THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Assessing and Managing Health Risks in Centralised Drinking Water Distribution Networks

From Theoretical Models to Practical Implementations

VICTOR RAFAEL DE JESUS VIÑAS COS

Department of Architecture and Civil Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

Assessing and Managing Health Risks in Centralized Drinking Water Distribution Networks: From Theoretical Models to Practical Implementations
VICTOR R. VIÑAS COS
ISBN 978-91-8103-160-7

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5618
ISSN 0346-718X

Department of Architecture and Civil Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
Schematic of the outbreak case study water distribution networks, and figures showing P_{inf} results and TOPSIS scores.

Chalmers digitaltryck
Gothenburg, Sweden 2025

Assessing and Managing Health Risks in Centralized Drinking Water Distribution Networks: From Theoretical Models to Practical Implementations
VICTOR R. VIÑAS COS
Department of Architecture and Civil Engineering
Chalmers University of Technology

ABSTRACT

In 2010, the United Nations General Assembly recognized access to safe and clean water and sanitation as a universal human right, further reinforced by its inclusion in the Global Sustainable Development Goals (SDG), specifically SDG 6. Despite significant progress, with 2.1 billion people gaining access to safely managed drinking water between 2000 and 2022, the risk of waterborne illnesses from contaminated drinking water remains a critical global public health issue. Centralized water distribution networks, essential for delivering safe drinking water, are vulnerable to contamination due to unforeseen events or operational errors. This thesis aims to establish a comprehensive microbial risk management framework for water distribution networks, focusing on maintaining the integrity of these systems to prevent contamination and ensure public health.

The research involved a comprehensive literature review to identify key microbial risks in water distribution networks. Two primary microbial risk events: (i) cross-connections and backflows, and (ii) intrusion, were simulated and evaluated. A fault-tree analysis (FTA) method was developed to estimate infection probabilities using national and local data, validated through outbreak simulations. Intrusion events were modelled and evaluated for various scenarios, with a stochastic approach proposed to estimate infection probabilities. Additionally, water quality analysis and metagenomics provided further input data. A risk management approach combining Quantitative Microbial Risk Assessment (QMRA) with machine learning and multi-criteria decision analysis (MCDA) was introduced to predict pipe failures and support decision-making.

The research revealed significant findings regarding microbial risks in water distribution networks. National and local risk estimates for cross-connection and backflow events showed median daily infection risks for *Campylobacter*, norovirus, and *Cryptosporidium* ranging from 10^{-6} to 10^{-3}. Surprisingly, the highest risk was observed in endemic cases, likely due to frequent but unreported incidents. Local assessments, such as in Gothenburg, indicated lower risks, highlighting the need for localized evaluations.

Simulations of intrusion events demonstrated unacceptable infection risks across all scenarios, with *Campylobacter* posing the highest risk. Stochastic modelling showed lower infection probabilities compared to earlier estimates, incorporating multiple pipe failures and variable contamination levels. Risk management measures, including improved physical separation and reduced leakage rates, significantly lowered infection probabilities. Outbreak case study aligned well with epidemiological data, validating the model's prediction. The research underscores the importance of robust microbial risk management frameworks to safeguard public health in water distribution networks.

Keywords: hydraulic modelling, gastrointestinal illness, water distribution network, quantitative microbial risk assessment (QMRA), health risk

LIST OF PUBLICATIONS

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Viñas, V**., Malm, A., & Pettersson, T. J. R. (2019). Overview of microbial risks in water distribution networks and their health consequences: quantification, modelling, trends, and future implications. *Canadian Journal of Civil Engineering, 46*(3), 149-159. doi:10.1139/cjce-2018-0216

II. **Viñas, V**., Sokolova, E., Malm, A., Bergstedt, O., & Pettersson, T. J. R. (2022). Cross-connections in drinking water distribution networks: Quantitative microbial risk assessment in combination with fault tree analysis and hydraulic modelling. *Science of the Total Environment*, *831*, 154874. doi:10.1016/j.scitotenv.2022.154874

III. Odhiambo, M., **Viñas, V**., Sokolova, E., & Pettersson, T. J. R. (2023). Health risks due to intrusion into the drinking water distribution network: hydraulic modelling and quantitative microbial risk assessment. *Environmental Science: Water Research & Technology, 9*(6), 1701-1716. doi:10.1039/D2EW00720G

IV. **Viñas, V.**, Modin, O., Odhiambo, M., Mohammadi, A., & Pettersson, T. J. R. (2024). Evaluating Microbial Contamination Risks in Drinking Water Networks Using Stochastic Modelling and Metagenomics. Manuscript.

V. Sangroula, U., **Viñas, V.**, Odhiambo, M. & Pettersson, T.J.R. (2024). Integrating Machine Learning and Multi-Criteria Decision Analysis for Health Risk Assessment in Water Distribution Networks. Manuscript.

The author of this thesis made the following contributions:

Paper I. Conceptualization; Methodology; Investigation; Writing – Original draft preparation; Writing – review & editing; Visualization.

Paper II. Conceptualization; Methodology; Formal analysis; Investigation; Writing – Original draft preparation; Writing – review & editing; Visualization.

Paper III. Formal analysis; Investigation; Visualization; Writing – review & editing.

Paper IV. Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation (sample collection, preparation); Writing – Original Draft; Writing – review & editing; Visualization.

Paper V. Methodology; Software; Validation; Formal analysis; Writing – Original draft preparation; Visualization

Other work and publications not appended to the thesis:

Vinas, V. (2018). *Assessing the health consequences of deficiencies in water distribution networks: a basis for future network management.* (Licentiate thesis). Chalmers University of Technology, Gothenburg.

Gontia, P., Thuvander, L., Ebrahimi, B., **Vinas, V.**, Rosado, L., & Wallbaum, H. (2019). Spatial analysis of urban material stock with clustering algorithms: A Northern European case study. *Journal of Industrial Ecology, 23*(6), 1328-1343. doi:https://doi.org/10.1111/jiec.12939

Rosén, L., Lindhe, A., Bergion, V., Sokolova, E., Bondelind, M., Malm, A., Pettersson, T., Moona, N., **Vinas, V.**, Lång, L.-O., Sörén, K., Sjöling, Å., & Murphy, K. (2020). *Risk-based decision support for safe drinking water supply (in Swedish)* (2020-05). Retrieved from Stockholm: https://vattenbokhandeln.svensktvatten.se/produkt/riskbaserat-beslutsstod-for-saker-dricksvattenforsorjning/

Viñas, V., Malm, A., & Pettersson, T. J. R. (2016). *An extensive review of research on microbial risks in the water distribution network.* Poster presentation at The 10th Nordic Drinking Water Conference, Reykjavík, Iceland.

Viñas, V., Odhiambo, M., Mohammadi, A. S., & Pettersson, T. J. R. (2022). *Microbial Water Quality Sampling to Inform Health Risk Assessments: Is it a Viable Strategy?* Poster presentation at the Water Safety Conference, Narvik, Norway.

Sangroula, U., **Viñas, V.**, Odhiambo, M. & Pettersson, T.J.R. (2024). *Strategies for Effective Leakage Risk Management in Water Distribution Networks*. The 6th International Conference for Water Safety 4-6 September, Montevideo, Uruguay.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and contributions of many individuals and organizations.

First and foremost, I would like to express my sincere gratitude to the Gothenburg Region (GR), which includes the municipalities of Ale, Alingsås, Göteborg, Härryda, Kungsbacka, Kungälv, Lerum, Lilla Edet, Mölndal, Partille, Stenungsund, Tjörn, and Öckerö, for their main financial support of the project, in addition to contributing with their technical expertise throughout my PhD work. I am also deeply grateful for the contributions from Chalmers institutionsmedel, Chalmers DRICKS, and Future City Water, project number 2020-01842.

I would like to extend my heartfelt thanks to my main supervisor, Thomas Pettersson, for his indescribable support throughout the years. Thomas, thank you for your unconditional understanding and guidance during the last nine years; I would not be here without you. I am also grateful to my co-supervisors, Annika Malm, Ekaterina Sokolova, and Olof Bergstedt, for their invaluable guidance, support, and encouragement throughout my research journey.

I would also like to extend my gratitude to all members of the DRICKS framework programme for drinking water research. It has been a privilege to participate in the engaging discussions, meetings, and seminars with such a dedicated and knowledgeable group. Your insights and collaboration have greatly enriched my research experience. I am also thankful to Catherine Paul and the Water Research School network for providing a platform for knowledge exchange and professional development.

I would like to express my gratitude to the Water Environment Technology (WET) division at Chalmers University for providing a fun and stimulating environment where I had the opportunity to meet and collaborate with other amazing researchers.

I am also deeply thankful to my coauthors Uchit Sangroula, Michael Odhiambo, Amir JKK Saeid Mohammadi, and Oskar Modin. Their valuable contributions have significantly enhanced the quality of my publications. Watching them work has been an incredible learning experience, and their dedication and expertise have inspired me to become a better engineer.

Additionally, I would like to thank my colleagues at AFRY for their support. I am grateful to Anders Janson, Anders Lilja, and especially Edwards Santana Miranda, who encouraged me the most to finish my PhD. Their understanding and flexibility in allowing me to take leave to complete my thesis, as well as the opportunity to join such a dynamic group, have been invaluable.

Finally, I would like to thank my friends and family for their unwavering support and encouragement throughout this journey. Sadly, you are too many to list individually.

Göteborg, January 2025

Victor Vinas

LIST OF ACRONYMS AND ABBREVIATIONS

AGI	Acute Gastrointestinal Illness
C	Infection Consequences
DALY	Disability-Adjusted Life Years
FTA	Fault-Tree Analysis
Log_{10}	Logarithmic reduction. Each log_{10} reduction corresponds to a tenfold (90%) decrease in the number of microorganisms, e.g., 2 log_{10} = 99% reduction.
MCDA	Multi-Criteria Decision Analysis
MPN	Most Probable Number
OR	Odds Ratio
PDD	Pressure-Dependent Demand
P	Probability of Failure
P_{inf}	Probability of Infection
$P_{illness}$	Probability of Illness
QMRA	Quantitative Microbial Risk Assessment
R	Risk, P x C. In the context of this thesis, it will mainly refer to Daily Risk of Infection.
RR	Risk Ratio
WDN	Water Distribution Network

TABLE OF CONTENTS

1 Introduction

In 2010, the United Nations General Assembly adopted a resolution (A/RES/64/292) which recognised access to safe and clean water and sanitation as a universal human right. This commitment was further reinforced by its inclusion in the Global Sustainable Development Goals (SDG), specifically SDG 6 “Clean water and sanitation”. According to the (WHO, 2022), water is considered safe when it “[…] does not represent any significant risk to health over a lifetime of consumption, including different sensitivities that may occur between life stages.” Between 2000 and 2022, access to safely managed drinking water was gained by 2.1 billion people (WHO/UNICEF JMP, 2023). In the same period, the number of people lacking at least basic drinking water services decreased from 1.2 billion to 703 million. Of those who gained access to safely managed drinking water services, two-thirds (1.4 billion) resided in urban areas. Despite these advances, the risk of waterborne illnesses connected to contaminated drinking water remains a significant global public health concern. It is estimated that more than 500,000 diarrhoeal deaths occur annually as a result of microbial contamination of drinking water. A significant proportion of these fatalities are among children.

In urban areas across the globe, centralized systems are commonly utilized to deliver safe drinking water to billions of consumers. A fundamental aspect of any centralized system is the water distribution network, which is responsible for conveying water from the drinking water treatment plant or an appropriate water source to the consumers’ taps. The distribution network is composed of several components, including pipes, valves, storage reservoirs, and pumps (WHO, 2022). These components function collectively to maintain water quality and ensure an adequate supply of water to consumers. However, these components are susceptible to unforeseen events or operational errors. Such incidents may result in the contamination of the drinking water. Furthermore, due to the location of the distribution network at the end of the supply chain, incidents are less likely to be detected and remediated in a timely manner (Risebro et al., 2007). Consequently, it is of the utmost importance to maintain the integrity of the network to prevent contamination of the treated water delivered to consumers.

The integrity of the distribution network is categorised into three components: physical, hydraulic, and water quality (National Research Council, 2006). The first of these, physical integrity, refers to the ability of the distribution system to act as a physical barrier against external contamination. This can be compromised if, for instance, there are cross-connections with non-potable water pipes or cracks in the pipes. Secondly, hydraulic integrity is the capacity of the system to maintain adequate flow, pressure, and water age. It is possible that certain events, such as pump shutdown and main breaks, may result in a reduction in pressure and flow, thereby compromising the hydraulic integrity of the system. A contamination event can only occur when both the physical and the hydraulic integrity of the system are compromised (Ercumen et al., 2014). Finally, water quality integrity concerns the biochemical processes occurring within the pipes that can result in a deterioration of the drinking water quality. A deterioration in water quality can result in contamination or increase the likelihood of contamination occurring. An example of a breach in water quality integrity is the complete decay of the disinfectant residual.

Waterborne disease outbreaks represent extreme consequences of contamination in the distribution network. Waterborne outbreaks are defined as “an incident in which two or more epidemiologically linked persons experience a similar illness after exposure to the same water source and epidemiologic evidence implicates the water as the likely source of the illness” (WHO, 2022). In Sweden, approximately 34% of the outbreaks with known causes are associated with the distribution network (Malm et al., 2010). This proportion is similar to the European Union level, where 31% of the outbreaks were caused by distribution network deficiencies (Risebro et al., 2007). Gastrointestinal illness (GI) is the most common illness associated to waterborne outbreaks (Messner et al., 2006).

