NRC Publications Archive Archives des publications du CNRC

Predicting risk of water quality failures in distribution networks under uncertainties using fault-tree analysis

Sadiq, R.; Saint-Martin, E.; Kleiner, Y.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1080/15730620802213504>

Urban Water, 5, 4, pp. 287-304, 2008-12-01

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=d745c4b6-b538-48e9-a2a5-c01ac5ea3593>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=d745c4b6-b538-48e9-a2a5-c01ac5ea3593>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

<http://irc.nrc-cnrc.gc.ca>

Predicting risk of water quality failures in distribution networks under uncertainties using fault-tree analysis

NRCC-50059

Sadiq, R.; Saint-Martin, E.; Kleiner, Y.

2008-11-13

A version of this document is published in / Une version de ce document se trouve dans:
Urban Water, v. 5, no. 4, December 2008, pp. 287-304

The material in this document is covered by the provisions of the Copyright Act, by Canadian laws, policies, regulations and international agreements. Such provisions serve to identify the information source and, in specific instances, to prohibit reproduction of materials without written permission. For more information visit <http://laws.justice.gc.ca/en/showtdm/cs/C-42>

Les renseignements dans ce document sont protégés par la Loi sur le droit d'auteur, par les lois, les politiques et les règlements du Canada et des accords internationaux. Ces dispositions permettent d'identifier la source de l'information et, dans certains cas, d'interdire la copie de documents sans permission écrite. Pour obtenir de plus amples renseignements : <http://lois.justice.gc.ca/fr/showtdm/cs/C-42>

National Research
Council Canada

Conseil national
de recherches Canada

Canada

Predicting risk of water quality failures in distribution networks under uncertainties using fault-tree analysis

Rehan Sadiq^{1*}, Elise Saint-Martin², and Yehuda Kleiner¹

Institute for Research in Construction of National Research Council (NRC-IRC), Ottawa, ON, Canada¹

Civil Engineering Determent, University of Ottawa, ON, Canada²

Abstract - The water quality in a distribution system is affected by many factors, including operational and environmental conditions in and around the distribution network. Lack of reliable data as well as knowledge gaps with respect to the impact of these factors on water quality make the quantification of water quality failure risk very challenging. Furthermore, the variability inherent in (sometimes) thousands of kilometers of distribution pipes presents added complexities. Major modes of water quality failures can be classified into intrusion of contaminants, regrowth of bacteria (biofilm), water treatment breakthrough, leaching of chemicals or corrosion products from system components, and permeation of organic compounds through plastic pipes. Deliberate contamination and negligence of operators have in recent years become an added concern.

In earlier works by Sadiq *et al.* (2004, 2007), an aggregative risk analysis approach using hierarchical structure was proposed to describe all possible mechanisms of contamination. In this paper a similar structure is used as a basis for a fault-tree approach. While fault-tree analysis is widely used for many engineering applications, in this paper we specifically explore how interdependencies among factors might impact analysis results. Two types of uncertainties are considered in the proposed analysis. The first is related to the *likelihood* of risk events, and the second is related to non-linear *dependencies* among risk events. Each basic risk event (input factor) is defined using a fuzzy probability (*likelihood*) to deal with its inherent uncertainty. The dependencies among risk events are explored using Frank copula and Frechet's limit. The proposed approach is demonstrated using two well-documented episodes of water quality failures in Canada, namely, Walkerton (ON) and North Battleford (SK).

Key Words. Distribution networks, fault-tree analysis, risk events, water quality failure, fuzzy probability, Frank copula, Walkerton, and North Battleford

* Corresponding author: <mailto:rehan.sadiq@nrc-cnrc.gc.ca>

1. INTRODUCTION

Water quality can be defined by a collection of upper and lower limits on selected performance indicators (Maier 1999). Therefore, a water quality failure (WQF) refers to an exceedance of one or more water quality indicators from specific regulations, or in the absence of regulations, exceedance of guidelines or self-imposed limits driven by customer service needs (Sadiq *et al.* 2004). Current and anticipated water quality regulations in Canada and the United States aim to minimize the impact of water quality failure in distribution systems. These regulations require utilities to continually upgrade and rehabilitate their distribution networks, water treatment processes and technologies, and to improve operations and/ or change source-waters.

The Bonn Charter of Safe Drinking Water (2004) discusses a generic framework for the effective management of water quality from source to tap that can be linked to World Health Organization (WHO) guidelines for drinking water. It emphasizes the need for proactive risk-based management of drinking water supplies. The main feature of the Charter is not only the end-of-pipe verification (regular water quality monitoring) implementation of management control systems throughout the system to assess risks at all points.

The quantification of risk of water quality failure in water distribution systems is a difficult task. Water distribution systems comprise many (sometimes thousands) of kilometers of pipes of different ages and various materials. Operational and environmental conditions are highly variable both temporally and spatially. Further, as pipes are mostly buried, limited data are available on their condition. Finally, some of the failure processes are not well understood and forensic investigation of contamination is very difficult because there is generally a time lag between the time of failure and the time at which the consequences (e.g. outbreaks) are observed.

Fault tree analysis is a commonly used tool to determine the cause(s) of system failure. The fault tree is constructed as a *tree of sub-events, spreading into bottom events, procreating the fault and finally heading to the top [or main] event* (Ming-Hung *et al.* 2006). A fault tree is a graphical representation of an event structure that clearly envisions the entire

system and each level of each event within it. This structure enables one to identify particular sub-events that have a high impact on the main event.

The difference between a general hierarchical model and a fault tree is that while the former simply display the order of events that lead to the main failure event (and aggregates them using any appropriate weighting scheme), the latter defines logical relationships ('*and*' and '*or*') between sub-events. Traditional fault tree analysis requires the assignment of crisp probabilities between events and the assumption of 'independence' between risk events. These two requirements carry inherent shortcomings. Firstly, one rarely knows 'precise' probabilities of causal relationships between events, and secondly, the required assumption of independence is often unrealistic.

Recently, Li (2007) proposed an object-oriented approach for water supply systems based on aggregative risk assessment (similar to Sadiq *et al.* 2004) and fuzzy fault tree analysis. Li (2007) used fuzzy evidential reasoning method to determine the risk levels associated with components, subsystems, and the overall water supply system. Fuzzy sets theory was used to evaluate the likelihood, severity, and risk levels associated with each hazard and Dempster-Shafer theory was used to aggregate the risk levels of multiple hazards throughout the hierarchical tree. Fuzzy fault tree analysis was used to evaluate the likelihood of the occurrence for a specific event in a water supply system. The dependencies in fault tree analysis were also studied using a 'dependency factor' similar to linear correlation.

The objective of this paper is to propose a generic fault tree approach to evaluate risk of water quality failure in a distribution network. The hierarchical structure of factors contributing to failure, presented in earlier work by Sadiq *et al.* (2004, 2007) is extended for the proposed fault tree structure. Further, the shortcomings discussed earlier, namely false precision in probability assignments and unrealistic independence assumptions, are addressed. Probabilities of basic risk events are defined linguistically using fuzzy numbers, and nonlinear interdependencies are handled using Frank copula (for known correlation) and Frechet's limit (for unknown correlation) (Ferson *et al.* 2004). The proposed fault tree is applied to two case studies to demonstrate the applicability of the approach.

