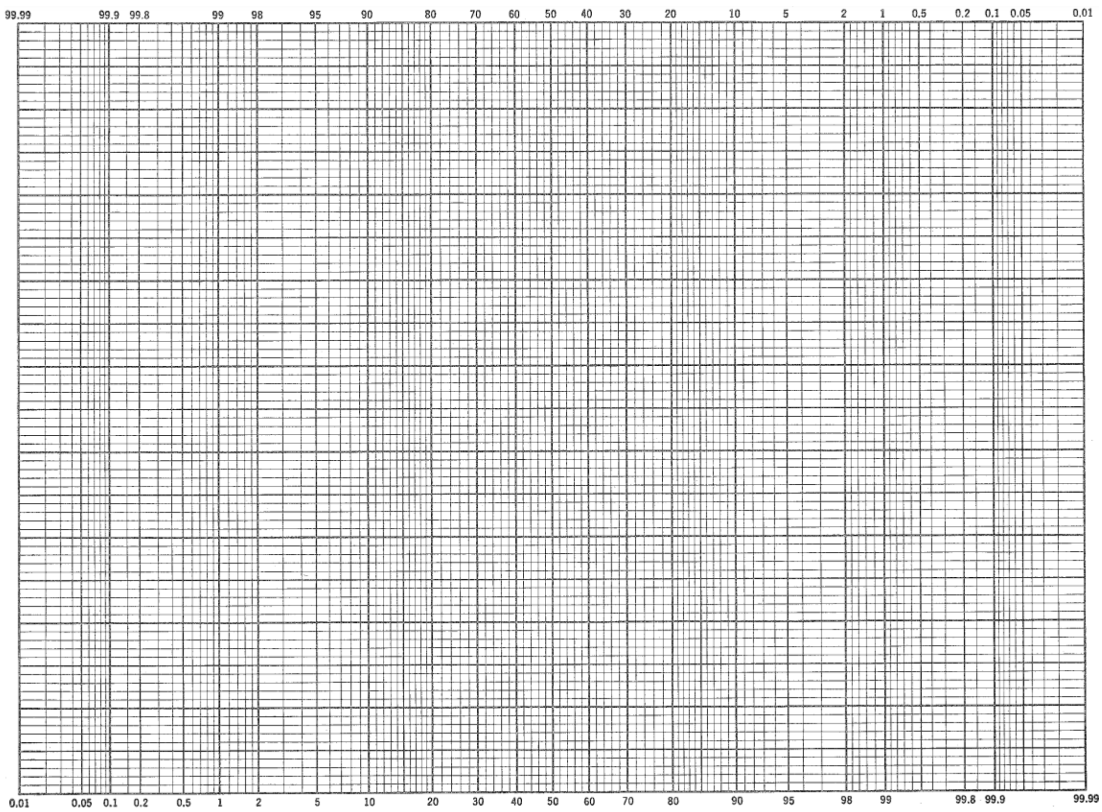
What is the probability that the groundwater exceeds the concentration of XYZ that endangers human health?

Click for Solution to Exercise 2↴

Exercise 3

A series of measurements of water quality data at a site are tabulated below (after McBean & Rovers, 1990, 1992). Assuming the data are normally distributed, estimate the mean and standard deviation of the data. Start by filling in the table below, then plot the values on probability paper. An image of arithmetic probability paper is provided after the table.

I Water quality data	II Rank	III Rank-ordered data	IV Plotting position
8.2			
5.3			
< 5			
< 5			
10.1			
9.3			
7.6			
< 5			
Use plotting position = m/(n + 1) used to plot the data			