There are different tools available to evaluate the (microbial) health risks associated to distribution networks. One method is through epidemiological studies, which are commonly performed during outbreak investigations. Their purpose is to determine the extent of the outbreak (how many were affected/infected) and identify both the causative agent of the disease and the events that led to the presence of the agent in the drinking water supply (Institute of Medicine, 2000). In the last three decades, randomised control trials (RCTs) and epidemiological studies have also been used to estimate the risk of disease at an endemic level. While outbreaks can be seen as extreme cases, the endemic level of disease is a sort of baseline level of disease in a population. Results for these studies have been mixed; some studies have found an increased risk of illness from drinking tap water while other authors have not found any association (Colford et al., 2005; Hellard et al., 2001; Malm et al., 2013; Nygård et al., 2007; Payment et al., 1991; Payment et al., 1997). However, there is mounting evidence that malfunctioning distribution networks, as well as specific system deficiencies (i.e., pipe breaks, water outages and inadequate residual disinfectant), increase the risk of endemic GI (Ercumen et al., 2014).

Another way to analyse and assess the risks in the distribution network is using computational models. Computational fluid dynamics (CFD) models have been developed to simulate contamination events and transport inside distribution systems (Mansour-Rezaei & Naser, 2013; Mora-Rodríguez et al., 2014). Quantitative microbial risk assessment (QMRA) models have been used in conjunction with hydraulic models to quantify the consequences of different microbial risks (Blokker et al., 2018; Jamal et al., 2020; Teunis et al., 2010b; Yang et al., 2011). Most of these models have important limitations that restrict their use, e.g., uncertainties in the input data, assumptions made about the conditions in the distribution network, e.g. turbulent flow, instantaneous mixing, and so on (Besner et al., 2011). However, they are useful for evaluating measures that can be implemented to manage the specific risks addressed in the model.

1.1 Research objectives

The primary objective of the research in this thesis is to establish a comprehensive microbial risk management framework for water distribution networks. This involves:

- Conducting an extensive literature review to understand the current state-of-the-art in microbial risks associated with water distribution networks.
- Developing a theoretical model for assessing and managing health risks in the distribution network, with a focus on cross-connection risks.
- Identifying potential input data for the Quantitative Microbial Risk Assessment (QMRA) model from the literature and real-world data.
- Developing a methodology for assessing the health risk of specific events, such as pipe breaks and intermittent water supply, in the distribution network.
- Validating and improving the developed models and methodologies using stochastic simulations and water quality samples.

1.2 Research questions and scope

The research will address the following questions:

RQ1. What are the current methodologies and tools for evaluating microbial risks in water distribution networks?

RQ2. How can these methodologies be improved to provide a more accurate and comprehensive assessment of health risks in water distribution systems?

RQ3. What are the most influential factors affecting the risk of infection in water distribution networks?

RQ4. How can the developed models and methodologies be integrated into a comprehensive microbial risk management framework for water distribution networks?

The scope of the research includes centralized drinking water distribution networks and focuses on microbial risks and their impact on health.

1.3 Research approach and limitations

The overarching aim of this thesis is to establish the groundwork for a microbial risk management framework, with a specific focus on the water distribution network. The management framework would be based on QMRA modelling, which is known to have some important knowledge gaps, e.g., uncertain input data, lack of system-specific information, among others (Besner et al., 2011). There have been previous attempts at overcoming some of these limitations (Islam et al., 2015; Kirmeyer et al., 2014; McInnis, 2004), however to this day they remain incomplete. This thesis is a first attempt at identifying missing links between the work already done and what needs to be addressed in the future to successfully develop a comprehensive framework.

1.4 Thesis structure

The thesis is structured as follows:

- Chapter 1: Introduction - Provides an overview of the research objectives, questions, approach, and structure.
- Chapter 2: Background - Presents a brief history of drinking water distribution networks, modern distribution networks, drinking water distribution modelling, standard operating procedures to preserve integrity, evidence for problems (RCTs & observational studies), and risk and QMRA framework.
- Chapter 3: Materials and Methods - Describes the prestudy, study sites, contamination likelihood, contamination load, contamination transport, dose-response and risk estimates, water sampling and risk reduction measures.
- Chapter 4: Results and Discussion - Presents the results of the research and discusses their implications.
- Chapter 5: Conclusions - Summarizes the findings of the research and discusses possibilities for future work.

The chapters are organized according to the phases of the research process, with each chapter contributing to the overarching objective of developing a comprehensive microbial risk management framework for water distribution networks. The five papers that form part of the thesis are integrated into the relevant sections of the thesis. **Paper I** serves to establish the basis for the literature review and to provide input data for the QMRA models. **Paper II** makes a contribution to the development of a conceptual model for the assessment of the probability of a cross-connection and backflow event. In addition, it employs a waterborne outbreak case study for the purpose of validating the results. **Paper III** presents a methodology for assessing intrusion risks in the distribution network. **Paper IV** offers an improvement of the developed intrusion model in **Paper III**, incorporating uncertainties and stochastic factors. **Paper V** introduces the use of machine learning-based probability estimations for pipe failures and decision support through multi-criteria decision analysis, with the objective of implementing risk-reducing measures within the risk management framework. An overview of the papers and the workflow for each can be found in 3 Materials and methods.

2 Background

"*And I wish to give an account of the other kinds of waters, namely, of such as are wholesome and such as are unwholesome, and what bad and what good effects may be derived from water; for water contributes much towards health*." (Hippocrates Part 7)

Important concepts for understanding the main contents of the thesis are introduced in this chapter. The focus is on the water distribution network and the microbial risks associated with it, in addition to introducing important concepts in epidemiology and the QMRA framework.

2.1 Brief history of drinking water distribution network

The evolution of water supply systems is a testament to human ingenuity and environmental understanding. From the Bronze Age civilizations of the Sumerians and Akkadians in Mesopotamia and the Minoans in Crete to the Roman Empire, advanced water distribution systems were developed (De Feo et al., 2013; Mays et al., 2013; Mays et al., 2007). These systems utilized canals connected to rivers, rainwater harvesting systems, wells, aqueducts, and underground cisterns.

For example, the Minoan palace of Knossos featured a network of terracotta piping located beneath the floors, which facilitated the distribution of water (De Feo et al., 2013). Similar terracotta pipes were discovered in other Minoan sites, including Tylissos, Gournia, and Vathypetro. The Minoan architectural style incorporated flat rooftops, light wells, and open courts that served as catch basins to collect rainwater, which then flowed to storage areas or cisterns.

The Hellenic world witnessed the evolution of underground aqueducts due to the continuous wars between ancient cities (Maliva & Missimer, 2012). These aqueducts were constructed in a subterranean manner and were composed of tubes or channels made of stone slabs or terracotta. Inverted siphons, pipes that operated under pressure, were commonly used to cross deep valleys. One of the most famous aqueducts from this period is the tunnel of Eupalinos on Samos Island, the first deep tunnel in history that was dug from two openings with the two lines of construction meeting near the middle (Mays, 2008).

The Roman Empire made extensive use of cisterns, with Pompeii being a notable example (Mays et al., 2013). The city's water distribution system was an intricate network of lead and terracotta pipes, aqueducts, and cisterns, collectively demonstrating the advanced engineering capabilities of the time. After the decline of the Roman Empire, water supply systems experienced fundamental changes. Medieval cities in Western Europe, as well as castles and monasteries, had their own wells, fountains, or cisterns (Mays et al., 2013). The eastern part of the empire retained the relevant Roman construction tradition, implemented mostly on the watering system of Constantinople and other major centres of the eastern Mediterranean.

In the Americas, the city of Xochicalco in central Mexico, which became one of the great Mesoamerican cities in the late classic period (ca. 650–900 AD), had no rivers or streams or wells to obtain water, so rainwater harvesting was the source of water (Mays et al., 2013).

In conclusion, the history of water supply systems in antiquity shows a progression from simple wells and cisterns to complex networks of underground aqueducts and pipes. These systems were adapted and improved upon by successive civilizations, demonstrating the importance of water management in the development of urban centres (Mays et al., 2007).

2.2 Modern distribution networks

Modern distribution networks are made up of many components: pipes, valves, pumps, reservoirs, hydrants and other appurtenances that connect the drinking water supply to consumers' taps (National Research Council, 2006; WHO, 2014). The presence or absence of a particular component is strongly influenced by the area that the distribution network is intended to serve. For example, the layout of the network will depend on existing roads and streets, existing and planned land use, and where water demand is concentrated (WHO, 2014).

The distribution network can be configured in three ways: branch, grid and combined (see Figure 1). The branch system, similar to the branches of a tree, has smaller pipes branching off from larger ones across the service area (National Research Council, 2006). The main advantage of this system is its lower capital cost compared to the grid system. However, it has a significant disadvantage: in the event of a failure, all downstream consumers would be affected, as the water can only follow one path to the consumers. This configuration is common in rural areas and small settlements. The grid system, on the other hand, creates loops throughout the service area, allowing water to follow two or more paths to consumers. This redundancy minimises the number of consumers affected if neighbourhoods need to be isolated (Svenskt Vatten, 2020). However, it is more expensive due to the increased system length resulting from the loops. This configuration is typical in large, densely populated areas.

In practice, most large distribution networks are a combination of loops and branches (Svenskt Vatten, 2020). The denser, central parts of a city are likely to have loops, while peripheral neighbourhoods are often served by a single branch. As mentioned above, the choice of system is primarily influenced by local topography, street layout and the type of community to be served.

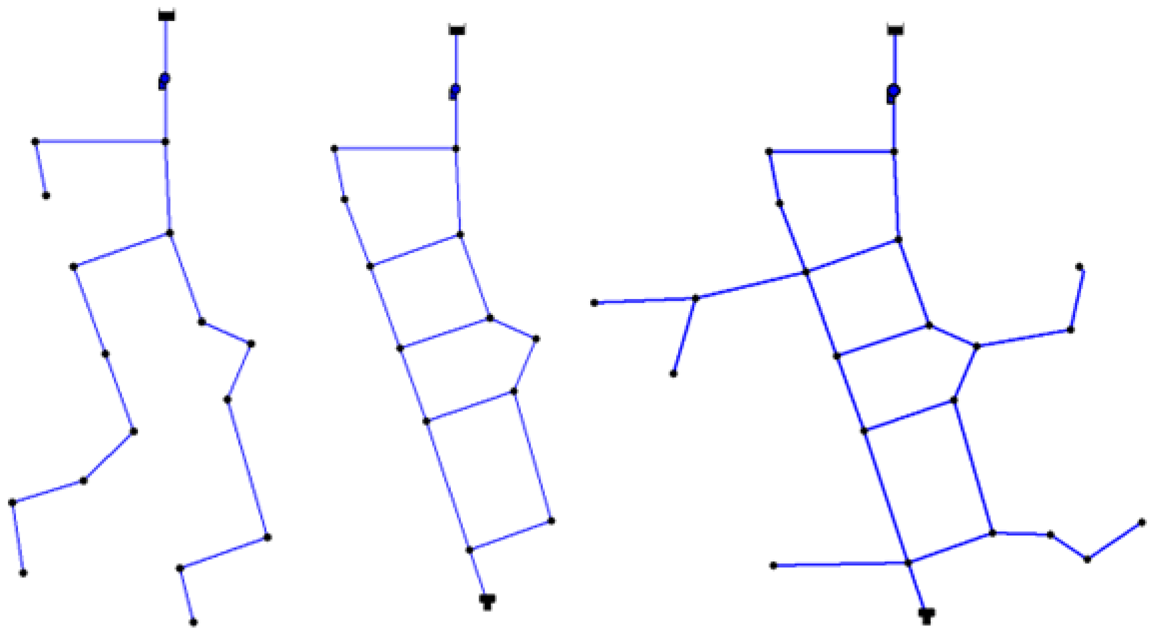

Figure 1. Schematic branch, grid and combined network configurations in which the distribution network can be structured.

2.3 WDN modelling

Hydraulic models serve as the foundation for the modelling of drinking water distribution networks (WDN). These mathematical representations of fluid flow systems are designed to emulate the behaviour of an existing system or predict conditions in a proposed system. They are capable of simulating the dynamic behaviour of WDNs, accounting for variations in water demand throughout the day and changes in operating conditions. The advent of personal computing power and the development of enhanced computational capabilities have led to the ubiquity of hydraulic modelling in the water industry, with applications in the planning, operation, and design of WDNs. Hydraulic modelling is employed on a routine basis to simulate the flow and pressure within the network under conditions of either steady state or quasi-steady state. In the context of hydraulic modelling, a steady state is defined as a condition where all parameters (e.g., flow, pressure) remain constant over time. In contrast, a quasi-steady state is defined as a condition where these parameters change slowly in comparison to the time scale of interest. Another significant aspect of hydraulic modelling is transient analysis, which pertains to the study of rapid changes in flow and pressure (Boulos et al., 2005).

In addition to hydraulic modelling, water quality modelling represents another crucial component of WDN modelling. The objective of water quality modelling is to utilise information derived from the hydraulic model (e.g., flows) in order to simulate the spatio-temporal variability of water quality within the network.

Hydraulic modelling is employed for a variety of purposes, including calculation, optimisation, water quality modelling, and asset management. In recent times, the scope of its applications has been broadened to encompass flushing, fire-flow analysis, transient analysis, multi-species quality modelling, resilience, and other areas. The calibration of these models is typically based on real-time monitoring of network parameters.

The Hardy Cross, Linear (Wood & Charles, 1972), Newton-Raphson (Larock et al., 1999), and Gradient algorithm (Todini & Pilati, 1988) methods are among the most popular methods for analysing looped pipe networks (Tsakiris & Spiliotis, 2014). All these methods are numerical iterative algorithms designed to solve a set of linear and non-linear equations. Depending on the unknown determinant, these methods are classified as head-based or flow-based methods (Swamee & Sharma, 2008).