2. WATER QUALITY FAILURES IN DISTRIBUTION SYSTEMS

The safety of drinking water is the supreme priority of water utilities. A typical modern water supply system comprises the water source (aquifer or surface water source including the catchment basin), transmission mains, treatment plant and distribution network, which includes pipes and distribution tanks. While water quality can be compromised at any point in the system, failure at the distribution level can be critical because it is closest to the point of use and, with the exception of a filtering device at the consumer level there are virtually no safety barriers before consumption.

The World Health Organization (WHO) has reported that infectious water borne diseases are the world's single largest source of human mortality. Although developed countries have a higher capability of controlling waterborne pathogens, water quality episodes still occur in North America and Europe (Edge *et al.* 2003). While biological agents are the most common source of major water quality failure, chemicals can also adversely affect drinking water quality. These chemicals can be categorized as organic and inorganic. Inorganic chemicals include arsenic, lead, and nitrate, to name a few. Organic chemicals like PAH (polynuclear aromatic hydrocarbon) and various DBPs (disinfection byproducts) can cause anemia, decreased level of blood platelets and organic chemicals like DBPs have a potential to increase cancer risk and miscarriages (AWWA 2007).

2.1 Water quality deterioration mechanisms

Water quality failures compromising either the safety or the aesthetics of water can occur due to various mechanisms and pathways throughout the distribution network. Contamination in distribution networks can generally be classified into the following major categories (Kleiner 1998):

- Intrusion of contaminants into the distribution system through system components whose integrity was compromised;
- material deterioration of system components through corrosion or leaching causing deterioration of distributed water quality;
- formation of disinfection byproducts (DBPs) and loss of disinfectant;

- compromised quality of water due to microbial (and chemical) breakthrough at the water treatment plant;
- permeation of organic compounds from the surrounding soil through plastic system components into the water supplies;
- biofilm regrowth in the distribution system; and
- deliberate contamination of the distribution system.

Intrusion of contaminants in the water distribution system can occur through pipes and storage tanks. Open finished water reservoirs are susceptible to microbial contamination from external non-point sources such as feces of infected animals or dead animals themselves (e.g., beaver, squirrels and rabbits) within the watershed (Kirmeyer *et al.* 2001). Intrusion of contaminants through water mains may occur during maintenance and repair events, through broken pipes and gaskets, and cross connections. A broken gasket, intended to seal pipe joints, can be a pathway for a variety of bacteria into the distribution network (Geldreich 1990). Regular maintenance and repair events as well as other anthropogenic and natural disasters may cause intrusion of contaminants in the water distribution network. Cross connections (an unprotected physical connection between a potable and a non-potable water system) can potentially introduce substances that may compromise the quality of the finished water (Kirmeyer *et al.* 2001).

The corrosion of metallic pipes and plumbing devices increases the concentration of metal compounds in the water. Metallic (e.g. cast iron) deterioration may often cause red water problems. In fact, red water is of the most commonly encountered water quality failure although the peril is a loss of aesthetics and damage to clothes rather than health issues. Contamination of water by compounds leached from pipe liners (plastic and epoxy lining) has also been observed (Sadiq *et al.* 2004).

Disinfection is the primary method to inactivate pathogens, however, . other concerns have been raised in the last three decades about the safety of disinfected water. Harmful DBPs are formed in the presence of natural organic matter and bromide (from the source) during chlorination.

Permeation is a phenomenon in which contaminants migrate through the wall of plastic pipes. Holsen *et al.* (1991) reported that most permeation events occur where plastic pipes are buried in soils that are contaminated with gasoline, diesel fuel, or solvents. Thompson and Jenkins (1987) also reported that polyethylene pipes are potentially susceptible to permeation of non-polar organic compounds. Similarly, all elastomeric and thermo-plastic materials are prone to permeation by hydrocarbons. In general, the risk of contamination through permeation is relatively small compared to other contamination mechanisms.

Biofilm is a deposit consisting of microorganisms, microbial products and detritus on the walls of pipes or tanks. Biological regrowth occurs when bacteria escape through the treatment plant and enter into the distribution system and subsequently rejuvenate and re-grow in storage tanks/ water mains.

At the point-of-entry (POE) to the distribution system, water quality can be compromised when treatment is insufficient to address poor raw water quality. This can happen when treatment units and/ or pumping equipment is ill-designed or functioning at suboptimal levels due to aging equipment, insufficiently trained staff, unusual loads due to unexpected environmental and anthropogenic conditions (e.g., heavy rainfall, winds and agricultural practices) or any combination thereof.

(Un)purposeful contamination refers to all forms of threats (terrorism, vandalism) and negligence, which may cause water contamination in the distribution system. Threat may also include pranks or some form of negligence (e.g. insufficient monitoring of the distribution system on the part of a staff member). These threats may vary due to vulnerabilities related to factors such as the geographical and physical conditions (e.g., closeness to the urban centers, topography etc.) of the source water, water treatment plant, and distribution system.

2.2 Recent episodes of water quality failures in Canada -

Episodes of contaminated supply of drinking water are more prominent in rural than in urban jurisdictions. Rural communities have less access to trained staff and state-of-the-art technologies to regulate and monitor their water treatment and distribution systems. Over time, these concerns may escalate and pose serious threats to human health. Two major

outbreaks related to water quality failures including Walkerton and North Battleford are discussed and analyzed in section 4.

3. MODELS AND METHODS

3.1 Fault tree analysis

To develop building blocks for the hierarchical framework of aggregative risk analysis, a Meta language was developed by Sadiq *et al.* (2004) and extended in Sadiq *et al.* (2007). In the hierarchical tree, each risk event (factor) is partitioned into its contributory sub-events, and each of these can be further partitioned into lower level contributory factors. The unit consisting of a risk event (parent) and its contributory factors (children) is called a “family” (Figure 1). A family consists of two generations and each of the children can be a parent of children of the next generation. An element with no children is called a “basic risk event (factor)”. In this paper, the terms “risk event”, “event” and “factor” are used interchangeably for all elements with offspring(s).

The junctions in the fault tree can be generally described by two logical operators, “and” and “or”. For “and” gate (*conjunction*) all children in a family are required to trigger a parent event. For example, contaminant intrusion in a distribution system requires - a pathway (break or leak), driving force (pressure gradient) and presence of contaminant. For “or” gate (*disjunction*) any one or more of the children can trigger the parent event. The fault tree thus describes the main event as a function of all sub-events using conjunctions and disjunctions.

The notation used for any risk event is $X_{i,j}^k$, where i is the ordinal number of risk event X in the current generation; j is the ordinal number of the parent (in the previous generation); and k is the generation order of X . The term “main risk event” is used for the top event of the fault tree (it has no parent and is denoted by $X_{1,0}^1$) and it indicates that it is the first and the only event of the first generation (Figure 1). In this analysis, the main event is the risk of water quality failure in distribution systems. For clarity, in the remainder of this paper, the notation for a basic risk event (an event with no children) will include an

apostrophe at the generation index, i.e., if event $X_{l,l}^3$ is a basic risk event, it will be denoted $X_{l,l}^{3'}$. Since every parent is preceded by a junction, J , the notation for a junction is $J_{i,j}^k$, where i, j, k indices are the same as for the subsequent parent (Figure 1). Junction $J_{i,j}^k$ represents a logical operator that can be either an “or” gate or “and” gate.