Click for solution to Exercise 3

Exercise 4

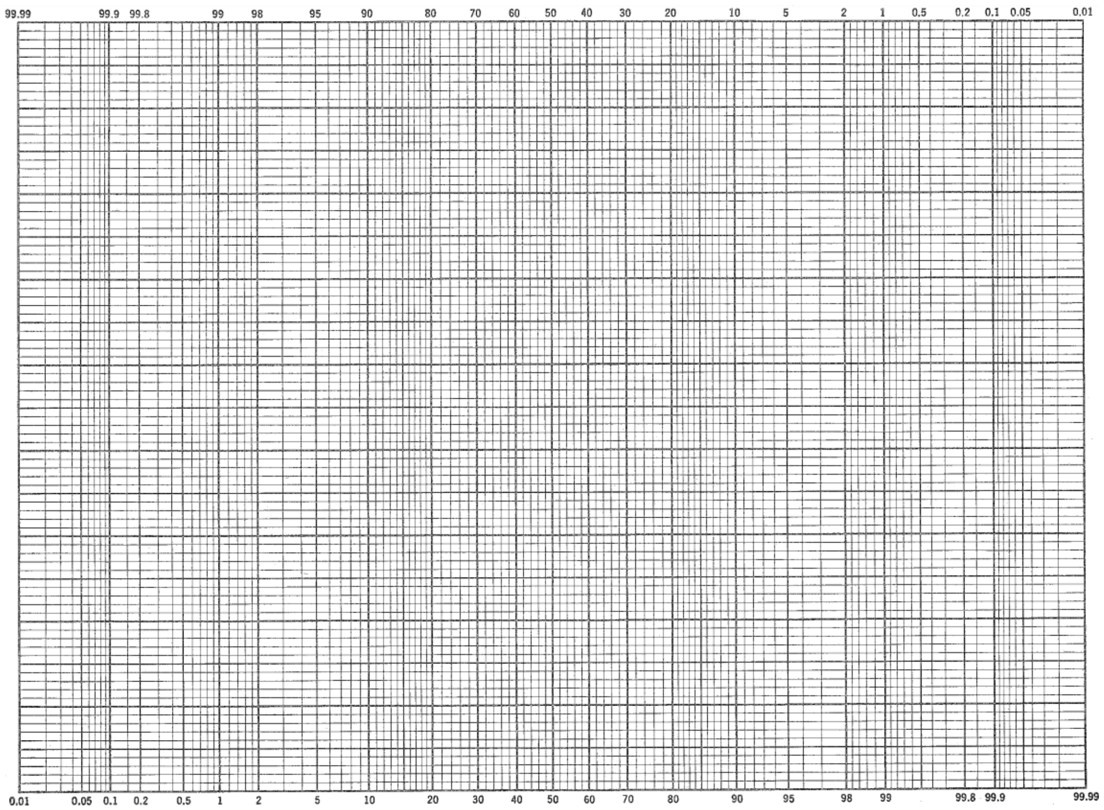
Seven data values of arsenic in groundwater (in μg/m^3) in a community water well near a gold mine, are available with statistics as shown here.

- 1.43, 19.3, 4.13, 37.6, 1.77, 1.01, 5.10
- Arithmetic Mean = 10.0
- Standard deviation of raw data = 13.7
- Geometric Mean = 1.51
- Standard deviation of log-transformed (natural logarithmic) data = 1.36

a. Assuming the arsenic concentrations are lognormally distributed, what is the probability that the arsenic concentration is larger than 45 μg/m^3?

b. Assume the data are characterized by the lognormal distribution and plot the data on probability paper (provided below). Why do the values not form a straight line when plotted on the probability paper?

c. Using the probability paper, and assuming the groundwater quality was sampled 100 times, what is the best estimate of the highest value that would occur in the 100 samples?



Click for Solution to Exercise 4

6 References

Barbalace, R. C. (2022). The chemistry of polychlorinated biphenyls. PCB The manmade chemicals that won't go away. *Environmental Chemistry.com*. https://environmentalchemistry.com/yogi/chemistry/pcb.html↗.

Beyer, W. H., (Ed.). (1966). *CRC handbook of tables for probability and statistics* (2nd ed.). John Wiley & Sons.