In addition to hydraulic modelling, prediction models can be employed to forecast failures in the WDN. A broad categorisation of these models reveals three predominant types: physical models (Makar, 2000), statistical models (Yamijala et al., 2009), and machine learning models (Fan et al., 2022; Warad et al., 2024). Physical models simulate the underlying mechanical and environmental processes, taking into account the pipe's internal and external loads, as well as material deterioration caused by environmental and operational conditions. Such models require large quantities of data from inspections, which may be technically difficult and expensive (Wilson et al., 2017). Statistical and machine learning models have been found to be more cost and time effective than physical models for predicting pipe failures, as they can provide accurate predictions even with minimal data (Kimutai et al., 2015; Winkler et al., 2018). Statistical models analyse historical failure data to identify patterns and estimate the probabilities of future breaks, aiming to establish relationships between explanatory variables and failure patterns through mathematical equations (Kleiner & Rajani, 2001). In recent years, machine learning models have become more widely adopted for pipe break analyses due to their ability to provide accurate predictions and capture complex relationships among the explanatory variables. In comparison to statistical models, machine learning models have been shown to possess enhanced predictive accuracy and a superior capacity to manage outliers (Forero-Ortiz et al., 2023).

The integration of these methods with Supervisory Control and Data Acquisition (SCADA) systems, Geographic Information Systems (GIS), and Computer-Aided Design (CAD) tools, along with the use of sensors, further enhances the capabilities of hydraulic modelling, allowing for more accurate and efficient management of WDNs.

2.4 Standard Operating Procedures to preserve integrity

Standard Operating Procedures (SOPs) are established protocols that provide guidance on how to carry out a range of tasks within the WDNs (Kanakoudis & Tsitsifli, 2019; WHO, 2014). These include:

- Maintaining Flows and Positive Pressure: This involves ensuring that water flows are maintained at optimal levels and that the pressure within the system is kept positive to prevent the ingress of contaminants.
- Operating Intermittent Supplies: This refers to the management of water supplies that are not continuous, requiring careful scheduling and monitoring to ensure adequate supply when needed.
- Maintaining Disinfection Throughout the Distribution System: This involves ensuring that water remains safe for consumption as it travels through the distribution system, which may include the addition of disinfectants.
- Mixing Water Supplies from Different Sources: This involves the careful blending of water from various sources to ensure consistent water quality.
- Inspection and Maintenance of Infrastructure: Regular inspection and maintenance of storage tanks, service reservoirs, valves, and other fittings are crucial to prevent leaks and ensure the efficient operation of the system.
- Water Leakage Management: This involves the detection and repair of leaks in the system to minimize water loss. The Infrastructure Leakage Index (ILI) is an example of a performance indicator used to measure real water loss from the supply network.
- Preventing Corrosion: This involves the use of various methods to prevent the corrosion of pipes and fittings, which can lead to leaks and water quality issues.
- Selection of Pipe Materials and Chemicals: This involves choosing appropriate materials for pipes and chemicals for water treatment based on several factors such as cost, durability, and safety.
- Installation of Backflow Prevention Devices: These devices prevent the reverse flow of water within the system, protecting the water supply from contamination.
- Ongoing Evaluation of Backflow Prevention Devices: Regular checks and maintenance of these devices are crucial to ensure they are working effectively.
- Repairing Water Main Breaks: This involves the prompt detection and repair of breaks in the water mains to prevent significant water loss and disruption to the supply.
- Construction and Commissioning of New Mains: This involves the planning, construction, and testing of new water mains before they are put into operation.
- Dewatering and Recharging Distribution Mains: This involves the removal and replacement of water within the mains, often carried out during maintenance or repair work.
- Controlling Permeation: This involves measures to prevent the ingress of contaminants into the water supply through the pipe walls.
- Collection and Testing of Water Samples: Regular sampling and testing of water are carried out to monitor water quality and ensure it meets health and safety standards.
- Calibrating Equipment and SCADA Systems: Regular calibration ensures that equipment and systems are working accurately and efficiently.
- Dealing with Customer Enquiries: This involves providing information and assistance to customers, addressing their concerns and queries about the water supply.

The general principles provided by SOPs are frequently codified in laws, international standards, handbooks, and other means to aid water providers in their implementation. For example, the Swedish

Water and Wastewater Association ("Svenskt Vatten") routinely releases checklists, handbooks, and reports on how to responsibly manage WDNs in Sweden.[1]

2.5 Evidence for problems (RCTs & observational studies)

From the descriptions in the previous section, it can be concluded that the drinking water distribution network is a complex system in many respects. This complexity arises from the dynamic nature of water use and the need for a variety of components to work together to maintain a delicate balance of appropriate operating conditions. Even Standard Operating Procedures (SOPs) have their own set of difficulties, ranging from logistical, technological and economic constraints that make it difficult to satisfy all the ideal controls in the network. Given these constraints, one might ask: Can the distribution network function as intended to deliver **safe** drinking water? Is there evidence that the distribution network could potentially be a source of disease to the population it serves?

The relationship between water and health has been discussed since ancient times. Throughout antiquity, tasty or tasteless, cool, odourless and colourless water were considered desirable qualities [*Then such waters as flow to the rising sun, must necessarily be clear, fragrant, soft, and delightful to drink, in such a city (…) The persons of the inhabitants are, for the most part, well coloured and blooming, unless some disease counteract.*][2] and stagnant, murky, swampy water was to be avoided.

The emphasis on desirable qualities in water stemmed from the belief that all disease was transmitted by vapours or mists filled with *miasmata* (vapours of decomposing organic matter). Water with poor qualities was thought to be contaminated with miasma, which would release vapours into the air and cause disease when inhaled. The miasma theory of disease remained the dominant explanation of disease transmission until the mid to late nineteenth century, when it began to be superseded by the germ theory of disease. The germ theory of disease gained prominence in part through experiments that disproved spontaneous generation (e.g. Louis Pasteur) and the identification of specific microorganisms as causative agents of disease (e.g. Koch's postulates). In addition, even before the aetiological agent of cholera was identified, John Snow (not to be confused with the fictional character) mapped cholera deaths in Soho, London, to support his conclusion that the source of the disease was contaminated water supplied by the Broad Street pump. John Snow's work is considered to be the foundation of modern epidemiology.

Epidemiology can be defined as the study of the distribution of disease and the factors that influence its frequency in human populations (Silman et al., 2018). Epidemiology seeks, among other things, to identify factors that affect health (e.g., agents that transmit disease, environmental factors), identify sensitive groups in a population, investigate outbreaks and control epidemics (Silman et al., 2018). There are two main measures of disease in a population: prevalence and incidence (Silman et al., 2018). Prevalence is the proportion of a population affected by a disease at a specific point in time, which is described in Equation (1).

$$\text{Prevalence} = \frac{\text{Cases}}{\text{Population (at risk)}} \tag{1}$$

Incidence is the number of new cases in a population within a specified time period. It can be expressed as a proportion (cumulative incidence Eq.(2)) or as a rate (Eq.(3)). Cumulative incidence can be used to calculate the level of risk for a certain population. The cumulative incidence is also known as attack rate.

[1] Svenskt Vatten homepage: Advice and Guidelines. https://www.svensktvatten.se/vattentjanster/dricksvatten/rad--riktlinjer/

[2] Hippocrates On Airs, Waters, and Places

$$\text{Cumulative incidence} = \frac{\text{New cases}}{\text{Population at risk for some amount of time}} \tag{2}$$

$$\text{Incidence rate} = \frac{\text{New cases over a time period}}{\text{Time each person was observed, summed for population}} \tag{3}$$

To estimate the increased risk of disease for a population exposed to a certain variable, three different ratios can be used. Risk ratio (RR) is the ratio of cumulative incidence in two population groups (Equation (4)). The risk ratio can also be seen as the relative risk of disease. If the RR = 1, the incidence is the same in the exposed group and the unexposed group: there is no association between exposure/risk factor and disease. If RR > 1, there is an increased risk of disease in the exposed group than in the unexposed group. If RR < 1, there is a reduction in risk of disease for the exposed group.

$$\text{Risk Ratio} = \frac{\text{Risk (cumulative incidence) in exposed group}}{\text{Risk (cumulative incidence) in unexposed group}} \tag{4}$$

Odds ratio measures the relative odds of an outcome occurring after an exposure. The numerator is the number of exposed cases divided by the number of unexposed cases. The denominator is the number of exposed non-cases divided by the number of unexposed non-cases. An OR > 1 indicates increased odds of developing the outcome when exposed to a given variable, and an OR < 1 indicates the opposite. An OR = 1 indicates that the odds of the outcome are not affected by the variable (Szumilas, 2010). Odds ratios can be expressed as in Equation (5):

$$\text{Odds Ratio} = \frac{\text{Odds of disease in exposed group}}{\text{Odds of disease in unexposed group}} \tag{5}$$

For rare diseases, i.e., where the prevalence is low, odds ratio are numerically similar to the risk ratio (Silman et al., 2018).

The findings of studies investigating the correlation between the consumption of tap water and gastrointestinal illness (GI) have been inconclusive. Some studies have indicated that tap water plays a significant role in the endemic level of GI, whereas other studies have not identified such an association. A synthesis of the findings from a selection of the studies is provided in Table 1.

Table 1. Summary of RCTs and observational studies performed in the drinking water distribution network to assess the association between distribution network and GI. Table 1 in **Paper I.**

Study	Location	Study design	Blinding	Size of the study*	Follow-up period	Results†	Attributable risk
Payment et al. (1991)	Canada	Cluster randomized controlled trial	No	607 / 2 408	12 months	IRR = 1.5 ($p < 0.01$)	≈35% excess GI in the tap water group compared to control
Payment et al. (1997)	Canada	Cluster randomized controlled trial	No	1 369 / 5 253	16 months	IRR = 1.15 ($p < 0.01$)	14%–19% excess risk of GI; 17%–40% in children 2–5 years old
Hellard et al. (2001)	Australia	Cluster randomized controlled trial	Yes	600 / 2 811	12 months	IRR = 0.99 (0.85–1.10)	No association found
Nygard et al. (2004)	Sweden	Ecological	n.a.	– / 7 280	n.a.	1. IRR = 1.11 (1.08–1.15) 2. IRR = 1.12 (1.08–1.16) 3. IRR = 1.13 (1.09–1.17)‡	Significant association of length of pipe directly proportional to increased risk of infection
Colford et al. (2005)	United States	Cluster randomized controlled trial	Yes	456 / 1 296	12 months	IRR = 0.98 (0.87–1.10)	No association found
Hunter et al. (2005)	United Kingdom	Case-control	n.a.	– / 427	n.a.	OR = 12.5 (3.5–44.7)	Significant association between low pressure event and disease ($p < 0.01$)
Nygard et al. (2007)	Norway	Cohort	n.a.	1 159 / –	n.a.	IRR = 1.58 (1.1–2.3)	Attributable fraction of 37% one week after exposure
Tinker et al. (2009)§	United States	Ecological	n.a.	– / 1 700 000	n.a.	1. OR = 1.00 (0.96–1.03) 2. OR = 0.99 (0.96–1.03) 3. OR = 1.07 (1.03–1.10) 4. OR = 1.05 (1.02–1.08)‖	Slight association directly proportional to the residence time and increased risk of disease
Malm et al. (2013)	Sweden	Ecological	n.a.	– / 500 000	n.a.	SIR = 1.08 (0.86–1.32)	No association found due to low pressure events
Säve-Söderbergh et al. (2017*a*)	Sweden	Cohort	n.a.	3 238 / 7431	n.a.	GI: OR = 1.1 (0.9–1.5) AGI: OR = 2.0 (1.2–3.3) Vomiting: OR = 1.9 (1.2–3.0)	Significant association for AGI and vomiting

*Sample size is given by: No. of households / No. of individuals.
†IRR, Incidence Risk Ratio; OR, Odds Ratio; SIR, Standardized Infection Ratio.
‡Result (1) is from univariate analysis; results (2) and (3) are from multivariate analyses.
§Included more water suppliers in a subsequent study (Tinker et al. 2010).
‖(1) OR between intermediate and short residence times and utility 1; (2) OR between intermediate and short residence times and utility 2; (3) OR between intermediate and long residence times and utility 1; (4) OR between intermediate and long residence times and utility 2.

2.6 Risk and QMRA framework

Risk is typically defined in several ways, depending on the subject of evaluation (Aven, 2010). According to the WHO (2022), risk "is the likelihood of identified hazards causing harms in exposed populations in a specific time frame, including the magnitude of that harm and/or the consequences." It is commonly expressed as Equation (6):

$$R = P \times C \tag{6}$$

where R is the risk, P is the likelihood of occurrence of an undesired event (e.g., pipe break) and C represents the consequences of the event (e.g., infection from drinking contaminated water).

A hazard is an agent with the potential to cause harm, which can originate from microbiological, chemical, radiological, or physical sources (WHO, 2022). The most apparent source of microbial risk is when the WDN is affected by the sewage system.

2.6.1 Health risks associated to microbiological contamination

Five major microbial risks have been identified for the distribution network (National Research Council, 2006). From higher to lower priority, they include: cross-connections and backflows; improper maintenance and operation of reservoirs; contamination during installation, rehabilitation, and repair of water mains; intrusion; and biofilms. The prioritization is based on the amount of evidence supporting their contribution to health risks for consumers. All these risks have been identified as causes of waterborne disease outbreaks (Craun, 2012; Hrudey & Hrudey, 2004; Hrudey & Hrudey, 2007; Risebro et al., 2007). A brief explanation of each risk is provided below.