Figure 2 provides a full description of the proposed fault tree structure. As mentioned earlier, this structure is an extension of a hierarchical structure proposed in earlier work by Sadiq *et al.* (2004, 2007). The risk events (at generation 2) of the fault tree are water quality deterioration mechanisms including - water quality at POE, intrusion of contaminants, material deterioration, (un)purposeful contamination, disinfectant loss and THM formation, permeation, and biofilm formation/ regrowth. Table 1 lists the description of each risk event in the proposed fault tree.

In conventional fault tree analysis, basic risk events are assigned crisp probabilities; however, such probabilities are often hard to come by due to insufficient statistical data and knowledge. Consequently, such crisp probabilities may lead to ‘precise’ but unrealistic results. In this paper we propose to use fuzzy linguistic probabilities (p) instead. These enable uncertainties to propagate throughout the structure of the fault tree.

3.2 Fuzzy set theory

Fuzzy logic provides a language with syntax and semantics to translate qualitative knowledge into numerical reasoning. The term *computing with words* has been introduced by Zadeh (1996) to explain the notion of reasoning linguistically rather than with numerical quantities. When evaluating events related to complex systems, decision-makers, engineers, managers, regulators and other stake-holders often view risk in terms of linguistic variables like *very high, high, very low, low* etc. Fuzzy set theory is able to describe effectively these types of uncertainties (encompassing vagueness), and linguistic variables can be used to approximate reasoning and subsequently be manipulated to propagate the uncertainties throughout the decision process.

Fuzzy-based techniques are a generalized form of interval analysis. A fuzzy number describes the relationship between an uncertain quantity x and a membership function μ , which ranges between 0 and 1 (Zadeh 1965). Fuzzy-based techniques can help in addressing

deficiencies inherent in binary logic and are useful in propagating uncertainties through models. Any shape of a fuzzy number is possible, but the selected shape should be justified by available information. Generally, triangular fuzzy numbers (TFN) or trapezoidal fuzzy numbers (ZFN) are used for representing linguistic variables (Lee 1996). Let the *probability (likelihood) of occurrence (p)* be defined by the triangular fuzzy number TFN_p . Figure 3 describes an 11-granular scale to define the *likelihood (p)* for basic risk events (Lee 1996). This proposed scale includes 11 linguistic constants, namely, (1) *absolutely low*, (2) *extremely low*, (3) *quite low*, (4) *low*, (5) *mildly low*, (6) *medium*, (7) *mildly high*, (8) *high*, (9) *quite high*, (10) *extremely high* and (11) *absolutely high*. The objective of using this scale is to provide decision-makers more flexibility in expressing their linguistic notions of probabilities. Figure 3 also provides a graphical depiction of the relationships between the TFN_p and the associated memberships. In addition, a *nil* event (i.e. if some phenomenon is surely absent), the granular (*q*) is assigned a $TFN_p(0, 0, 0)$. For example, assume that for a given scenario there are only two types of pipe materials - plastic and metallic. If the water distribution system does not have plastic pipe, the *likelihood (p)* of a plastic pipe being present is *nil*, and can be assigned a $TFN_p(0, 0, 0)$. On the other extreme, the *likelihood (p)* is *absolute* for a metallic pipe being present and can be assigned a $TFN_p(1, 1, 1)$. Each of the basic risk events can be assigned a fuzzy *likelihood* using this 11-granular system. The top table in Figure 6 provides an example of fuzzy probabilities assigned to basic risk events.

3.3 Uncertainty in dependency relationships

As mentioned earlier, traditional fault tree analysis uses a default assumption of “independence” among the risk events to determine the joint probability (risk) of a parent event. This assumption simplifies the analysis, but may not be very realistic.

The relationship between risk events may be positively or negatively correlated (or independent). In the case of two independent events *X* and *Y*, the joint probability of their conjunction is a simply a product of their individual probabilities (Ferson *et al.* 2004), or

$$P(X \wedge Y) = x y \quad (1)$$

where $x = P(X)$ and $y = P(Y)$, and the symbol ' $\wedge$ ' denotes a conjunction. The joint probability of their disjunction is obtained using following equation:

$$P(X \vee Y) = 1 - (1 - x)(1 - y) = x + y - x y \quad (2)$$

where the symbol ' $\vee$ ' denotes a disjunction.

We denote the interdependence of two events by Pearson correlation $r \in [-1, 1]$, where $r = 0$ means *complete independence*, $r = +1$ means *perfect dependence* (consonant events, likely influenced by the same causes) and $r = -1$ means *opposite dependence* (disjoint events). The strongest positive dependence between two events ($r = +1$) is when one event ensures the occurrence of another. When two events are mutually exclusive ($r = -1$), one event ensures the absence of the other. Pearson correlation r can be used to assign a level of inter-dependencies between the two extremes *opposite* and *perfect* (Ferson *et al.* 2004). Figure 4 shows a possible graphical relationship between Pearson correlation and a linguistic dependence scale ranging from *opposite* to *perfect*.

A particular formulation for correlation is derived from the Frank copula (1979). To formulate the probability of a conjunction of two events using the Frank model of correlation, the following equation is used:

$$P(X \wedge Y) = and_{Frank}(x, y, r) = \begin{cases} \min(x, y) & \text{if } r = +1 \\ x y & \text{if } r = 0 \\ \max(x + y - 1, 0) & \text{if } r = -1 \\ \log_s [1 + (s^x - 1)(s^y - 1)/(s - 1)] & \text{otherwise} \end{cases} \quad (3)$$

where $s = \tan\left(\pi \frac{(1-r)}{4}\right)$. This is a continuous function and the limits are when $r = +1$, 0, or -1

as r tends to be these values respectively. Similarly, disjunction of two events using the Frank model is:

$$P(X \vee Y) = or_{Frank}(x, y, r) = \begin{cases} \max(x, y) & \text{if } r = +1 \\ 1 - (1 - x)(1 - y) & \text{if } r = 0 \\ \min(x + y, 1) & \text{if } r = -1 \\ 1 - \log_s [1 + (s^{1-x} - 1)(s^{1-y} - 1)/(s - 1)] & \text{otherwise} \end{cases} \quad (4)$$

There exist many different methods to express correlation (dependence) but the Frank model (copula) is the most common. It is selected to express dependencies among risk events in the proposed fault tree analysis. Although exact correlation (r) is rarely available, meaningful analysis can be performed even if the type (sign) of correlation is known e.g., ‘positive’ ($r \in [0, +1]$) or ‘negative’ ($r \in [0, -1]$). In the case of completely unknown correlation, the Frechet’s limit $r \in [-1, +1]$ can be used to represent the maximum uncertainty in dependence relationships. Figure 5 demonstrates various types of possible dependency relationships as a function of the correlation coefficient using Frank model (Ferson *et al.* 2004).

Figure 6 provides an example in which the sample fault tree depicted in Figure 1 is analyzed. Final risk values (R) are shown using trapezoidal fuzzy number (ZFN), and a summary of results is provided in the table of Figure 6. A total of 10 trials are performed, using different assumptions of dependencies. A few points should be noted. Trial 6 represents the highest uncertainty related to interdependencies among risk events, which results in a maximum uncertainty in the calculated risk. Similarly, Trial 1 represents the lowest uncertainty related to interdependencies among risk events, which results in the lowest uncertainty in the calculated risk. Positively dependent risk events result in higher risk and negatively dependent events tend to reduce the calculated risk. These last two points can be intuitively understood.