Dourson, M., & Gadagbui, B. (2021). The dilemma of perfluorooctanoate (PFOA) human half-life. *Regulatory Toxicology and Pharmacology: RTP*, *126*, 105025. https://doi.org/10.1016/j.yrtph.2021.105025↗.

Farrow, C., & McBean, E. (2016). Human health risk assessment: Arsenic exposure risks in Bangladesh. *Journal of Environmental Science and Engineering Technology*, *4*(1), 22–28. http://dx.doi.org/10.12974/2311-8741.2016.04.01.3↗.

Joseph, T., Dubey, B., & McBean, E. (2015a). A critical review of arsenic exposures for Bangladeshi adults. *Science of the Total Environment*, *527–528*, 540–551. https://doi.org/10.1016/j.scitotenv.2015.05.035↗.

Joseph, T., Dubey, B., & McBean, E. (2015b). Human health risk assessment from arsenic exposures in Bangladesh. *Science of the Total Environment*, *527–528*, 552–560. https://doi.org/10.1016/j.scitotenv.2015.05.053↗.

McBean, E. (2013). Improved human health risk characterization for regions with arsenic-contaminated groundwater. *Journal of Groundwater Research*, *2*/*2*, December.

McBean, E. (2019a). *Risk assessment: Procedures and protocols.* Wiley.

McBean, E. (2019b). Removal of emerging contaminants: The next water revolution. *Journal of Environmental Informatics Letters*, *1*(1), 1–7. http://jeiletters.org/index.php?journal=mys&page=article&op=viewFile&path[]=771&path[]=pdf_1_1↗.

McBean, E. A., & Rovers, F. A. (1985). Analysis of variances as determined from replicates versus successive samplings. *Groundwater Monitoring & Remediation*, *5*(3), 61–64. https://doi.org/10.1111/j.1745-6592.1985.tb00930.x↗.

McBean, E. A., & Rovers, F. A. (1990). Flexible selection of statistical discrimination tests for field-monitored data. In D. M. Nielsen & A. I. Johnson (Eds.), *Groundwater and vadose zone monitoring* (pp. 256–265). American Society for Testing and Materials Special Publication. https://doi.org/10.1520/STP23418S↗.

McBean, E., & Rovers, F. A. (1992). Estimation of the probability of exceedance of contaminant concentrations. *Groundwater Monitoring & Remediation*, *12*(1), 115–119. https://doi.org/10.1111/j.1745-6592.1992.tb00416.x↗.

Murray, G., McBean, E., & Sykes, J. (1995). Estimation of leakage rates through flexible membrane liners. *Groundwater Monitoring and Remediation*, *15*(4), 148–154. https://doi.org/10.1111/j.1745-6592.1995.tb00562.x↗.

Murray, G., McBean, E. & Sykes, J. (1996). Risk-based engineering design for a landfill leachate collection and liner system. *Groundwater Monitoring and Remediation*, *16(*2), 139–146. https://doi.org/10.1111/j.1745-6592.1996.tb00134.x↗.

Sharma, M., McBean, E., & Gowing, A. (2007). Bioconcentration of dioxins and furans in vegetation. *Water, Air and Soil Pollution*, *179*(1), 117–124. http://dx.doi.org/10.1007/s11270-006-9218-7↗.

Subramanian, R., White, P., Cogliano, J. (2006). Comparison of Cancer Slope Factors Using Different Statistical Approaches. *Risk Analysis, 26(3),* 825-30. https://doi.org/10.1111/j.1539-6924.2006.00769.x↗.

Unwin, J., Miner, R., Srevers, G., & McBean, E. (1985). Groundwater quality data analysis. *National Council for Air and Stream Improvement Technical Bulletin*, *462*. https://www.ncasi.org/resource/technical-bulletin-no-0462-groundwater-quality-data-analysis/↗.

Wilson, R. (1979). Analyzing the daily risks of life. *Technology Review*, *41*, 41–46.