Cross-connections are defined as points in the distribution network where non-potable water elements (e.g., wastewater pipe) may come into contact with the drinking water (USEPA, 2002). A backflow can occur when the pressure of the non-potable water source exceeds that in the distribution network and there are inadequate cross-connection controls present (e.g., absence of a backflow prevention valve) (WHO, 2014). Cross-connections are considered one of the most serious public health risks in the distribution network (National Research Council, 2006; WHO, 2014).

Routines exist to ensure correct hygienic practices during the installation, rehabilitation, and repair of water mains (Säve-Söderbergh et al., 2017; WHO, 2014, 2022). However, contamination can occur if these procedures are not followed (e.g., inadequately disinfecting newly laid pipes) (National Research Council, 2006). Unhygienic practices during the installation, rehabilitation, and repair of water mains were classified as a high-priority issue (National Research Council, 2006).

Water quality in reservoirs can be compromised in several ways. Physical breaches, such as cracks in the walls/roofs of the reservoir and cross-connections, can allow contamination to enter from the exterior (Falco & Williams, 2009; Kristianstads kommun, 2015). Inadequate hydraulic design can lead to long residence times, resulting in a complete loss of disinfectant and microbial regrowth (Clark et al., 1996; National Research Council, 2006; Seyoum & Tanyimboh, 2014). Improper management of reservoirs was considered a high-priority issue in the first assessment report by the National Research Council (2006).

According to Besner et al. (2011), intrusion is defined as the contamination of drinking water due to adverse pressure conditions and physical breaches in the system. Three conditions are necessary for microbial contamination to occur: presence of pathogens surrounding the distribution network (source); occurrence of pressure transients or low-pressure events (adverse pressure conditions); and deteriorated physical conditions of the pipes (physical breach). Intrusion was considered a medium priority issue by the National Research Council (2006); however, it is gradually being recognized as a major contributor to the waterborne disease burden (Besner et al., 2011; Islam et al., 2015; Murphy et al., 2016).

Biofilms are complex assemblies of microorganisms, extracellular polymeric substances, and organic and inorganic matter (Kauppinen et al., 2012). They are known to act as potential reservoirs for pathogens within the distribution network (Berry et al., 2006; Nocker et al., 2014; Wingender & Flemming, 2011). Pathogenic organisms that intrude the distribution network (e.g., via cross-connections and backflow into the system) can attach to biofilms and subsequently detach due to shear stresses caused by increased water flow. Biofilms can harbour a variety of pathogens, including *Cryptosporidium* oocysts (Angles et al., 2007), enteric viruses (Skraber et al., 2005; Storey & Ashbolt, 2003), opportunistic pathogens (Farkas et al., 2012; Pryor et al., 2004), and bacterial pathogens (September et al., 2007; Wingender & Flemming, 2011). The National Research Council (2006) classified biofilms as a medium-priority issue.

2.6.2 QMRA

One of the most valuable methods available for quantifying microbial risks is the quantitative microbial risk assessment (QMRA) framework (Petterson & Ashbolt, 2016)(Petterson et al. 2016). According to WHO (2016), a QMRA consists of four basic steps:

1 Problem formulation: This step determines the scope and purpose of the assessment. Hazards, exposure pathways and health outcomes are investigated;
2 Exposure assessment: This includes quantifying pathogen sources, magnitude, and frequency of the exposure for the different scenarios being analysed;
3 Health effects assessment: This involves estimating the health impact from the identified hazards and the population of the study (e.g., drinking water consumers);
4 Risk characterization: This step combines the exposure and health effects assessments to quantify the risk of infection. This can be represented as number of consumers infected per year, DALYs. A sensitivity analysis can also be performed in this step to determine which parameters influence the most the QMRA results.

An alternative QMRA framework presents a 5-step approach: 1) hazard identification, 2) dose-response assessment, 3) exposure assessment, 4) risk characterisation and 5) risk management (Haas et al., 2014; WHO, 2016). In the risk management step, costs and measure effectiveness are important components to make a decision after performing the previous four steps. It is important to note that to perform a relevant assessment, uncertainties must be considered in each step. Otherwise, the results may not be representative of reality (Bouwknegt et al., 2014). A detailed explanation of each step from a distribution network perspective follows below.

1. Problem formulation

The main aim with problem formulation is to determine the scope and the purpose of the risk assessment. A successful problem formulation requires effective identification of potential hazards, contamination pathways, and exposure outcomes. (WHO, 2016). However, already in this step some levels of assumptions are needed to successfully formulate the problem.

An important assumption made early in the assessment is the choice of reference pathogens, as it is not possible to assess all waterborne pathogens in a single QMRA (WHO, 2016). The choice of reference pathogens assumes that all other pathogens of the same type are controlled in the same way as the reference. Reference pathogens should be those that are most representative of the local conditions. For example, Campylobacter, norovirus, and Cryptosporidium would be appropriate choices of reference pathogens for a QMRA carried out in Sweden: epidemiological studies support their selection (Abrahamsson; Guzman-Herrador et al., 2015).

To determine the exposure pathway, it is necessary to define which hazardous events or scenarios will be included in the assessment. This definition is necessary because the potential pathways will differ depending on the risk being assessed. For example, if the risk assessor is interested in studying intrusion and reservoir contamination, transport within the distribution to the consumer can be identified in the same way (see Figure 2). However, the pathogen source and pathway will differ completely.

The last step in the problem formulation is to determine which health outcome will be used to assess the risk (WHO, 2016). These can be expressed as, e.g., a yearly probability of infection or disability-adjusted life years (DALYs). The choice of health outcome will depend on the objective of the risk assessment.

2. Exposure assessment

The primary objective of exposure assessment is to quantify the sources of contamination and the exposure pathways, as identified in step 1 - problem formulation (Haas et al., 2014). The quantification of potential pathogen concentrations in the distribution network has been achieved using both theoretical models (e.g., (Vairavamoorthy et al., 2007)) and source characterization (Besner et al., 2010; Karim et al., 2003). Most of the quantification has focused on the risk of intrusion and contamination during maintenance or repair work (Blokker et al., 2018; Teunis et al., 2010b; Yang et al., 2015). Additionally, hydraulic models are essential for simulating the transport of water from the source to the consumers' tap.

Risk reduction measures to limit exposure to pathogens during a contamination event in the distribution network are limited compared to options available for the source water and treatment plant (Risebro et al., 2007). According to QMRAs already performed in the network, the following parameters achieve some kind of reduction of the pathogen concentration: disinfectant residual, dilution factor, flushing (Blokker et al., 2018; Teunis et al., 2010b; Yang et al., 2015).

In the context of QMRAs in the distribution network, consumers are primarily exposed through the consumption of unboiled tap water (McInnis, 2004; Teunis et al., 2010b; Yang et al., 2011). To quantify this parameter, consumption pattern studies have been conducted in different settings (Hynds et al., 2012; Roche et al., 2012; Säve-Söderbergh et al., 2017). Local consumption pattern studies are needed to accurately describe exposure levels in the population and thus make QMRA results more reliable.

3. Health effects assessment

After determining the population that will be exposed to a certain pathogen concentration, the next step will be to assess the health risk outcomes of these. The health effects assessment uses dose-response models to relate the pathogen dose to a probability of infection or disease (Haas et al., 2014). Dose-response models are currently available for numerous pathogens: *Campylobacter* (Medema et al., 1996; Teunis et al., 2005; Teunis et al., 1999); *Salmonella* (Teunis et al., 2010a; Teunis et al., 1999); *E. coli* O157:H7 (Teunis et al., 2004; Teunis et al., 2008a); adenovirus (Teunis et al., 2016); norovirus (Messner et al., 2014; Teunis et al., 2008b); *Cryptosporidium* (Teunis et al., 1999; Teunis et al., 2002); and *Giardia* (Teunis et al., 1999; Zmirou-Navier et al., 2006).

4. Risk characterization

Risk characterization consists in combining the exposure assessment and the health effects to generate a quantitative estimation of risk. Risk estimates are commonly expressed as yearly probability of infection, probability of illness and/or disability-adjusted life years (DALYs). These estimates can be used in diverse ways, e.g., compare the calculated health risk with a regulatory target. For example, in United States the acceptable annual probability of infection allowed is set to 1/10,000 consumers (National Research Council, 2006). Consequently, if a QMRA is performed for the distribution network and the estimated risk of infection is 2/10,000 consumers per year; it could be concluded that the risk is unacceptable if compared to the US health target.

To properly characterize the risk, uncertainties must be included in the analysis (Bouwknegt et al., 2014; Lindhe, 2010). A common way to perform uncertainty analyses in distribution network health risk assessments is through Monte Carlo simulations (Khanal et al., 2006; Nilsson et al., 2005; Teunis et al., 2010b; Torres et al., 2009). In Monte Carlo simulations, probability distributions are used as input data, selecting random numbers from the distribution for each calculation. This process is performed for a certain number of iterations (e.g., 1 000, 10 000), obtaining a probability distribution as the result (Lindhe, 2010).

A sensitivity analysis may also be carried out for different reasons: refine the assessment, identify sources of uncertainty, and determine risk reduction measures, among others (WHO, 2016 984). Monte Carlo simulations can also be used for this purpose; however, the standard method involves changing the input variable and noting the extent of the change in the result.

5. Risk management

The risk characterisation results may be used to inform decisions on managing health risks (Haas et al., 2014). For example, if the risk is deemed unacceptable, mitigation measures are needed. After implementing the mitigation measure(s), a new risk estimate is calculated with the updated information. The adjusted risk estimate is then compared to the health target and evaluated accordingly. Possible mitigation measures during a contamination incident in the distribution network include (Blokker et al., 2018; Säve-Söderbergh et al., 2017):

- Isolation of the affected area
- Chlorination
- Flushing
- Boil water advisory
- Emergency water sources

Decision-making is an integral part of any successful project. It can involve either individuals or groups and can be based on a single criterion or multiple criteria (Zolghadr-Asli et al., 2021). One widely used decision-making tool is Multi-Criteria Decision Analysis (MCDA), a structured process that aims to prioritize alternatives while considering multiple varying criteria (Ngubane et al., 2024; Taherdoost & Madanchian, 2023). In general, an MCDA process involves the following steps:

1. Defining the decision problem, objectives, alternatives, and evaluation criteria
2. Structuring the decision problem
3. Incorporating stakeholder preferences
4. Evaluating and scoring the alternatives
5. Analyzing the results to provide recommendations

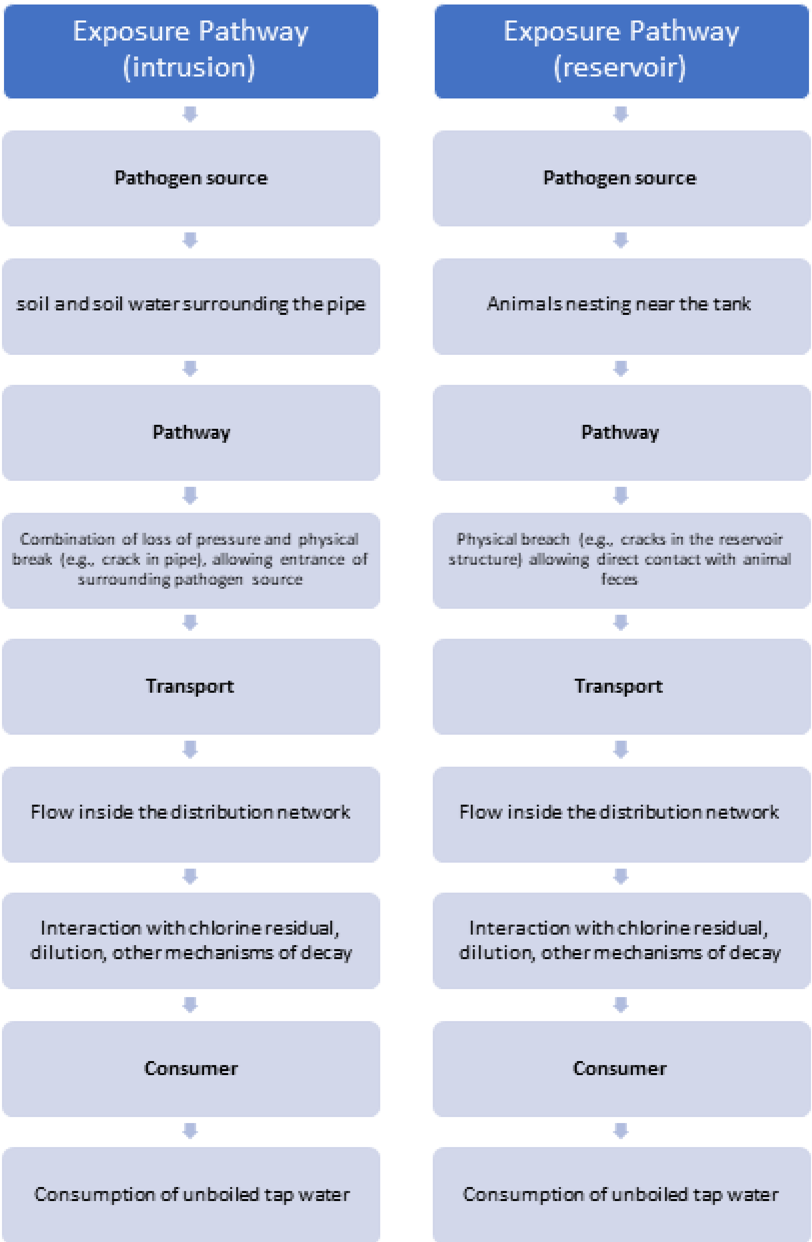

Figure 2. Example of exposure pathways for intrusion and reservoir.