3.4 Interpretation of calculated risk

Two measures - *similarity measure* (S) and *entropy* (E) are employed to interpret the final results of the analysis. The similarity measure is an empirical value that indicates which of the 11 failure likelihoods, shown in Figure 3 has the closest resemblance to the calculated risk R . The similarity measure is calculated using the reciprocal of the absolute averaged distance between the calculated fuzzy number, and each of the 11 granulars q_i .

$$S = \text{MAX}(s_i), \text{ where } s_i = \frac{1}{[|d - a_i| + |e - b_i| + |f - b_i| + |g - c_i|]/4} \quad ; \quad i = 1, 2, \dots, 11 \quad (5)$$

where $[d, e, f, g]$ represents the calculated risk R as ZFN, and $[a_i, b_i, c_i, d_i]$ represents the granular q_i as ZFN (for TFN, $b_i = c_i$).

The calculated risk R is assigned a linguistic constant based on q_i , which corresponds to $\text{MAX}(s_i)$, i.e., the granular with the highest similarity. The table in Figure 6 provides similarities for all 10 trials.

Entropy E represents a spread or variance of the calculated risk R . Entropy E is simply described by a convex set of granulars q_i (linguistic constants), which intersect with the calculated risk R and can be mathematically written as:

$$E = \bigcup_{R \cap q_i \neq \emptyset} q_i \quad (6)$$

Entropy (E) describes the uncertainty in the final risk value and consists of a set of linguistic constants that R spans. For example, in trial 6, the final risk is (most similar to) *quite low* (QL), but can vary from *absolutely low* (AL) to *mildly low* (ML) (Figure 6).

It should be noted that in this paper we have not provided a detailed description of the fault tree analysis procedure. This is a well-established procedure that is provided in detail in many textbooks, e.g., see Vesely (1987). Rather we focused on the impact of interdependencies of contributing factors (events) on the final risk results using fault tree analysis procedure.

4. CASE STUDIES

4.1 Walkerton (Ontario)

Walkerton is a community of about 5,000 people within the Municipality of Brockton in Bruce County, northwest of Toronto. It is principally an agricultural area. In mid May 2000, Walkerton experienced one of the most detrimental episodes of water quality failure in recent years. More than 2300 people were affected by contaminated water originating from the water distribution systems in and around the Walkerton municipality. The affected population showed symptoms of gastrointestinal illness caused by *E. coli* (O157:H7). The bacteria exist in both human and animal intestines, and are usually derived from fecal matter (Zhao 2002).

Health officials issued a boiled water advisory once they started receiving reports from local doctors about bloody diarrhea cases. Water samples taken during the crisis were

tested by the Public Health laboratory and confirmed that the drinking water was contaminated with *E. coli*. Further, the water utility for Walkerton (Grey-Bruce Regional Health Unit) confirmed that the chlorinator had not been properly functional in the past for an undetermined period of time. By the end of May, seven deaths related directly to this bacterial outbreak were recorded. Medical experts declared that the Walkerton outbreak was the world's second worst instance of *E. coli* (O157:H7) contamination spread through drinking water. The forensic investigation into the incident found that cattle manure from a neighboring farm washed off during heavy rainfall and contaminated Well #5 of the Walkerton Public Utilities Commission (WPUC).

At the time of the incident, the town's drinking water supply came from a groundwater supply fed from three drilled wells, identified as well #5, #6, and #7 (Zhao 2002). Each well had a chlorination unit. The wells supplied water through a distribution system of about 42 kilometers in length (50-400 mm in diameter) and two system storage facilities.

The outbreak occurred due to a combination of factors; the principal contributing risk events are the *intrusion of contaminants* into the distribution system, *vulnerability* of the system, *inadequate treatment*, and *(un)purposeful contamination* (i.e., negligence and insufficient monitoring) due to negligence and disregard for public safety.

The intrusion of contaminants into the system was caused by the contamination of well #5 due to surface water runoff from the surrounding agricultural activity, which occurred during heavy rainfall (Zhao 2002). The surface water runoff became contaminated with *E. coli* bacteria originating from calf feces. Well #5 was nearer to the surrounding agricultural activity than well #6 and #7. Therefore, the *vulnerability* of well #5 in terms of physical and geographical conditions was relatively high.

The wells were insufficiently monitored and raw water samples were not taken at the required 10-days intervals (Woo and Vicente 2002). It was noted that chlorine residuals were not monitored daily as required and the residual monitoring records incorrectly showed residuals exactly of 0.5 mg/L and 0.75 mg/L (Hrudey *et al.* 2003). Poor laboratory results were concealed and modified (Woo and Vicente 2002). The inquiry revealed that chlorine treatment at well #5 was inadequate and consequently poorly disinfected water was entering the water distribution network.

The water treatment plant had improperly trained staff (as in many small communities), which contributed to poor water quality testing and monitoring as well as reporting of false data (Woo and Vicente 2002). Other possible causes included pipe installations, broken pipes that were later repaired, cross-connections at private wells, loss of pressure due to a fire event, and potential contamination of two water storage standpipes (Hrudey *et al.* 2003).

4.2 North Battleford (Saskatchewan)

In April 2001, North Battleford experienced an outbreak of *C. parvum* affecting a community of approximately 14,000 people (Office of Laboratory Security 2001). There were approximately 5800-7100 reported cases of diarrheal illness linked to this incident. A commission of inquiry stated that a major contributing factor was a high concentration of fecal matter found in the source water. The fecal matter originated from neighboring agricultural activity and was known to contain high concentrations of *C. parvum* (Woo and Vicente 2003).

A second possible contributing factor was identified as the location of the sewage treatment plant. The plant was located 3500m upstream of the raw water intake. Results from studies conducted by the Saskatchewan Environment and Resource Management (SERM) indicated that the geographical conditions were favorable to allow the sewage to migrate toward the water intake. It was reported, "...the topology of the area, along with river turbidity and flow caused the sewage effluent to reach the surface more often than not" (Woo and Vicente 2003). Consequently, the *C. parvum* outbreak may have occurred in part because of contamination of the raw surface water.

There were two water treatment facilities in North Battleford. The groundwater treatment system was comprised of chlorination and filtration, whereas surface water treatment system was comprised of flocculation, sedimentation, sand filtration, and chlorination. Consequently, a third possible source of contamination was identified as insufficient treatment related to the solids contact unit (SCU) of the surface water treatment plant. It can be concluded that it was a failure of multi-barriers to protect the drinking water supplies from contamination.

Woo and Vicente (2002) reported, “...the major contributing factor [at the technical and operational management level] dealt with the lack of education and a culture of non-compliance”. Further, there was inadequate treatment at the time of the outbreak. As mentioned earlier, small rural communities often do not have the best technological equipment and know-how. Consequently, the SCU of the surface water treatment plant was not functioning at the required level, causing poor settling (Stirling *et al.* 2001). There was evidence that floc formation in the SCU had not been established prior to the outbreak in April 2001 (Woo and Vicente 2002). This made it possible for *C. parvum* to pass into the finished drinking water. In addition, the treatment units were found to be in poor condition due to the aging of the surface water treatment plant. The Commission report stated, “...components of the surface water treatment were discovered to be either missing or offline prior to the outbreak” (Woo and Vicente 2002).