7 Boxes

Box 1 Table of Areas Under the Normal Distribution Curve

The cumulative area under the normal distribution curve is calculated as shown here.

$$F(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}z^2}\, dz$$

where:

F = function of

z = variable of interest

e = Euler's number, 2.71828 (dimensionless)

Areas under the normal distribution curve.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0159	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3896	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4430	0.4441
1.6	0.4452	0.4463	0.4474	0.4485	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4762	0.4767
2.0	0.4772	0.4788	0.4783	0.4778	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4865	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4980	0.4980	0.4981
2.9	0.4981	0.4982	0.4983	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998	0.4998
4.0	0.499968									

Return to where text links to Box 1

Box 2 Table of Percentiles of Student's t-Distribution

Percentiles of student's t-distribution for a one-sided test are shown in the table below.

Percentiles of student's t-distribution (df = degrees of freedom) (for one-sided test) (Beyer, 1966).

F/df	0.60	0.75	0.90	0.95	0.975	0.990	0.995	0.9995
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	31.598
3	0.277	0.765	0.633	2.353	3.182	4.541	5.841	12.941
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	6.859
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	5.405
8	0.262	0.706	2.397	1.860	2.306	2.896	3.355	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.259	0.695	1.356	1.782	2.170	2.681	3.055	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.291

Return to where text links to Box 2↑

8 Exercise Solutions

Solution Exercise 1

The t-test allows calculations of the confidence levels of the mean when the estimate of the standard deviation, S, is known but the population standard deviation, σ, is not known. If σ is known and we assume a Gaussian or normal distribution for the data, 95 percent confidence limits for μ are given by the following equation.

$$\bar{\mathrm{x}} - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \bar{\mathrm{x}} + \frac{1.96\sigma}{\sqrt{n}}$$

where:

$\bar{x}$ = estimate of the mean

σ = standard deviation

n = number of values in the sample

μ = population mean

When σ is replaced by S, the only change needed to calculate 95 percent confidence limits for μ is to replace the number 1.96 in the equation by a quantity represented as $t_{\alpha/2}$ (where $\alpha/2$ is used to indicate a two-sided test, above and below the mean, as further described below). Thus, the equation is as shown here.

$$\bar{x} - \frac{t_{\alpha/2}S}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{t_{\alpha/2}S}{\sqrt{n}}$$

where:

α = specified level of significance

S = sample estimate of the standard deviation

The mean, $\bar{\mathrm{x}}$, is the sum of the values divided by the number of values. The standard deviation, S, is calculated by squaring the difference between each data point and the mean, summing the squared values, dividing by the number of data points, and taking the square root of that value. From the data provided, $\bar{x} = 1.60$ and $S = 0.94$.

For n = 8, the degrees of freedom are 8 – 1 = 7. Using the table provided in Box2↥, go to 7 degrees of freedom and, since the values in the table are for a one-sided test, for $\alpha = 5\%$ we use the column for 0.975 (i.e., half of 0.05 subtracted from 1). The table indicates the student's t-value, $t_{\alpha/2} = 2.365$. The confidence bounds are calculated as shown here.

95% confidence limit: $\mu = 1.60 \pm \frac{2.365(0.94)}{\sqrt{8}} = 1.60 \pm 0.79$ or between 0.81 and 2.39

Return to Exercise 1↥

Solution Exercise 2

For a Gaussian distribution, the probability that a value in the distribution is less than a specified value is calculated as shown here.

$$z = \frac{x - \bar{x}}{S}$$

where:

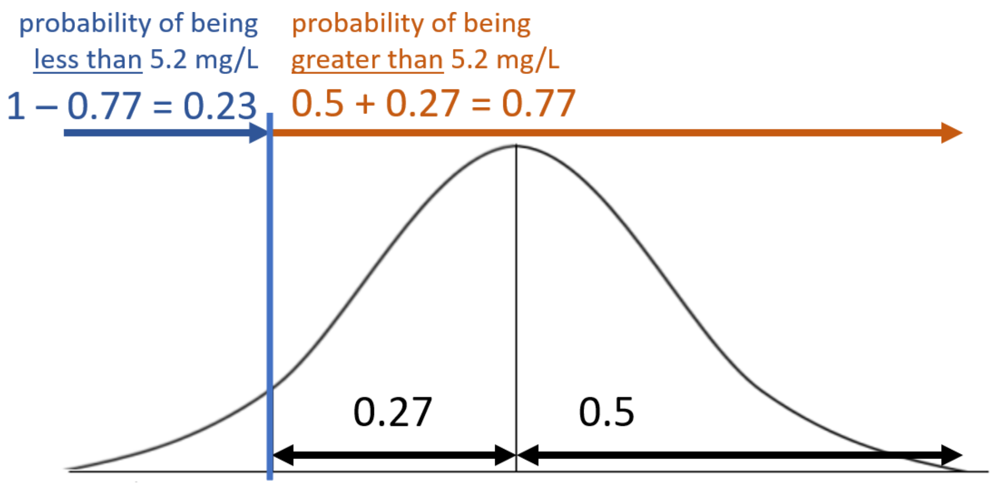
x = sample value
$\bar{x}$ = sample estimate of the mean
S = standard deviation of the sample values

So, the z value associated with measuring a value less than 5.2 mg/L is calculated as follows.

$$\frac{5.2 - 12}{9} = -0.75$$

Then, from the table of area under the normal distribution curve provided in Box 1 the z value for 0.75 is found by reading down to 0.7 in the first column and over to the column for 0.05 where the value is 0.2734, or in round numbers 0.27. This is a fraction of the entire area under the curve which is 1.

The position of 0.27 in the normal distribution can be envisioned as shown in the image below. It is measured to the left of the mean because the z value is negative. The probability that a sampled value for the chemical XYZ exceeds 5.2 mg/L is 0.5 + 0.27 = 0.77.



the z value of 0.75 corresponds to ~0.27 of the area under the Gaussian Distribution curve

Return to Exercise 2↥

Solution Exercise 3

Step 1

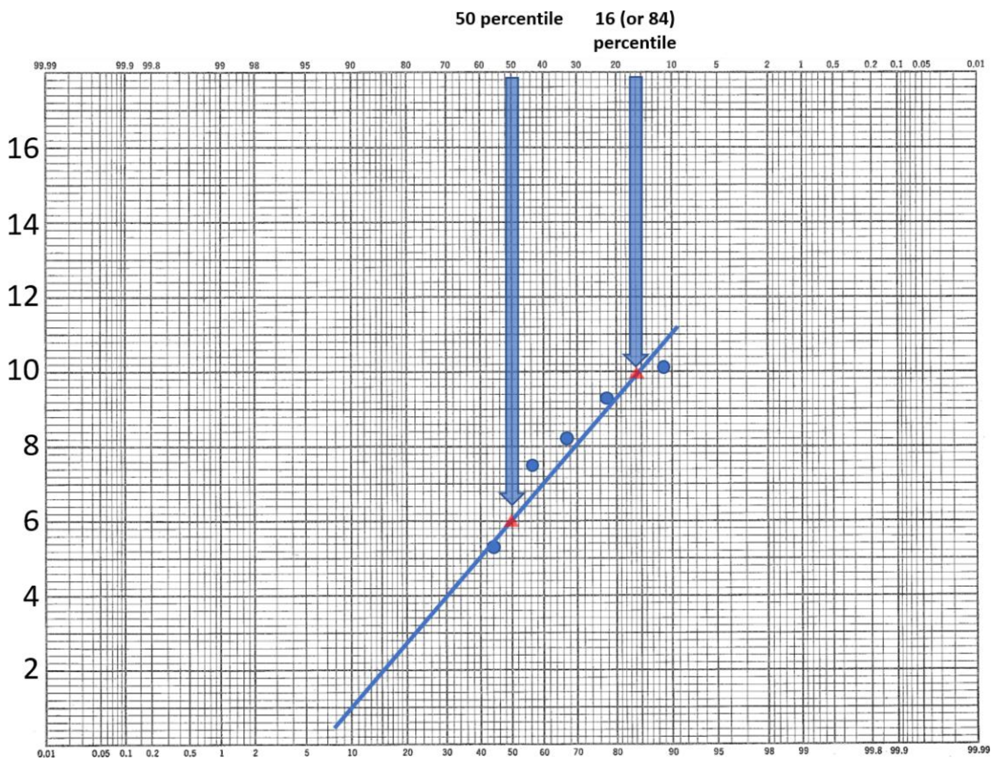
Rank order the detect data and, utilizing a plotting position formula such as the Weibull formula, then determine the plotting position as shown in the table below.