3 Materials and methods

An overview of the work carried out is shown Figure 3. A literature review (**Paper I**) was conducted to identify the most relevant risks to address. Two different microbial risk events were simulated for this project: cross-connections and backflows, and intrusion. **Paper II** focused on a fault-tree analysis (FTA) method developed to estimate probabilities of infection (P_{inf}) using nationally aggregated and local data for three distinct risk levels. Additionally, an outbreak due to a cross-connection and backflow was simulated and used as validation for both the FTA and the outbreak simulation. **Paper III** focused on simulating intrusion due to a pipe break in a real WDN and was evaluated for multiple scenarios. **Paper IV** expanded on the intrusion case study by proposing a stochastic approach to estimating P_{inf}, in addition to providing additional sources of input data in the form of water quality analysis for indicators and metagenomics. **Paper V** introduced a risk management approach that combined QMRA with machine learning for predicting pipe failures and MCDA to support decision-making.

For the purposes of the WDN risk assessment, the health risk was calculated using Equation (6) presented in 2.6 Risk and QMRA framework.

The pathogen concentrations arriving at the consumer taps for the microbial risk events were used as the dose component in the dose-response modelling. All risk events were assessed for the same reference pathogens, i.e., *Campylobacter*, norovirus, and *Cryptosporidium*, using published dose-response relationships. The tolerable health risk was set to a daily target not exceeding 10^{-6}, based on the rationale of (Signor & Ashbolt, 2009) to use as a daily risk target rather than an annual target to capture potentially elevated risk events of short duration.

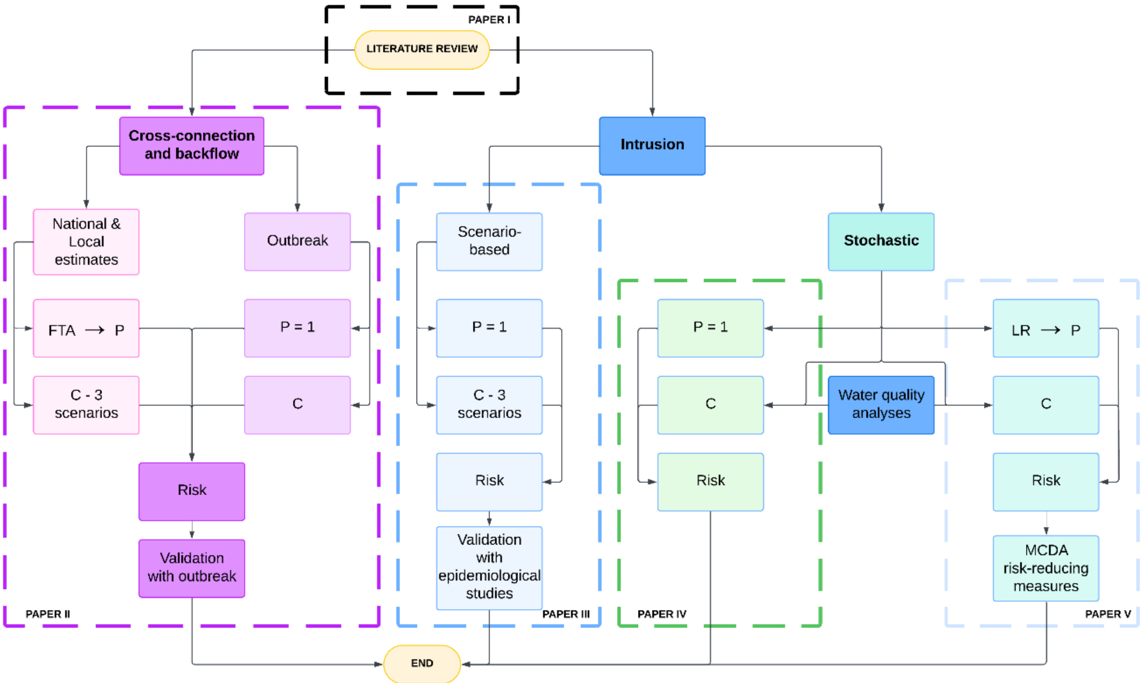

Figure 3. Overview for the workflow of each risk event and its connection to **Papers I-V**.

3.1 Prestudy

As a first step, a narrative review (reported in **Paper I**) was conducted to assess the state of the art on the health risks of WDNs and to identify potential areas where QMRAs could be improved. The literature review focused on two main areas: (1) RCTs and observational studies assessing the contribution of WDNs to the endemic health burden, and (2) QMRAs conducted in WDN. The literature search was conducted in three databases: Scopus, Web of Science and PubMed. The following keywords were used for the initial literature search: (1) 'drinking water' (2) 'distribution system', 'distribution network' (3) 'disease outbreak', 'gastrointestinal disease'. The keywords were combined in different ways using Boolean operators. Keywords in group (3) were truncated for some searches. The same search strategies were used for the three selected databases. After the initial search strategy in Scopus, the results were refined using additional keywords, e.g., water contamination, epidemic, risk assessment, etc. In addition, the reference lists of articles retrieved from the search were screened to identify additional potentially relevant records.

3.2 Study sites

In this thesis, two microbial risks were assessed: cross-connection and backflow, and intrusion. For cross-connections and backflows in **Paper II**, the FTA method was initially applied at the national level in Sweden and at the local level in Gothenburg. The Gothenburg network is located in south-west Sweden, spanning a total of 1 800 km. For further validation and comparison of the FTA predictions with traditional hydraulic modelling, a second network was selected (Figure 4). This network, situated in an undisclosed small municipality in Sweden, had previously experienced a waterborne outbreak associated with a cross-connection and backflow. Through PCR analysis, the causative agent was identified as norovirus. The network at this site spanned 39.7 km, comprising 81 nodes with pipe diameters ranging from 28 to 144 mm.

For intrusion calculations in **Paper III** - **Paper V**, different methods were tested in a WDN located in Scania County in the southern part of Sweden (Figure 5). This site comprised a specific section within a larger WDN with a total length of 13.5 km pipelines of varying diameters, ranging from 25 to 200 mm, and including 457 nodes. The annual water supply in the study area was 50 000 m^3. The total estimated water losses through leakage in the WDN accounted for 27% of total water production. The WDN has a total of 280 metered service connections serving approximately 714 persons, considering an estimated daily demand of 140 l/p/d in Sweden and excluding non-revenue water (water losses). The pipe network was laid between 1.5 m to 2 m below the ground surface within the same trench as stormwater and wastewater pipes. It was estimated that most sewer pipes were at the same elevation as drinking water pipes, with approximately 0.5 m distance in the horizontal direction between the pipes. Furthermore, it was estimated that 100% of the pipes in the WDN were below the groundwater level (GWL) based on studies of GWLs in the area.

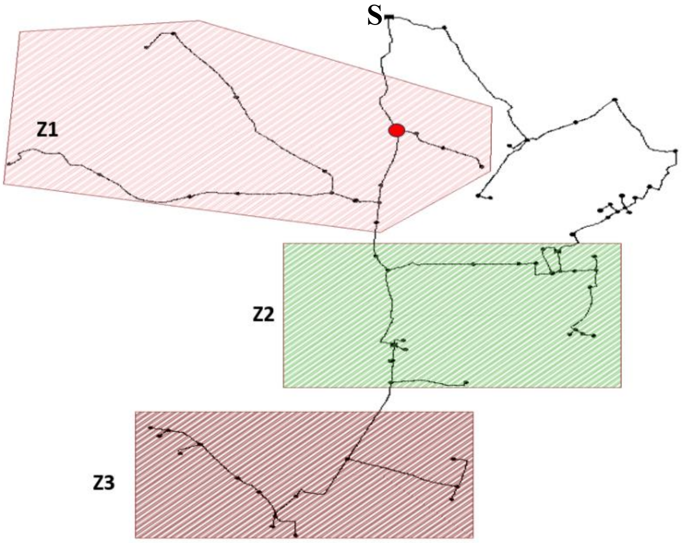

Figure 4. Schematic representation of the studied distribution network divided in zones Z1, Z2 and Z3. The contamination node (red dot) was located in Z1 and the water supply node (marked with an S) was located directly upstream of this node. The segment of the network not highlighted in any zone was discarded from the analysis, since the contamination did not reach the consumers in that area. For security reasons, the network scale and configuration has been distorted. Taken from **Paper II**.

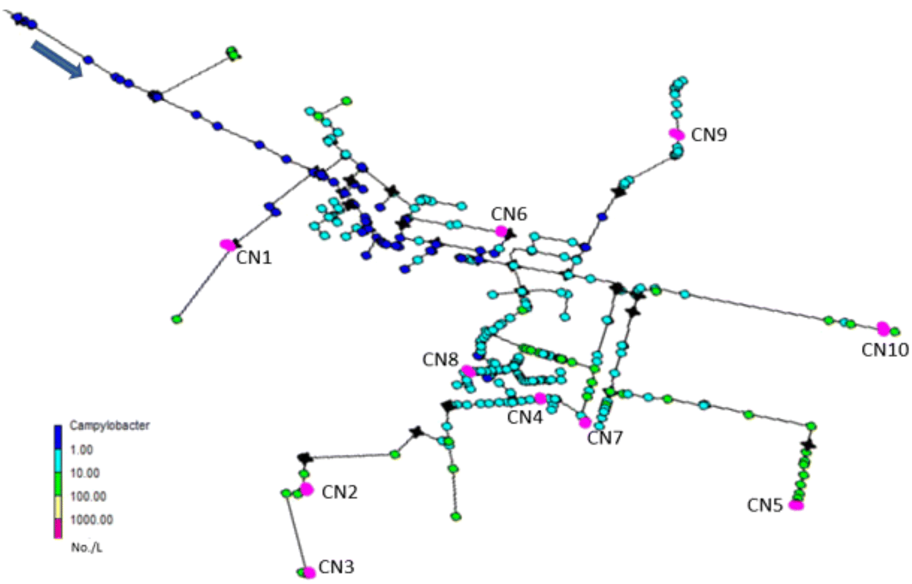

Figure 5. Case study section of the distribution network in Scania used for intrusion simulations. Critical nodes are shown in purple, whilst intrusion nodes are denoted by black diamonds. Taken from **Paper III**.

3.3 Probability of Failure

3.3.1 Fault tree for cross-connection and backflow failure

The probability of a contamination event due to cross-connection and backflows occurring was calculated using the FTA method developed in **Paper II**. This calculation was based on nationwide and local data from various sources (see Table 2). The following parameters were considered:

- **Total Length of Distribution Network in Sweden**: The Swedish Water & Wastewater Association (2000) reports a total length of 67,000 km.

- **Number of Swedish Municipalities**: According to the Government Offices of Sweden, there are 290 municipalities.
- **Average Length of Distribution Network per Municipality**: Calculated to be 231 km.
- **Contamination Events and Probabilities**:

  Disturbances Reported (2000-2008): Malm et al. (2010) documented 11 incidents, resulting in a probability of 1.82 x 10^{-5} km^{-1} yr^{-1} for endemic disease contamination events.

  Outbreaks Reported (1980-2009): Nine outbreaks were reported, with a probability of 4.48 x 10^{-6} km^{-1} yr^{-1} for contamination events leading to outbreaks.

  Severe Outbreaks (1980-2009): The probability of a contamination event leading to a severe outbreak was calculated to be 8.96 x 10^{-7} km^{-1} yr^{-1}.

  Gothenburg Specific Data: For the Gothenburg distribution network, the probability of a contamination event was found to be 3.31 x 10^{-8} km^{-1} yr^{-1}.

Table 2. Summary of national estimates and Gothenburg for probability of a contamination event (P) for the different cases.

Parameter	Value	Description
Total length of distribution network in Sweden	67 000 km	The Swedish Water & Wastewater Association (2000)
No. of Swedish municipalities	290	Government Offices of Sweden
Average length of distribution network/municipality	231 km	
No. of disturbances reported 2000-2008	11 incidents	Malm et al. (2010)
No. of disturbances reported 2000-2008 (km^{-1} yr^{-1})	1.82 x 10^{-5}	Probability of a contamination event for endemic disease
Outbreaks reported 1980-2009	9 outbreaks	Malm et al. (2010)
Outbreaks reported 1980-2009 (km^{-1} yr^{-1})	4.48 x 10^{-6}	Probability of a contamination event that leads to outbreak
Outbreaks reported 1980-2009 (km^{-1} yr^{-1})*	8.96 x 10^{-7}	Probability of a contamination event that leads to severe outbreak
Gothenburg (km^{-1} yr^{-1})	3.31 x 10^{-8}	Probability of a contamination event for Gothenburg distribution network. Calculation available in **Paper II** - supplementary data.

3.3.2 Logistic regression model for pipe failure

In the context of the intrusion simulations conducted in **Papers III** and **IV**, no probability of contamination was calculated, i.e., P was assumed to be 1. Conversely, in **Paper V**, a logistic regression model was used to estimate the likelihood of pipe breaks that could lead to intrusion. Logistic regression was mathematically formulated as in Equation (7) (Gasso, 2019):

$$p = \frac{1}{1 + e^{-(w_o + \sum_{i=1}^{m} w_i x_i)}} \tag{7}$$

In this equation, p is the output probability of each sample, x_i denotes the value of the i^{th} feature, w_i is the weight of the i^{th} feature and w_o is the bias constant. The weighted sum of the features, combined with the bias, is passed through the sigmoid function (p), which transforms it into a value between 0 and 1. The transformed value is then compared to the threshold value of 0.5, in order to determine the predicted class:

$$Class = \begin{cases} 0 \;\; if \;\; p \leq threshold \\ 1 \;\; if \;\; p > threshold \end{cases} \tag{8}$$

where less than or equal to 0.5 classifies the pipe as intact (0) and more than 0.5 classifies as a break (1).

The explanatory variables employed in the logistic regression model were selected from three distinct categories: physical, environmental and operational. An overview of the explanatory variables adopted for the model is provided in Table 3.