The subsequent analyses in this paper intend to illustrate the effects of the contributing risk events and their associated interdependencies on the overall risk of water contamination in distribution networks in both case studies.

4.3 Examination procedure

The risk events selected for inclusion in the two case studies for outbreak scenarios were *intrusion of contaminants*, *treatment adequacy*, *(un)purposeful contamination*, *vulnerability* of the system with respect to *physical/ geographical conditions*. In each case study, fuzzy *likelihood* values assigned to the basic risk events were fixed for all outbreak scenarios, but the *dependency* assumptions were varied to investigate the impacts of unknown dependencies on the risk analysis.

4.3.1 Baseline scenario

For each case study, a baseline scenario was established, which represents normal operating conditions (not considering the outbreaks). A baseline scenario sets a benchmark, against which three outbreak scenarios (per case study) were subsequently examined.

It should be noted that relatively large uncertainties (related to input values for basic risk factors) were anticipated in the baseline scenarios. Ironically, there was less uncertainty involved in the outbreak scenarios due to the rigorous investigations that ensued. Table 1

provides the fuzzy *likelihood* values (defined based on the 11-granular system) assigned to each basic risk event for baseline scenarios for the two case studies.

4.3.2 Interdependencies in baseline scenario

Interdependencies in baseline scenarios were assumed only for basic risk events as provided in Table 2. An independence assumption was made for all higher-level risk events. In addition for simplicity, interdependencies among non-relevant basic risk events were also assumed independent ($r = 0$).

4.3.3 Outbreak scenarios and interdependencies

Three outbreak scenarios were explored for each of the two case studies where various levels of *dependencies* were assigned to explore conditions at the time of each outbreak. The differences between these outbreak scenarios were related only to variations in the interdependency assumptions. Table 1 provides the fuzzy *likelihood* values (defined on the 11-granular system) assigned to each basic risk event for all the outbreak scenarios.

For example, based on a notion that there is a relationship between *vulnerability* and *threat*, *positive weak* dependency was assumed for ‘family’ *(un)purposeful contamination* (Figure 2). It can be argued that *vulnerability* of the system due to *insufficient monitoring* may possibly increase the *likelihood* of *threat* such as vandalism. Similarly, *vulnerability* due to *physical/ geographical conditions* of the system (e.g. proximity to industrial activity or hospital) may entice sabotage or other forms of *threats* to the system. Similarly, *positive mildly strong* dependency was assumed between the adequacy of *treatment units* and *untrained staff* (it is generally observed that small communities have limited resources for highly trained staff and they also lack sophisticated treatment). As mentioned earlier, the set of *dependencies* for the three outbreak scenarios are modified for risk events like contaminant intrusion, treatment adequacy, deliberate acts and the system vulnerability.

Scenario 1 (Sc1) assumed the same set of dependencies as that of the baseline scenario, i.e., two of the four risk events had some form of *positive* dependency, while for the other two *no information* was assumed. In Sc2, the set of *dependency* of all major risk events were assumed *positive strong*. Sc3 was investigated as a worst-case scenario where the least

amount of information, i.e., *no information* was provided for the *dependencies*. Table 3 provides a summary of the *dependencies* assumed for risk events of the fault tree for baseline and outbreak scenarios.

4.4 Risk analyses results

4.4.1 Walkerton

The results of all outbreak scenarios are superimposed graphically in Figure 7, to enable easy comparison with the baseline scenario. In Sc1, the calculated risk is [0.26, 0.51, 0.57, 0.78], clearly higher than the baseline risk of [0.02, 0.20, 0.20, 0.47]. This corresponds to a *medium* risk (based on the similarity measure as defined in section 3.2) of a water quality failure in the distribution system. This higher risk of Sc1 is due only to the basic risk event uncertainties, since Sc1 has the same set of dependencies as the baseline scenario. The entropy of Sc1 is *quite low to high*.

In Sc2, the entropy increases to *quite low to extremely high*. The increase in entropy suggests that the assumption of *positive* dependency increases the uncertainty significantly. The calculated risk is *mildly high* and ranges from 0.28 to 0.84. In this scenario, the resulting higher risk is due to a combination of the basic risk events uncertainties and the particular set of dependencies. In Sc3, complete ignorance is assumed about the relevant interdependencies, resulting in calculated risk described by a ZFN [0.23, 0.46, 0.65, 0.83]. The overall risk is *mildly high* and the total entropy ranges from *quite low to extremely high*. The table in Figure 7 provides a summary of risk estimates and associated entropy results for all scenarios of the Walkerton case study.

4.4.2 North Battleford

A similar set of dependencies is assumed for North Battleford. The input *likelihood* (i.e. for basic risk events) is different, however, based on the conditions specific to North Battleford present at the time of the outbreak. For instance, the estimated risk associated with treatment in Walkerton is *quite low* (based on similarity measure) while for North Battleford it is estimated to be *low*. Evidence during the inquiry of the outbreak in North Battleford revealed that inadequate treatment units of the water treatment plant contributed to the

increased risk. The units were practically obsolete or functioning at sub-optimal levels due to the age of the treatment plant.

Figure 8 illustrates the results of all scenarios in the North Battleford case study. In Sc1, the calculated risk R of water quality failure in the distribution systems is described by a ZFN [0.27, 0.42, 0.49, 0.70], which can be interpreted as *medium* risk. The entropy is found to be between *low* to *quite high*.

The entropy for Sc2 increases to *quite low* to *quite high*. This reinforces the notion that there is a slightly higher uncertainty associated with *positive* dependencies. The risk R remains *medium* and ranges from 0.30 to 0.71. Since the risk does not increase significantly from Sc1 as is observed in the Walkerton case study, it is assumed that in this case positive dependencies have only a limited impact on the calculated risk.

In Sc3, the risk R remains *medium* and is described by a ZFN [0.24, 0.35, 0.60, 0.76]. Entropy is *quite low* to *quite high*, as in Sc2. In this case, no significant difference in the risk R emerges when the assumptions for dependencies are changed. It appears that the risk events with relatively high *likelihood* overshadow other events, which results in a similar risk R . The table in Figure 8 contains a summary of the calculated risk R and associated entropy E for all simulated scenarios for the North Battleford.

5. CONCLUSIONS

Water in distribution networks can be contaminated via several pathways and the quantification and characterization of this risk is a complex process. Thousands of kilometers of pipes of different ages and materials, uncertain operational and environmental conditions, unavailability of reliable data, and lack of understanding of some factors and processes affecting pipe performance make predicting the risk of distribution system failure extremely challenging. These complexities also make risk analysis for a water distribution system highly uncertain.

Two types of uncertainties were explored using a fault tree analysis of a water quality failure in distribution systems. The first was uncertainty with respect to the basic risk events, and the second was uncertainty with respect to interdependencies among risk events. A range

of values was defined based on informed assumptions and empirical estimates. Interdependencies were varied in the analysis to understand the relationships among risk events and the impact of these assumptions on the final calculated risk. The input values of the events were represented by fuzzy probabilities (*likelihood*). Frank copula and Frechet's limit were used to describe uncertainties in interdependencies among risk events. The interpretation of risk analysis results was carried out using *similarity* and *entropy* measures.