I Water quality data	II Rank	III Rank-ordered data	IV Plotting position
8.2	1	10.1	0.111
5.3	2	9.3	0.222
< 5	3	8.2	0.333
< 5	4	7.6	0.444
10.1	5	5.3	0.556
9.3	6	< 5	0.667 (censored data)
7.6	7	< 5	0.778 (censored data)
< 5	8	< 5	0.889 (censored data)

Use plotting position = m/(n + 1)

Step 2

Plot the values using the top axis as shown in the image below. Fit a straight line to the data points and read the mean and standard deviation from the graph. The mean concentration corresponds to the 50th percentile. Given that one standard deviation for the normal distribution corresponds to approximately 34 percent of the area under the normal distribution curve, the standard deviation corresponds to the difference between the 50th and the 16th percentile on the top axis (or the 84th and 50th percentile on the bottom axis), thus the mean = 6.0 and the standard deviation = 9.9 – 6.0 = 3.9 (red triangles).



Return to Exercise 3

Solution Exercise 4

a. First, calculate the natural log of the concentration values, then calculate their mean and standard deviation.

$$0.36, 2.96, 1.42, 3.63, 0.57, 0.01, 1.63 \rightarrow \text{mean} = 1.51, \text{std dev} = 1.36$$

Calculate the natural log of the value of interest (45).

Natural log of 45 → 3.81

As shown in the solution to Exercise 2, calculate the z value associated with the arsenic concentration being greater than 45 μg/m^3.

$$z = \frac{x - \bar{x}}{S}$$

where:

z = variable of interest with zero mean, unit standard deviation

x = sample value

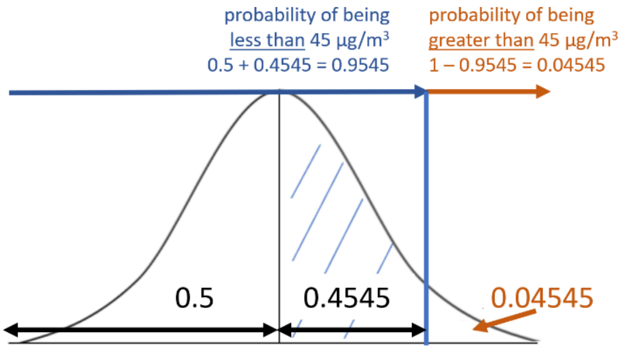
$\bar{x}$ = mean of sample values

S = standard deviation of the sample values

Substituting:

$$z = \frac{x - \bar{x}}{S} = \frac{3.81 - 1.51}{1.36} = 1.69$$

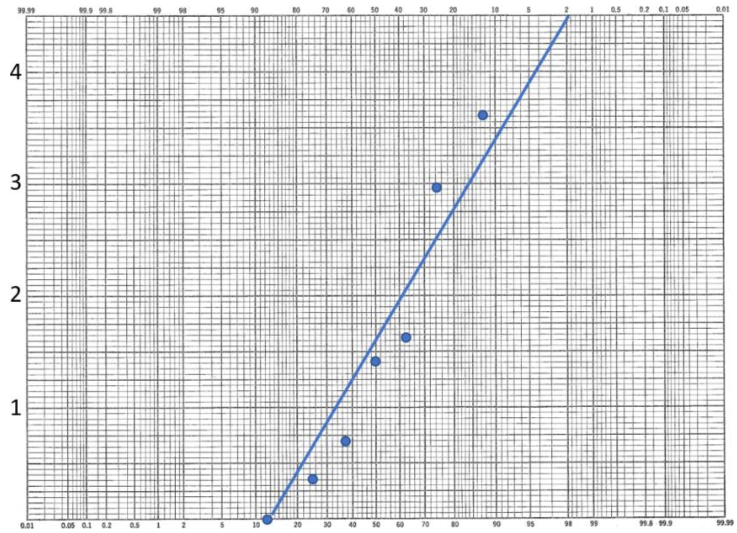
Then, from the table of area under the normal distribution curve provided in Box 1 the z value for 1.69 is found by reading down to the value of 1.6 in the first column and over to the column for 0.09 where the two-sided probability value is 0.4545. This is a fraction of the entire area under the curve which is 1. It is positive so is plotted to the left of the mean. Thus, the probability of arsenic being greater than 45 μg/m^3 is 4.55 % as shown in the image below.



the z value of 1.69 corresponds to 0.4545 of the area under the Gaussian Distribution curve

b. The transformed, ranked data with the associated Weibull plotting position and the graph are show in the images below. The data are not perfectly lognormal, rather it a log-normal representation is an approximation of the data distribution. because there is uncertainty associated with concentration variation depending on the sampling time, the field sampling process, and the laboratory analysis procedure.

Rank	Rank-ordered log transformed concentration data	Weibull plotting position
1	3.63	0.125
2	2.96	0.250
3	1.63	0.375
4	1.42	0.50
5	0.57	0.625
6	0.36	0.75
7	0.01	0.875



c. The rank of 1 of 100 values would be = m(n+1) or 1(101) = 1% so the concentration value where the fitted line crosses the 1% value provides an estimate of the expected maximum concentration. The fitted line could be shifted a bit up or down and the slope could be larger or smaller, but the projection of the fitted line to the 1 % value (i.e., 1/100) on the graph would likely fall around 5. Taking the exponential of 5 results in a concentration of 148 μg/m^3.

Another way to think of this is to use the table in Box 1 to find the z value for 0.495 which represents the 99 percentile (i.e., 1 in 100 samples) and solve for x, then log transform that value. The z value for 0.495 is 2.575. Rearranging the expression in part (a) to solve for x, then determining C for a probability of 0.01 is shown here.

$$z = \frac{x - \bar{x}}{S}$$

$$x = zS + \bar{x}$$

$$x = 2.575\ (1.36) + 1.51 = 5.012$$

$$C = e^{x} = e^{5.012} = 150.2\ \mu g/m^3$$

Return to Exercise 4↥

9 About the Author

After completing his Doctor of Philosophy at MIT (Massachusetts Institute of Technology) in 1973, Edward McBean did a post-doc at Cornell University, followed by several decades as a faculty member at the University of Waterloo, a decade as President of CRA Engineering Incorporated, and two decades as a Professor and Canada Research Chair in Water Supply Security and University Leadership Chair Professor, Water Security at the University of Guelph. Ed has been a recipient of a number of awards including being a Canadian Academy of Engineering Fellow, a Diplomate, Water Resources Engineering from American Academy of Water Resources Engineering and American Society of Civil Engineers, a European Union Academy of Sciences Fellow, a recipient of the American Academy of Water Resources Engineering and the American Society of Civil Engineers Outstanding Research and Innovation Award, a recipient of the Research and Development Medal from the Professional Engineers of Ontario, the lifetime achievement award prize from Ton Duc Tang University in Vietnam. Dr. McBean has published more than 400 articles in the technical literature.

Please consider signing up for the GW-Project mailing list to stay informed about new book releases, events, and ways to participate in the GW-Project. When you sign up for our email list it helps us build a global groundwater community. Sign up↗.

THE GROUNDWATER PROJECT