Table 3. Explanatory variables for the logistic regression model.

Variable	Name	Type	Description
Physical	Length	Numerical	Pipe length in m
	Diameter	Numerical	Pipe diameter in mm
	Age	Numerical	Pipe age in years
	Material	Categorical	Material of pipe section
Environmental	Elevation	Numerical	Elevation from mean sea level
	Soil type	Categorical	Major soil type at pipe location
	Roads and Trainlines	Categorical	Close proximity to the pipe
Operational	Pressure	Numerical	Average operational pressure
	Historical breaks	Numerical	Previous recorded breaks

The municipality provided physical pipe information and break records in EPANET and GIS formats, respectively. These datasets were then matched using QGIS geoprocessing features, and average operational water pressure was simulated over a 24-hour period in EPANET. Soil type data was obtained from the Geological Survey of Sweden (SGU). Categorical variables were created to account for the proximity of pipes to external infrastructure, such as road intersections and train lines, using GIS tools to generate buffer zones. These buffers were then overlaid with spatial data to establish relationships. Historical breaks were used to construct the target variable for predicting pipe failures over a 20-year period.

The dataset available for training represented the entire water distribution network (WDN). The logistic regression model was trained on the larger WDN, excluding the case study section, while the testing was carried out only on the case study section. Consequently, it was possible to test the model and evaluate its performance on data not encountered during the training session.

Data cleaning involved removing samples with missing factors and outliers. Numerical variables were standardised (adjusted to have a mean of zero and a standard deviation of one), and categorical variables were converted into a format suitable for analysis using a method called one-hot encoding (representing categories as binary vectors of 0s and 1s (James, 2013)). To address data imbalance, random oversampling was applied to the training dataset, enhancing the representation of the minority class (Werner de Vargas et al., 2023).

3.4 Contamination load

Each risk event (**Paper II - V**) required a different procedure to calculate the contamination load, i.e., the number of pathogens entering the network. This was due to, among other things, different faecal sources, different mechanisms causing the contamination event and the type of model used. A summary of the contamination loads used for each risk event and case is presented in Table 4.

Table 4. Summary of pathogens evaluated for each risk event, including contamination load assumed per scenario.

Event	Scenario	Reference pathogens	Contamination load (No. / L) [min, max]	Reference	Article
Cross-connection and backflow	Endemic	*Campylobacter*, norovirus, *Cryptosporidium*	71.5 (13, 130) 27.5 (5, 50) 2.53 (0.5, 5)	Henze (2001) Hewitt et al. (2011) Metcalf & Eddy (2003)	**Paper II**
Cross-connection and backflow	Outbreak (elevated)	*Campylobacter*, norovirus, *Cryptosporidium*	1 410 (130, 1 750) 577 (70, 700) 5.78 (0.7, 7)		**Paper II**
Cross-connection and backflow	Outbreak (extreme)	*Campylobacter*, norovirus, *Cryptosporidium*	20 062 (1 750, 25 000) 600 (500, 700) 8.5 (7, 10)		**Paper II**
Cross-connection and backflow	Outbreak	Norovirus	73.6 copies/min (55.6 – 83.33 copies/min)	Atmar et al. (2008)	**Paper II**
Intrusion	Pipe break, water shortage	*Campylobacter*, norovirus, *Cryptosporidium*	1.3 $\log_{10}$ / 100 mL 4.4 $\log_{10}$ / 100 mL* 0.3 $\log_{10}$ / 100 mL	Soller et al. (2010) Eftim et al. (2017) Ottoson et al. (2006)	**Paper III**
Intrusion	Pipe break	*Campylobacter*, norovirus, *Cryptosporidium*	Triangular distribution considering pathogen in WW, frequency, dilution factor	See Table 5 for more detailed information	**Paper IV** / **Paper V**

*1/1000 infectious gene copies (Seitz et al., 2011).

3.4.1 Pathogen load due to cross-connection and backflows

Three scenarios were considered in order to estimate the contamination load for use with the fault tree in **Paper II**: endemic risk level, elevated risk level and extreme risk level. The levels were taken from a sensor study carried out in Swedish networks (Jonsson et al., 2018). The different scenarios were based on an assumption of the amount of *E. coli* entering the distribution network:

1. Endemic risk level: 5-50 CFU/100 mL (0.005-0.05 % untreated wastewater)
2. Elevated risk level: 700 CFU/100 mL (0.7% untreated wastewater)
3. Extreme risk level: 5000 CFU/100 mL (10% treated wastewater)

The *endemic risk level*, abbreviated as F1 for national estimate and Got1 for Gothenburg estimate, represents the background level of risk present in the distribution network due to minor incidents. The *elevated risk level* (F2; Got2) represents an outbreak with the same magnitude of contamination as the Everöd outbreak (Kristianstads kommun, 2015). The *extreme risk level* (F3; Got3) was based on the Nokia outbreak (Laine et al., 2011).

Pathogen levels were estimated from the literature on the ranges present in raw and treated wastewater (Henze, 2001; Hewitt et al., 2011; Metcalf & Eddy, 2003)(see Table 4). The reference pathogens chosen were *Campylobacter*, norovirus, and *Cryptosporidium*. These are the most relevant enteric pathogens for Swedish networks and represent bacterial, viral, and protozoan organisms (Abrahamsson, 2009).

To account for uncertainty, pathogen concentrations were treated as a triangular distribution (min, mode, max). The minimum and maximum values were taken from the range reported in the literature (if available) and the most likely value (mode) was the mean of the range. If the pathogen concentration in the literature was given as a point estimate, the maximum value from the less critical scenario was used to generate the range, e.g., the minimum concentration for *Campylobacter* in the elevated case was assumed to be the same as the maximum value in the endemic case. The mode was then the mean of the range multiplied by 1.5 to account for the skewness towards the maximum in the effluent. The pathogen concentrations for each risk case used for the QMRA are shown in Table 4.

3.4.2 Norovirus load during outbreak due to cross-connection and backflow

The contamination load for the outbreak case study in **Paper II** was only assessed for norovirus. To estimate the norovirus load, virus shedding in faeces, 10^9 norovirus/g of faeces (min-max: 1×10^5 – 1.64×10^{12} copies/g of faeces), was taken from a human experimental study (Atmar et al., 2008). The amount of stool produced by a person in one day was 106 g (min-max: 80 g – 120 g) (Cummings et al., 1992). The median load at the septic tank outlet was 1.06×10^{11} copies/day; the minimum load was 1.06×10^7 copies/day; the maximum load was 1.74×10^{14} copies/day.

The norovirus load was then applied a reduction factor, ranging between 10^{-6} and 10^{-3}(Åström et al., 2016). The reduction factor represented the transport of pathogens from the septic tank through the soil material to the private well, from which the contaminated water was fortuitously pumped into the distribution network. Due to the considerable uncertainty associated with the pathogen transport from the septic tank to the WDN, each reduction factor was considered a different scenario (scenarios C1-C4). For scenario C1, the norovirus load intruding into the network was 73.6 copies/min (55.6 – 83.33 copies/min). For each subsequent scenario, the norovirus load increased tenfold (equivalent to a 1 $\log_{10}$ increase).

3.4.3 Scenario-based estimation of intrusion loads

In **Paper III**, the intrusion of soil water into the WDN was modelled with the aid of three distinct scenarios. The assumed layout of the distribution network pipes when generating these scenarios is shown in Figure 6. The three scenarios for intrusion evaluated in **Paper III** were:

- Scenario 1: Pipe Break – Without Leakage ($PB_{\text{No leakage}}$)

In this scenario, a pipe break was assumed to occur on the first pipe branch from the main distribution network to the selected area. Excess pressure in the WDN network was modelled by adjusting the status of the affected pipe. The status of the affected pipe was set to closed for 5 hours, and the lowest pressures observed at every node were recorded. Intrusion was assumed to occur only at the repair section, with no leakage in the rest of the WDN due to the assumption of new pipes without physical deficiencies. The intrusion volume was estimated by determining the volume of the affected section of the WDN given the diameter and length of the section.

- Scenario 2: Pipe Break – With Leakage and Varying Intrusion Potential (PB_{Leakage})

In the event of a main pipe break, the simulation was done by setting an extra demand of 20 l/s to the end node of the broken pipe to simulate high free flow due to the pipe break. The flow was set to last for 2 hours to reflect the leakage duration before it was identified, and the affected pipe was shut off. The affected pipe was then closed for an additional 3 hours to simulate the repair duration. The lowest pressures (H_p) at every node were extracted and used for determining the intrusion potential (ΔH_{int}). This procedure was repeated to test the effect of different model variables.

- Scenario 3: Insufficient Water Supply

Insufficient water in the WDN due to water shortage or process failures at the drinking water treatment plant was simulated using a flow control valve located downstream of the reservoir. The valve was run only during the simulation period of failure, between 06:00 and 11:00. Different flow settings across the valve were tested, and it was determined that flows of 2.5 l/s or lower would result in low pressures in the study area. The valve setting was thus set to 2.5 l/s, and the resulting lowest nodal pressure distributions were extracted and used for determining the intrusion volume.

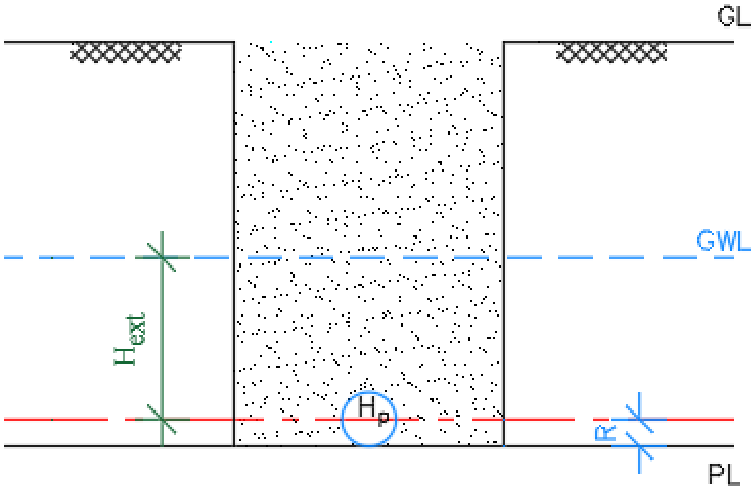

Figure 6. The position of a WDN pipe (cross-sectional view) in relation to the ground level (GL) and groundwater level (GWL) where: H_{ext} represents the external piezometric head (elevation of groundwater - surface), PL is the pipe level, R is the pipe radius and H_p is the internal piezometric head. GWL is in meters above sea level (m.asl) while PL, and GWL are measured from the GL. Taken from **Paper III** – Supplementary information.

The unit intrusion volume was determined using the orifice equation as shown in Equation (9):

$$Q_{int} = A_{tot} \cdot C_d \cdot \sqrt{2g \cdot \Delta H_{int}} \tag{9}$$

where:

- Q_{int} is the intrusion volume per time and length units,
- A_{tot} is the total area of all holes on the studied section of the WDN,
- C_d is a coefficient dependent on the sharpness of the edge of the hole,
- g is the gravitational acceleration,
- ΔH_{int} is the intrusion potential.

The total area of the leakage holes (A_{tot}) on every pipe was determined by re-arranging the orifice equation:

$$A_{tot} = \frac{Q_{out}}{C_d \cdot \sqrt{2g \cdot \Delta H_{normal}}} \tag{10}$$

Intrusion volume into the WDN was calculated for every affected pipe section and node in all scenarios. Sections of the WDN with similar characteristics were grouped, and the intrusion of each section was lumped into one node per section. A total of 32 intrusion nodes were identified for Scenarios 2 and 3.

Pathogen intrusion into the WDN was determined based on the intrusion volume of soil water and assumptions on pathogen concentrations in the intruded water. The pathogens selected were

Campylobacter, norovirus, and *Cryptosporidium*. Several dilution factors were used to estimate pathogen concentrations in soil water based on field studies and literature values. The average concentrations of the studied pathogens in untreated wastewater were used to calculate the numbers of pathogens entering the system. These numbers were adjusted with a reduction factor to reflect contamination transport through the soil.

3.4.4 Stochastic estimation of intrusion loads

In **Paper IV**, the intrusion loads were calculated using a different method to that employed in **Paper III**, although the same WDN was used in both cases. The Water Network Tool for Resilience (WNTR) (Klise et al., 2020) was utilized to iterate over the most critical pipes in the network and simulate pipe breaks for each. WNTR is a hydraulic modelling package developed in Python, which allows for greater versatility when, for example, generating contamination scenarios. For each pipe break iteration, low pressures experienced in the WDN nodes were logged. This process allowed for the creation of a Pandas DataFrame containing relevant information such as the name of the broken pipe, time of break, leak area and resulting low pressures in the network for a whole 24-hour period.

To estimate the intrusion load, the pressure results were combined with estimations of pathogen concentrations from water sampling results and literature values. A summary of the input parameters for intrusion load estimation is presented in Table 5. The estimation process involved over 1200 iterations, incorporating varying pipe breaks and time of break, in addition to a range of dilution factors and pathogen levels in faeces.

Table 5. Summary of input parameters for estimation of pathogen concentrations for intrusion simulation.

Parameter	**Distribution / Model**	**Comment**
Pathogen in faeces	*Campylobacter* Triangular (0, 20, 200) norovirus* Triangular (1000, 34000, 1000000) *Cryptosporidium* Triangular (0.5, 10, 398)	(Soller et al., 2010)
Dilution factor	Uniform (2 $\log_{10}$; 4 $\log_{10}$ size: 1000)	Based on *E. coli* results from sampling
Frequency	Uniform (0.1, 1 size: 1000)	Based on the detection of *E. coli* & pathogens from sampling
Intrusion volume	Hydraulic model	Calculated using Module 1 – pipe break from **Paper IV** – supplementary material

*Norovirus concentrations were further reduced by an infectivity coefficient, assuming 1 in 1000 norovirus are infectious (Seitz et al., 2011).