The proposed approach was applied to two case studies - Walkerton (Ontario) and North Battleford (Saskatchewan). Factors contributing to water quality failure were assumed to be comprised of water quality at POE, intrusion of contaminants into the distribution system, material deterioration, (un)purposeful contamination, and water quality deterioration related to disinfection, and biofilm regrowth. In the case of Walkerton, the calculated risk was at the most *mildly high*, but entropy was *quite low* to *extremely high* in the three outbreak scenarios. In North Battleford, similar observations were made, though the calculated risk remained *medium* for all three scenarios of outbreak.

The calculated risk values, R , for the different outbreak scenarios in the two case studies were mostly *medium*. Overall, the lower estimates of R were primarily attributed to smaller values of the fuzzy *likelihood* assumed for the basic risk events. Both water quality failure events were primarily linked to the 'point of entry' contamination, therefore comparatively higher likelihood values of basic risk events were assumed in that category. There was not enough evidence to make assumptions of higher likelihood values for other basic risk events. Even though events have already happened and proved to culminate in major outbreaks, this does not imply that the risk analysis results obtained in this study are not valuable. Rather, this analysis further enhances the importance and underscores the need for more comprehensive risk analysis with consideration of different types of uncertainties including a thorough investigation of the structure of the fault tree and the assumed logical relationships among various risk events as proposed in this paper.

Note that the calculated risk values, R , represent only the most likely estimates and should be interpreted with entropy (E) results, which account for uncertainties. The uncertainties were considerably large in both case studies. For example in Scenario 3, the estimated risk was as high as *Extremely High* and *Quite High* for Walkerton and North Battleford case studies, respectively. Risk-based decision-making is a process of identifying necessary and appropriate actions. The proposed

risk-based methodology provides an effective tool for water utilities to make informed decisions. It offers a scientifically rigorous and administratively effective way to respond to the pressures in a timely way. Therefore, even the calculated risk were mostly *medium*, the conservative estimates were alarming, and required swift actions from the water utilities. Further, it is important to recognize that risk-based decision-making should not intended to be just a cost-saving tool, rather should be used to prioritize the rehabilitation and corrective actions required.”

A major advantage of the proposed methodology is that it can accommodate both qualitative and quantitative data. Some data may be supported by rigorous observations, while other data may be based on loosely supported or anecdotal-based beliefs. These two types of data should have different weights in the aggregation and uncertainty propagation process. The approach can also accommodate and process data obtained from sources of different reliabilities. These two aspects are not discussed in this paper and are the subject of future research.

6. REFERENCES

- American Water Works Association (AWWA) 2007. *Chlorine Disinfectants/Disinfectant By-Products (DBPs)*, <http://www.awwa.org/Advocacy/pressroom/dbps.cfm> (viewed on May 8, 2007).
- Edge, T., Byrne, J.M., Johnson, R., Robertson, W. and Stevenson, R. 2003. *Waterborne Pathogens*, National Water Research Institute, Environment Canada.
- Ferson, S., Hajagos, J., Berleant, D., and Zhang, J. 2004. *Dependence in Dempster-Shafer Theory and Probability Bounds Analysis*, Sandia National Laboratories, US.
- Frank, M. 1979. On the Simultaneous Associativity of $F(x, y)$ and $x + y - F(x, y)$, *Aequationes Mathematicae* 19: 194-226.
- Geldreich, E.E. 1990. *Microbiological Quality Control in Distribution Systems*, In *Water Quality and Treatment* (ed. W.P. Frederick), American Water Works Association, McGraw-Hill, New York.
- Holson, T.M., Park, J.K., Jenkins, D. and Shelleck, R.E. 1991. *Contamination of Potable Water by Permeation of Plastic Pipe*, *Journal of American Water Works Association*, 83(8): 53-56.
- Hrudey, S.E., Huck, P.M., Payment, P., Gillham, R.W., and Hrudey, E.J. 2003. *A Fatal Waterborne Disease Epidemic in Walkerton, Ontario: Comparison with Other Waterborne Outbreaks in the Developed World*, *Water Science and Technology*, 47(3): 7-14.
- Kirmeyer, G.J., Friedman, M., Martel, K., and Howie, D. 2001. *Pathogen Intrusion into Distribution System*, AWWARF, Denver, Colorado.
- Kleiner, Y. 1998. *Risk Factors in Water Distribution Systems*, British-Columbia Water and Waste Association 26th Annual Conference, Whistler, British-Columbia, 26-29 April 1998.
- Lee, H.-M. 1996. *Applying Fuzzy Set Theory to Evaluate the Rate of Aggregative Risk in Software Development*, *Fuzzy Sets and Systems*, 79: 323-336.
- Li, H. 2007. *Hierarchical Risk Assessment of Water Supply Systems*, Ph.D. thesis, Loughborough University, Leicestershire, UK
- Maier, S.H. 1999. *Modeling Water Quality for Water Distribution Systems*, Ph.D. thesis, Brunel University, Uxbridge, UK.
- Ming-Hung, S., Ching-Hsue C., and Jing-Rong C. 2006. *Using Intuitionistic Fuzzy Sets for Fault-Tree Analysis on Printed Circuit Board Assembly*, *Microelectronics Reliability*, 46(12): 2139-2148.
- Office of Laboratory Security 2001. *Material Safety Data Sheet – Infectious Substances*, Public Health Agency of Canada, 2001 (<http://www.phac-aspc.gc.ca/msds-ftss/index.html>, viewed on May 7 2007).
- Sadiq, R., Kleiner, Y., Rajani, B., 2004, *Aggregative Risk Analysis for Water Quality Failure in Distribution Networks*, *Journal of Water Supply Research and Technology: Aqua*, 53(4): 241-261.

- Sadiq, R., Kleiner, Y., Rajani, B., 2007, *Water Quality Failures in Distribution Networks - Risk Analysis using Fuzzy Logic and Evidential Reasoning*, Risk Analysis- An International Journal, 27(5): 1381-1394.
- Stirling, M.D., Aramini, J. Ellis, Lim, G., Meyers, R., Fleury, M., and Werker, D. 2001. *Waterborne Cryptosporidiosis Outbreak North Battleford, Saskatchewan, Spring 2001*, Public Health Agency of Canada (PHAC), Canada Communicable Disease Report, Vol. 27-22, 15 November 2001 (<http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/01vol27/dr2722ea.html>, viewed on May 7 2007).
- The Bonn Charter of Safe Drinking Water 2004. *International Water Association (IWA)*, September 2004, pp. 18.
- Thompson, C. & Jenkins, P. 1987. *Review of Water Industry Plastic Pipe Practices*, AWWARF, Denver, Colorado.
- Vesely, W.E. 1987. *Fault Tree Handbook*, Nuclear Regulatory Commission, pp. 218.
- Woo and Vicente, D. and Vicente, K.J. 2003. *Socio-technical Systems, Risk Management, and Public Health: Comparing the North Battleford and Walkerton Outbreaks*, Reliability Engineering & Safety Systems, 80(3): 253-269.
- Zadeh, L.A. 1965. Fuzzy sets, *Information and Control*, 8 (3): 338-353.
- Zadeh, L.A. 1996. *Fuzzy Logic Computing with Words*, IEEE Transactions of Fuzzy Systems, 4(2): 103-111.
- Zhao, Z. 2002. *Risk Analysis of the Walkerton Drinking Water Contamination Tragedy*, Master's Thesis, University of Waterloo, Ontario, Canada.