3.5 Contamination transport

Once the contamination loads had been calculated as outlined in Section 3.4, the available WDN models were employed to simulate the contamination transport within the networks. This facilitated the quantification of contamination levels (outbreak simulation in **Paper II** and intrusion simulations in **Papers III-V**) reaching the consumers' taps. For both WDN case studies, the model included demand patterns and pressure levels at the supply node, representing expected operational conditions relevant to each scenario. For the outbreak calculation in **Paper II**, the cross-connection was assumed to last for 1 hour per day, for 9 to 10 days. In the hydraulic model, the quality time step was set to 1 minute, and the simulation run for 168 hours. For the intrusion calculations in **Paper III**, the duration of the contamination event varied between 2 to 8 hours, depending on the scenario being evaluated.

The simulation was also run for 168 hours. For **Paper IV** and **V**, the intrusion duration was set to 1 hour, quality time step of 5 minutes and the simulation run for 72 hours.

3.5.1 Dose-response and risk estimates

This subsection describes the methodology used to calculate dose-response and risk estimates for the contamination events simulated in **Papers II-V**. The resulting pathogen concentrations arriving at the consumers' taps were combined with Swedish consumption studies to calculate a dose to use for the dose-response modelling. The dose was calculated as:

$$D = C_{node} \cdot V \tag{11}$$

where D is the dose, C_{node} is the pathogen concentrations at the consumer tap and V is the volume of water ingested per person.

The volume of water ingested was determined from two different consumption studies available for Sweden: a log-normal distribution used in **Paper II** (Westrell et al., 2006) and a gamma distribution used in **Papers III-V** (Säve-Söderbergh et al., 2018).

The cross-connection and backflow contamination event using FTA (**Paper II**) and all intrusion contamination events (**Papers III-V**) were assessed for the same reference pathogens, while only norovirus was evaluated in the outbreak simulation in **Paper II**. The Exact Beta-Poisson model was used to assess the dose-response for each reference pathogen. Using this dose-response model, the probability of infection was calculated with:

$$P_{inf} = 1 - exp^{-r \cdot D} \tag{12}$$

where P_{inf} is the probability of infection, r is a sample from the Beta distribution with α, β parameters for each pathogen and D is the dose ingested.

The Beta distribution parameters for each reference pathogen were: *Campylobacter* (α = 0.024, β = 0.011) (Teunis et al., 2005), norovirus (α = 0.04, β = 0.055) (Teunis et al., 2008b) and *Cryptosporidium* (α = 0.115, β = 0.176) (Teunis et al., 2002). The population was assumed to be homogenous, i.e., all consumers would respond according to the chosen dose-response function.

In **Paper II**, the daily probability of infection (C) was calculated by combining P_{inf} with cross-connection and backflow events of durations varying between 1 to 24 hours. For intrusion events in **Paper III** to **V**,

$$C = 1 - \left(1 - P_{inf}\right)^{n} \tag{13}$$

where C is the daily probability of infection, P_{inf} is the probability of infection for an event of n duration and n is duration of the contamination event expressed in days.

In the outbreak simulation presented in **Paper II**, an additional step was taken to estimate the amount of symptomatic illnesses based on the daily risk of infection from norovirus estimated using both FTA and the hydraulic simulation of the outbreak. The proportion of asymptomatic infections due to norovirus GII.4 was assumed to be 40.7% (95% CI: 32.8%–49.0%), taken from an outbreak study in Japan (Miura et al., 2018). The probability of developing symptoms when infected, $P_{illness}$, was calculated as:

$$P_{illness} = P_{(illness|infection)} \cdot C \tag{14}$$

where $P_{illness}$ is the probability of illness, $P_{(illness\,|\,infection)}$ is the conditional probability of developing the illness after infection, and C is the daily probability of infection.

3.6 Water sample collection and analysis

For **Paper IV**, water samples were collected between 2018 and 2022 during scheduled and unplanned pipe repairs in a large Swedish drinking water distribution network.

Plastic bottles were autoclaved at 121°C for 21 minutes before use in the field. The minimum amount of water collected was 100 mL for coliform enumeration and 500 mL for DNA extraction. Depending on the size of the repair pit, two or more samples were collected from the same site and considered different sample points. Coliform enumeration was carried out within 4 hours of collecting the sample.

Total coliforms and *Escherichia coli* concentrations were determined using the Colilert® test (IDEXX Laboratories, Westbrook, Maine). Highly turbid samples were diluted (1:10 to 1:100 dilution) using autoclaved Milli-Q water. The samples were incubated at 36°C for 24 hours; with an additional incubating time of 1 – 4 hours when wells were uncertain.

Table 6. List of waterborne pathogens run through genomic databases to identify if their sequences were present in the samples collected.

Pathogen	Pathogens of interest (common)	Pathogens of interest (emerging)
Bacteria	*Vibrio cholera,* serovarieties O1 and O139	*Helicobacter pylori*
	Salmonella spp.	*Aeromonas hydrophila*
	Shigella spp.	*Leptospira* sp.
	Escherichia coli (possibility to distinguish haemorrhagic serotypes)	
	Campylobacter spp.	
	Legionella pneumophila	
	Mycobacterium avium complex (MAC)	
	Pseudomonas aeruginosa	
Virus	Adenovirus	*Alphatorquevirus*
	Astrovirus	*Cyclovirus*
	Norovirus	*Erythroparvovirus*
	Hepatovirus A	*Bocaparvovirus*
	Rotavirus	*Protoparvovirus*
Protozoa	*Cryptosporidium* sp.	*Microsporidia*
	Giardia intestinalis	*Cyclospora cayetanensis*
	Entamoeba histolytica	*Cystoisospora belli*
	Toxoplasma gondii	
	Naegleria fowleri	
	Acanthamoeba spp.	
Helminths		*Schistosoma* sp.

Microorganisms in the water samples were concentrated by filtering 25-255 mL of water through 0.2 µm pore size cellulose membranes. Subsequently, the membranes were preserved at -80°C, for a maximum of six months, until DNA extraction. The FastDNA™ Spin Kit for Soil (MP Biomedicals, USA) was used according to the manufacturer's instructions to isolate DNA from the particulates attached to the membranes. The extracted DNA was then stored at -20°C, and the samples with DNA concentrations greater than 8 000 ng/ml were sequenced. Library preparation and 150 bp paired-end shotgun sequencing using Illumina's NovaSeq 6000 system was carried out by Eurofins Genomics. The raw sequence reads were deposited at the NCBI short read archive with accession number PRJNA993188. The raw sequence reads were quality filtered using fastp v0.23.2 (Chen et al., 2018) and the microbial community composition of the samples was analysed using MetaPhlan v4.0.6 (Blanco-Míguez et al., 2023). Diversity of taxa in a sample and dissimilarity in community composition between samples were calculated with the mathematical framework based on Hill numbers (Chao et al., 2014) using qdiv v2.2.1 (Modin et al., 2020).

A comprehensive list of pathogens (shown in Table 6) was compiled from the literature to identify through common genetic sequence databases. (Magana-Arachchi & Wanigatunge, 2020; Ramírez-Castillo et al., 2015; WHO, 2022).

3.7 Risk reduction measures and stakeholder preferences

In this thesis, work relating to risk reduction measures was undertaken in **Paper III** and **Paper V**. In **Paper III**, a range of risk reduction measures were subjected to evaluation in order to ascertain the viability of managing the infection risks associated with the simulated intrusion scenario. The default scenario to use as baseline was Scenario 2, as described in Section 3.4.3. Model parameters that were modified to represent the implementation of risk reduction measures included increasing the $\log_{10}$ reduction when estimating the contamination load intruding into the WDN, reducing the leakage rate by 50% and 75% as well as reducing the repair time for the pipe break to 2 hours.

In order to ascertain potential measures for implementation as risk reduction strategies, a review of the literature was undertaken, and discussions were held with relevant stakeholders.Ten industry experts were selected for the study, including water specialists, utility managers and water supply consultants. These individuals were selected on the basis of their experience and knowledge in the maintenance and rehabilitation of water distribution systems.Informed consent was obtained from all study participants.

To facilitate the evaluation process, stakeholders were provided with a spreadsheet to rate each strategy against the evaluation criteria on a scale of 1-5, with 1 being the least favourable and 5 being the most favourable.The expert ratings were used to create a decision matrix, where each row represents a strategy, and each column represents an evaluation criterion. The alternatives and evaluation criteria considered are presented in Table 7 and Table 8.

The TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method was used to rank and prioritize these strategies (Park et al., 2023). The TOPSIS method uses an Ideal Solution (IS) and a Negative Ideal Solution (NIS) to find the best alternative. The best alternative is closest to the IS and farthest from the NIS, and the Euclidean distance is used to measure the distance from both solutions. Higher scores indicate the most favorable strategies according to the evaluation criteria.

Table 7: Different alternatives for reducing risks in water distribution systems

Strategies	**Strategy description**
Pressure management	Regulate water pressure through pressure-reducing valves.
Pipe repair and rehabilitation	Conduct repairs and apply techniques such as pipe lining.
Pipe replacement	Replace old or deteriorating pipes with new ones.
Increase inspection and testing	Enhance the frequency and thoroughness of pipe inspections.
Increase metering	Install additional meters to better monitor water flow and usage.
Active leakage control	Implement strategies for continuous measurement and management of leaks, including the use of specialized tools.
Cross connection control	Implement measures to identify and manage unauthorized or improper cross-connections.
Public awareness and engagement	Educate and engage the community on the importance of leak reporting.

Table 8: Evaluation criteria for the different strategies.

Evaluation Criteria (C)	**Criteria description**
Cost	The financial expenditure required to implement and maintain the strategy.
Executability	The ease and feasibility of implementing the strategy within the existing infrastructure.
Risk reduction	The extent to which the strategy lowers the overall risk.
Social concern	The strategy’s impact on the community and stakeholders (social, environmental and public health).
Reliability	The consistency and dependability of the strategy over time.

4 Results and discussion

Key results from the different risk event calculations are presented in the following chapter. The results are interpreted in the larger context of microbial risk assessments in drinking water distribution networks.

4.1 Risk estimates

From the literature review conducted in **Paper I**, two microbial risks were identified to be investigated in this thesis: cross-connection and backflows in **Paper II**, and intrusion events in **Paper III-V**. In **Paper II**, both national (Sweden) and local (Gothenburg) risk estimates of cross-connection and backflow events were calculated using FTA. Figure 7 illustrates the estimated national daily risks of infection for endemic, elevated and extreme cases. The median values of daily risk for *Campylobacter* ranged between 10^{-4} and 10^{-2}, for norovirus between 10^{-5} and 10^{-3}, and for *Cryptosporidium* between 10^{-5} and 10^{-4}. The greatest spread was observed for norovirus in all three cases, while the largest proportion of values close to the absolute maximum risk of infection was seen for *Campylobacter*. *Campylobacter* also demonstrated the least spread of all the cases studied. Surprisingly, the national daily risk of infection was highest for the endemic case (F1) compared to the elevated (F2) and extreme (F3) cases for all reference pathogens. This may be explained by the higher frequency of incidents in the network than reported outbreaks, i.e., higher P for incidents that do not seem to cause outbreaks.

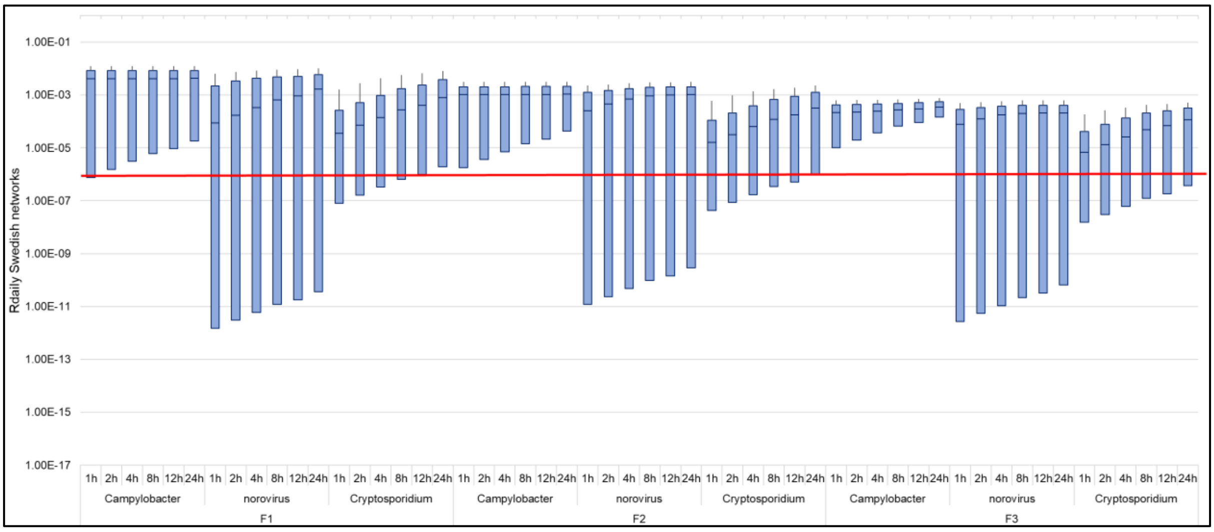

Figure 7. Box plots for daily risks of infection from cross-connection and backflows in an average Swedish distribution network (**Paper II**). The red line is the target value for daily risk of infection (10^{-6}). Times (in hours) on the x-axis represent the different durations of contamination events considered for the calculations, e.g., 1 h is a cross-connection and backflow event lasting 1 h. The bottom and top of the box are the 25th and 75th percentile of the values, respectively. The line inside the box is the median. The upper whisker shows the largest value (excluding outliers), and it is calculated as 1.5*IQR. The value for the bottom whisker is 0 (not shown).