Family	Junction	Corresponding risk events
2	OR ($J_{1,1}^2$)	$X_{1,1}^{3'}$ $X_{2,1}^{3'}$
3	AND ($J_{2,1}^2$)	$X_{3,2}^{3'}$ $X_{4,2}^{3'}$ $X_{5,2}^{3'}$
1	AND ($J_{1,0}^1$)	$X_{1,1}^2$ $X_{2,1}^2$

Figure 1. Building blocks of hierarchical framework to perform risk analysis

Figure 2. Proposed fault tree for water quality failure in a distribution network (Contd.)

Granulars (q^*)	Likelihood (p)	Triangular fuzzy number (TFN_p)
1	<i>Absolutely low (AL)</i>	(0, 0, 0.1)
2	<i>Extremely low (EL)</i>	(0, 0.1, 0.2)
3	<i>Quite low (QL)</i>	(0.1, 0.2, 0.3)
4	<i>Low (L)</i>	(0.2, 0.3, 0.4)
5	<i>Mildly low (ML)</i>	(0.3, 0.4, 0.5)
6	<i>Medium (M)</i>	(0.4, 0.5, 0.6)
7	<i>Mildly High (MH)</i>	(0.5, 0.6, 0.7)
8	<i>High (H)</i>	(0.6, 0.7, 0.8)
9	<i>Quite High (QH)</i>	(0.7, 0.8, 0.9)
10	<i>Extremely High (EH)</i>	(0.8, 0.9, 1.0)
11	<i>Absolutely High (AH)</i>	(0.9, 1.0, 1.0)

* q is assigned a value of (0, 0, 0), in case of complete absence (i.e., p is *nil*)

q is assigned a value of (1, 1, 1), in case of absolute presence (i.e., p is *absolute*)

Figure 3. A scale of fuzzy *likelihood* of a failure (after Sadiq *et al.* 2004)

Figure 4. Correlation and dependency relationship

Figure 5. Uncertainty in dependencies represented as a function of correlation (r)

Basic Risk events	Crisp <i>likelihood</i>	Fuzzy <i>likelihood</i>
$X_{1,l}^{3'}$	0.1	(0, 0.1, 0.2) (EL)
$X_{2,l}^{3'}$	0.2	(0.1, 0.2, 0.3) (QL)
$X_{3,l}^{3'}$	0.6	(0.5, 0.6, 0.7) (MH)
$X_{4,l}^{3'}$	0.4	(0.3, 0.4, 0.5) (ML)
$X_{5,l}^{3'}$	0.2	(0.1, 0.2, 0.3) (QL)

Trials	$J_{1,l}^2$	$J_{2,l}^2$	$J_{1,0}^1$	Calculated risk	Similarity	Entropy
	OR	AND	AND	R	S	E
1	Independent	Independent	Independent	(0.002, 0.013, 0.046)	AL	AL to EL
2	(+) Strong	(+) Strong	(+) Strong	(0.009, 0.05, 0.115, 0.214)	EL	AL to L
3	Positive	Positive	Positive	(0.002, 0.01, 0.2, 0.3)	QL	AL to ML
4	Unknown	Positive	Positive	(0.002, 0.01, 0.2, 0.3)	QL	AL to ML
5	Unknown	Unknown	Positive	(0, 0, 0.2, 0.3)	QL	AL to ML
6	Unknown	Unknown	Unknown	(0, 0, 0.2, 0.3)	QL	AL to ML
7	Negative	Negative	Negative	(0, 0, 0.014, 0.053)	AL	AL to EL
8	Unknown	Negative	Negative	(0, 0, 0.014, 0.053)	AL	AL to ML
9	Unknown	Unknown	Negative	(0, 0, 0.06, 0.15)	AL	AL to QL
10	Unknown	Negative	Positive	(0, 0, 0.048, 0.105)	AL	AL to QL

Figure 6. Risk analysis results for an example

Scenarios	Calculated risk (R)	Linguistic value assigned to R	E (Entropy) of risk estimates
Sc1	[0.2608,0.5129,0.5666,0.7818]	<i>Medium</i>	<i>Quite Low - Quite High</i>
Sc2	[0.2796,0.5390,0.5851,0.7916]	<i>Mildly High</i>	<i>Quite Low - Quite High</i>
Sc3	[0.2265,0.4583,0.6536,0.8294]	<i>Mildly High</i>	<i>Quite Low - Extremely High</i>

Note: In the case of equal similarity for two granulars, the largest is selected to be conservative

Figure 7. Risk analysis results for Walkerton (ON)

Scenarios	Calculated risk (R)	Linguistic value assigned to R	E (Entropy) of risk estimates
Sc1	[0.2726,0.4210,0.4896,0.6993]	<i>Medium</i>	<i>Quite Low - High</i>
Sc2	[0.2987,0.4600,0.5140,0.7107]	<i>Medium</i>	<i>Quite Low - Quite High</i>
Sc3	[0.2400,0.3526,0.5966,0.7587]	<i>Medium</i>	<i>Quite Low - Quite High</i>

Note: In the case of equal similarity for two granulars, the largest is selected to be conservative

Figure 8. Risk analysis results for North Battleford (SK)

Table 1. Description of all risk events and fuzzy *likelihood* values (q_i) assumed for basic risk events

Event	Name	Description of risk event	Walkerton		North Battleford	
			Baseline	Outbreak	Baseline	Outbreak
$X_{1,0}^1$	Water quality failure	Overall risk evaluation				
$X_{1,1}^2$	WQ at POE	Risk at the point-of-entry to the distribution system				
$X_{2,1}^2$	Intrusion of contaminants	Susceptibility of contaminants intrusion into the distribution system				
$X_{3,1}^2$	Material deterioration	Potential material deterioration causing leaching and corrosion				
$X_{4,1}^2$	(un)purposeful contamination	Vandalism, terrorism, also includes include negligence in the form of insufficient monitoring				
$X_{5,1}^2$	Disinfection related	Potential disinfectant loss and DBPs formation				
$X_{6,1}^2$	Permeation	Permeation of contaminants through plastic material				
$X_{7,1}^2$	Biofilm growth	Potential growth of biological products and microorganisms				
$X_{1,1}^3$	Treatment adequacy	Potential for inadequate treatment				
$X_{2,1}^{3'}$	Source water	Raw water contamination potential	3	5	3	6
$X_{3,2}^3$	Pathway	Possible contaminants pathways				
$X_{4,2}^{3'}$	Loss of pressure	Likelihood of pressure loss	2	2	2	2
$X_{5,2}^{3'}$	Presence of contaminants	Likelihood of contaminants' presence	1	4	1	4
$X_{5,3}^3$	Leaching	Potential for leaching from lining materials				
$X_{6,3}^3$	Internal corrosion	Potential of corrosion of metallic distribution pipes/plumbing devices				
$X_{7,4}^3$	Vulnerability	Vulnerability of the distribution system against malicious attacks				
$X_{8,4}^{3'}$	Threat	Direct and indirect threats to the distribution system	2	2	2	2
$X_{9,5}^3$	DBP Formation	Potential for formation of disinfection by-products				
$X_{10,5}^3$	Disinfection loss	Potential for disinfectant loss in the distribution system				
$X_{11,6}^{3'}$	Plastic	Plastic material present or not	1	1	1	1
$X_{12,6}^{3'}$	Hydro carbon	Hydro carbons are present or not around the pipe	1	1	1	1
$X_{13,6}^{3'}$	Conds. conducive to chemical permeation	Exposure of hydrocarbons to plastic material over long time	1	1	1	1
$X_{18,7}^{3'}$	Organic matter	Presence of optimal organic matter for the growth of biofilm	1	2	1	1
$X_{19,7}^{3'}$	Detention time	The time water stays in the system (water age)	2	3	2	2
$X_{20,7}^{3'}$	Conds. conducive to biofilm formation	Temperature, nutrients etc.	2	3	2	2
$X_{1,1}^{4'}$	Treatment units	Conditions of the treatment units in the plant before water enters the distribution system.	3	7	4	7
$X_{2,1}^{4'}$	Untrained staff	Lack of knowledge, experience and education	4	7	4	8

Event	Name	Description of risk event	Walkerton		North Battleford	
			Baseline	Outbreak	Baseline	Outbreak
$X_{3,3}^{4'}$	Broken pipes & gaskets	Risk of water quality failure due to broken pipes and gaskets throughout the water distribution system	2	4	2	2
$X_{4,3}^{4'}$	Maintenance and repair events	Risk of water quality failure due to the intrusion of contaminants during maintenance and repair of pipes and other components of the distribution system.	2	4	2	2
$X_{5,3}^{4'}$	Cross-connections	Risk of water quality failure due to the intrusion of contaminants at the cross-connections of the distribution system.	2	2	2	2
$X_{6,5}^{4'}$	Metallic surface	Risk of water quality failure due to leaching in the absence of a metallic surface.	11	11	1	1
$X_{7,5}^{4'}$	Conds. conducive to leaching	Conditions conducive to leaching leading to risk of water quality failure.	1	2	2	2
$X_{8,6}^{4'}$	Metallic surface	Risk of water quality failure due to corrosion in the presence of a metallic surface.	11	11	1	1
$X_{9,6}^{4'}$	Conds. conducive to corrosion	Risk of water quality failure due to corrosion likelihood of conditions conducive to corrosion.	2	3	1	1
$X_{10,7}^{4'}$	Physical & Geo. vulnerability	Risk of water quality failure due to the surrounding physical and geographical vulnerability of the system.	2	2	1	3
$X_{11,7}^{4'}$	Insufficient monitoring	Risk of water quality failure due to insufficient monitoring in a distribution system.	3	4	3	5
$X_{12,9}^{4'}$	Organic matter	Risk of water quality failure due to DBP formation likelihood of organic matter.	1	2	1	2
$X_{13,9}^{4'}$	Concentration of residual disinfectant	Risk of water quality failure due to DBP formation under likelihood of residual disinfectant.	2	3	2	2
$X_{14,9}^{4'}$	Conds. conducive to DBP formation	Risk of water quality failure due to DBP formation under likelihood of conducive conditions.	2	3	2	3
$X_{15,10}^{4'}$	Organic matter	Risk of water quality failure related to disinfectant loss under likelihood of organic matter.	1	2	1	2
$X_{16,10}^{4'}$	Concentration of residual disinfectant	Risk of water quality failure related to disinfectant loss under likelihood of residual disinfectant.	2	3	2	2
$X_{17,10}^{4'}$	Conds. conducive to disinfectant loss	Risk of water quality failure related to disinfectant loss under likelihood of conducive conditions.	2	3	2	3

*Shaded boxes represent the risk events that have children. The fuzzy *likelihood* of these risk events is calculated while propagating uncertainties through the fault tree.

** The number assigned for basic risk events are based on the 11-granular system and represents the fuzzy likelihood as shown in Figure 3.

Table 2. Assumed interdependencies among risk events for ‘baseline’ scenario for both case studies

Junction	Description	Gate	Dependence	Justification
$J_{1,0}^1$	Risk of water quality failure	OR	Independent	No correlation between second-generation risk events.
$J_{1,1}^2$	Water Quality at POE	AND	Independent	No correlation between treatment adequacy and source water
$J_{2,1}^2$	Intrusion of contaminants	AND	Independent	No correlation between third generation risk events.
$J_{3,1}^2$	Material deterioration	OR	Independent	No correlation between leaching and corrosion.
$J_{4,1}^2$	(un)purposeful contamination	AND	Positive Weak	Vulnerability of the system encourages threat.
$J_{5,1}^2$	Disinfection	OR	Positive M. Strong	Low value correlation between DBP formation and disinfectant loss (inversely proportional).
$J_{6,1}^2$	Permeation	AND	Independent	Third generation events are independent of each other, e.g., detention time does not affect temperature.
$J_{7,1}^2$	Biofilm growth	AND	Independent	No correlation between third generation risk events.
$J_{1,1}^{3'}$	Treatment adequacy	AND	Positive M. Strong	Poor treatment units because of probability of untrained staff in smaller communities.
$J_{3,2}^3$	Pathway	OR	Independent	Third generation events are dependant of each other. Maintenance and repair events increase the probability of broken pipes and gaskets.
$J_{5,3}^3$	Leaching	AND	Independent	No correlation between third generation risk events for leaching.
$J_{6,3}^3$	Internal corrosion	AND	Independent	No correlation between third generation risk events.
$J_{7,4}^3$	Vulnerability	OR	Independent	No correlation between physical/geographical vulnerability of the system and insufficient monitoring.
$J_{9,5}^3$	DBP formation	AND	Independent	No correlation between third generation risk events.
$J_{10,5}^3$	Disinfectant loss	AND	Independent	No correlation between third generation risk events.

Note: Highlighted events are the identified as major risk events.

Table 3. Summary of interdependencies assumptions among risk events for both case studies

Junction	Description	Gate	Dependency assumptions			
			Baseline	Sc 1	Sc 2	Sc 3
$J_{1,0}^1$	Risk of water quality failure	OR	Independent	Independent	Independent	Independent
$J_{1,1}^2$	Water Quality at POE	AND	Independent	Independent	Independent	Independent
$J_{2,1}^2$	Intrusion of contaminants	AND	No Info	No Info	Positive Strong	No Info
$J_{3,1}^2$	Material deterioration	OR	Independent	Independent	Independent	Independent
$J_{4,1}^2$	(un)purposeful contamination	AND	Positive M. Strong	Positive M. Strong	Positive Strong	No Info
$J_{5,1}^2$	Disinfection	OR	Positive M. Strong	Positive M. Strong	Positive M. Strong	Positive M. Strong
$J_{6,1}^2$	Permeation	AND	Independent	Independent	Independent	Independent
$J_{7,1}^2$	Biofilm growth	AND	Independent	Independent	Independent	Independent
$J_{1,1}^{3'}$	Treatment adequacy	AND	Positive V. Strong	Positive V. Strong	Positive Strong	No Info
$J_{3,2}^3$	Pathway	OR	Independent	Independent	Independent	Independent
$J_{5,3}^3$	Leaching	AND	Independent	Independent	Independent	Independent
$X_{6,3}^3$	Internal corrosion	AND	Independent	Independent	Independent	Independent
$J_{7,4}^3$	Vulnerability	OR	No Info	No Info	Positive Strong	No Info
$J_{9,5}^3$	DBP formation	AND	Independent	Independent	Independent	Independent
$J_{10,5}^3$	Disinfectant loss	AND	Independent	Independent	Independent	Independent

Note: Highlighted events are identified as major risk events.