It is not uncommon for local conditions to vary considerably between networks, despite their comparable size. This is reflected in the evaluation of the Gothenburg system using FTA, whereby the infection risks for all scenarios were found to be below the acceptable target of 10^{-6} (Figure 8). This finding is inconsistent with the anticipated outcome based on the national estimates, as the assumption of a positive correlation between network length and incident probability would suggest otherwise. As the national data combines networks with varying degrees of functionality, averaging their performance, distribution networks that are considered well-functioning (such as in Gothenburg) would require local risk assessments to obtain a realistic outcome from the analysis.

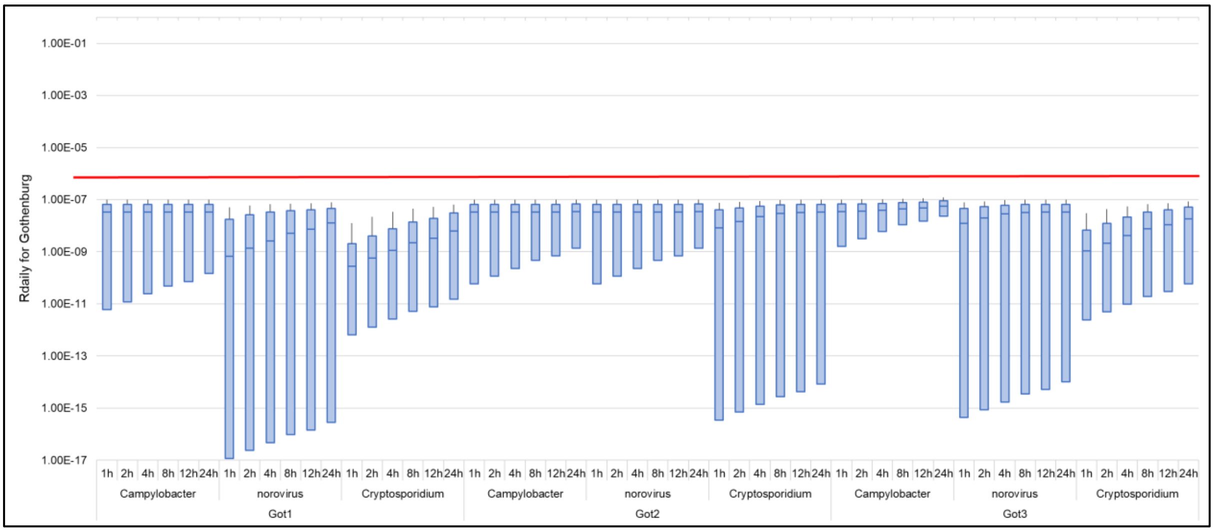

Figure 8. Box plots for daily risks of infection from cross-connection and backflows in the Gothenburg distribution network (**Paper II**). The red line is the target value for daily risk of infection (10^{-6}). Times (in hours) on the x-axis represent the different durations of contamination events considered for the calculations, e.g., 1 h is a cross-connection and backflow event lasting 1 h. The bottom and top of the box are the 25th and 75th percentile of the values, respectively. The line inside the box is the median. The upper whisker shows the largest value (excluding outliers), and it is calculated as 1.5*IQR. The value for the bottom whisker is 0 (not shown).

For the outbreak case study, the calculated daily risks of infection for the scenarios simulated using hydraulic modelling are shown in Figure 9. For scenarios C1 and C2, the daily risk of infection was in the order of 10^{-2} and 10^{-1}, respectively; and for scenarios C3 and C4, the median value for risk of infection was 1 (i.e., 100% infected).

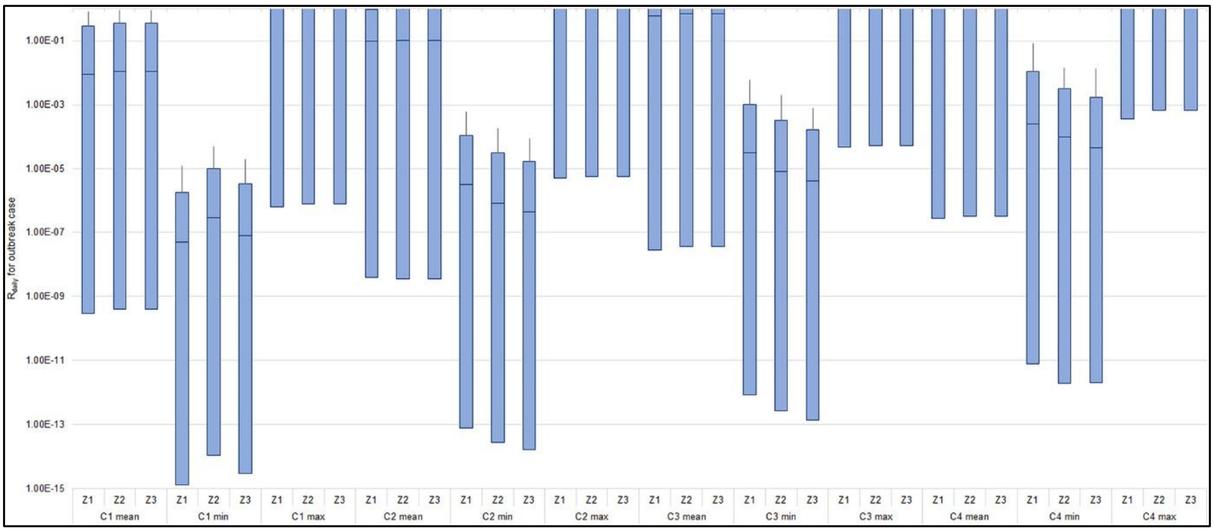

Figure 9. Daily risk of infection of norovirus during the studied outbreak for each zone (Z1-Z3) and scenario (C1–C4). Mean, min and max refer to the mean, minimum and maximum levels of contamination that intrude the network during the hydraulic simulation. The bottom and top of the box are the 25th and 75th percentile of the values, respectively. The line inside the box is the median. The upper whisker shows the largest value (excluding outliers), and it is calculated as 1.5*IQR. The value for the bottom whisker is 0 (not shown). Taken from **Paper II**.

In **Paper III**, the estimated 75th percentile of the daily risk of infection due to intrusion (Figure 10a-c) was above the target value of 10^{-6} for all scenarios, all pathogens, and all tested dilution effects between pathogen concentration in wastewater and soil water (with some few exceptions for *Cryptosporidium* and 4 $\log_{10}$ unit reduction). While the risks for all pathogens were considered unacceptable, *Campylobacter* was observed to have the highest risks, and norovirus the lowest.

As illustrated in Figure 11a, the probability of infection exhibited a notable decline, accompanied by a simultaneous reduction in the leakage rate. This resulted in a reduction in the total area of leakage holes, which in turn led to a decrease in the intrusion volume during low/no pressure events. Nevertheless, it is evident that even minimal intrusion volumes within the WDN can still give rise to unacceptable risks during low-pressure events (Figure 11a). The probability of infection exceeded the health target limit of 10^{-6} even when the leakage rate and, consequently, the intrusion volume were reduced from the initial value (27%) to half (14%) and then to a quarter (7%).

Regarding the impact of the duration of the pipe break event, it was observed that in this section of the WDN, the intrusion was more significantly influenced by the time period during which the event was detected than by the actual time the repair work of the pipe was undertaken. This was due to the low flow observed when the downstream WDN was isolated from the main water supply. The observation of different durations of low-pressure events in scenario 2 revealed a proportional relationship between the probability of infection and the duration of the low-pressure event (Figure 11b).

In **Paper IV**, the calculated infection estimates (Figure 12) generally indicated a lower P_{inf} in comparison to the risk estimates obtained in **Paper III** (Figure 10). This discrepancy can be attributed to several factors inherent to the model setup.

The new stochastic intrusion model built in the Python environment (**Paper IV**) incorporates the possibility for multiple pipes to fail, thus representing a more comprehensive approach than that presented in **Paper III**, where only one major pipe was at risk of breaking. This allowed for the inclusion of pipe failures that may have a less pronounced impact on the critical nodes in the P_{inf} estimates. For instance, an isolated branch situated downstream of the critical node may be considered in multiple iterations of the new simulations. Conversely, pipes that may have a more direct impact on the node, such as pipes in close proximity to the node, were also considered. Since the choice of pipe is randomised, it is expected that the variation in risk will reflect whether the node is in a more sensitive area of the network or if it is only directly impacted by a few select pipes.

Additionally, in **Paper IV** the variability in the level of contamination present in the soil water surrounding the pipe was introduced as probability distributions rather than as point estimates as in **Paper III**. The parameters employed to estimate the level of contamination comprised probability distributions for the typical concentrations of pathogens in wastewater, the dilution factor for the wastewater in soil based on the results of sampling and metagenomics, as well as the prevalence of pathogens in the soil water based on metagenomics. It is possible that this approach may have introduced a degree of bias in the concentration estimates. Nevertheless, the application of randomisation should serve to offset any potential bias and instead result in concentrations that are more reflective of the conditions typical of distribution networks in Sweden.

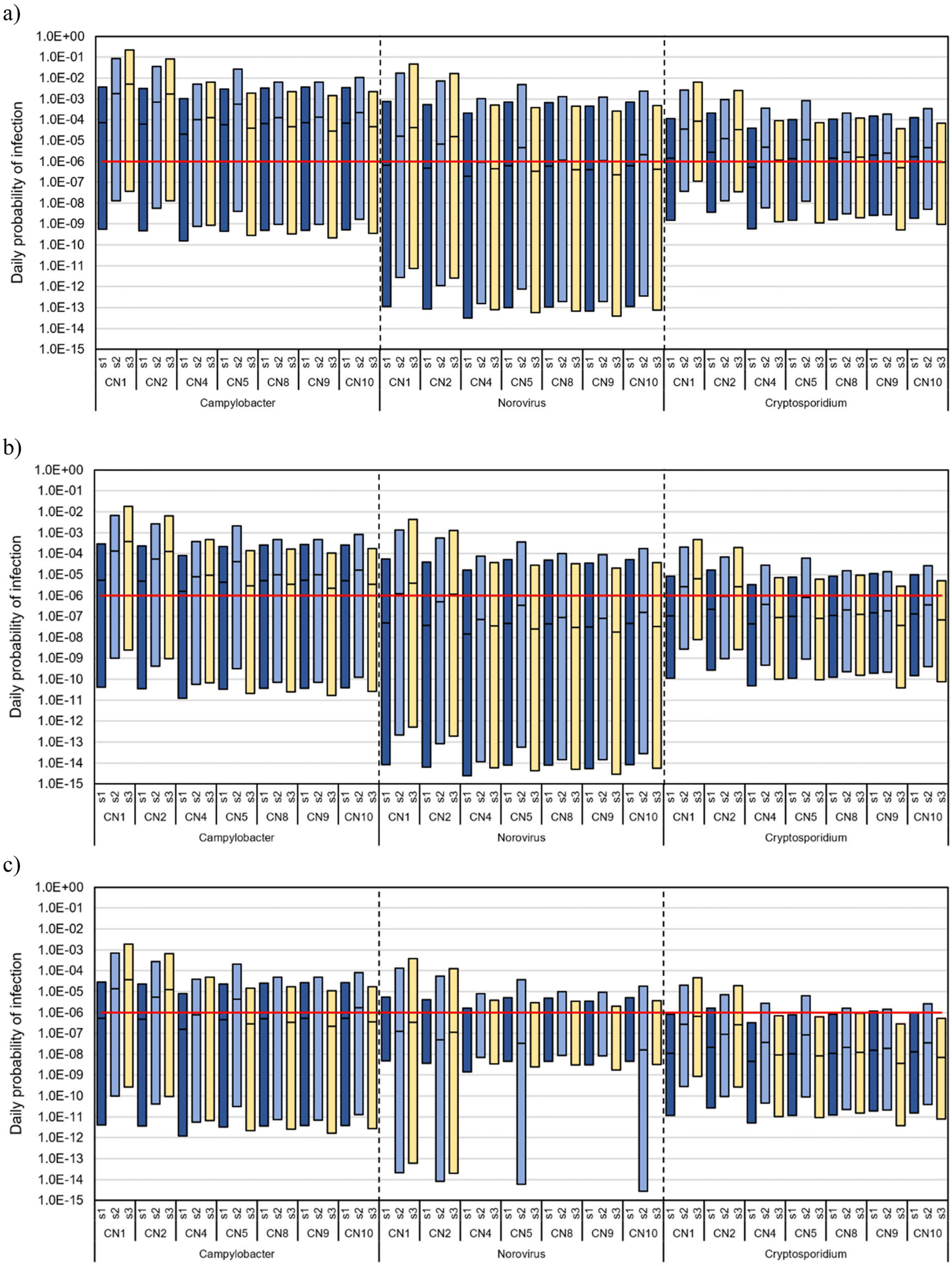

Figure 10. Box plot for daily probability of infection for selected critical nodes (CN1–10) for *Campylobacter*, norovirus, and *Cryptosporidium* across three scenarios: 1 (dark blue), 2 (light blue), and 3 (yellow). The plots labelled a, b, and c represent a reduction/dilution factor of 2, 3, and 4 $\log_{10}$ units, respectively, between the pathogen concentrations in wastewater and soil water. The target limit for the daily probability of infection (10^{-6}) is represented by the red line, while the 25th percentile is represented by the bottom line of the box, the 50th percentile is the line in the box, and the 75th percentile is the top line of the box. Taken from **Paper III**.

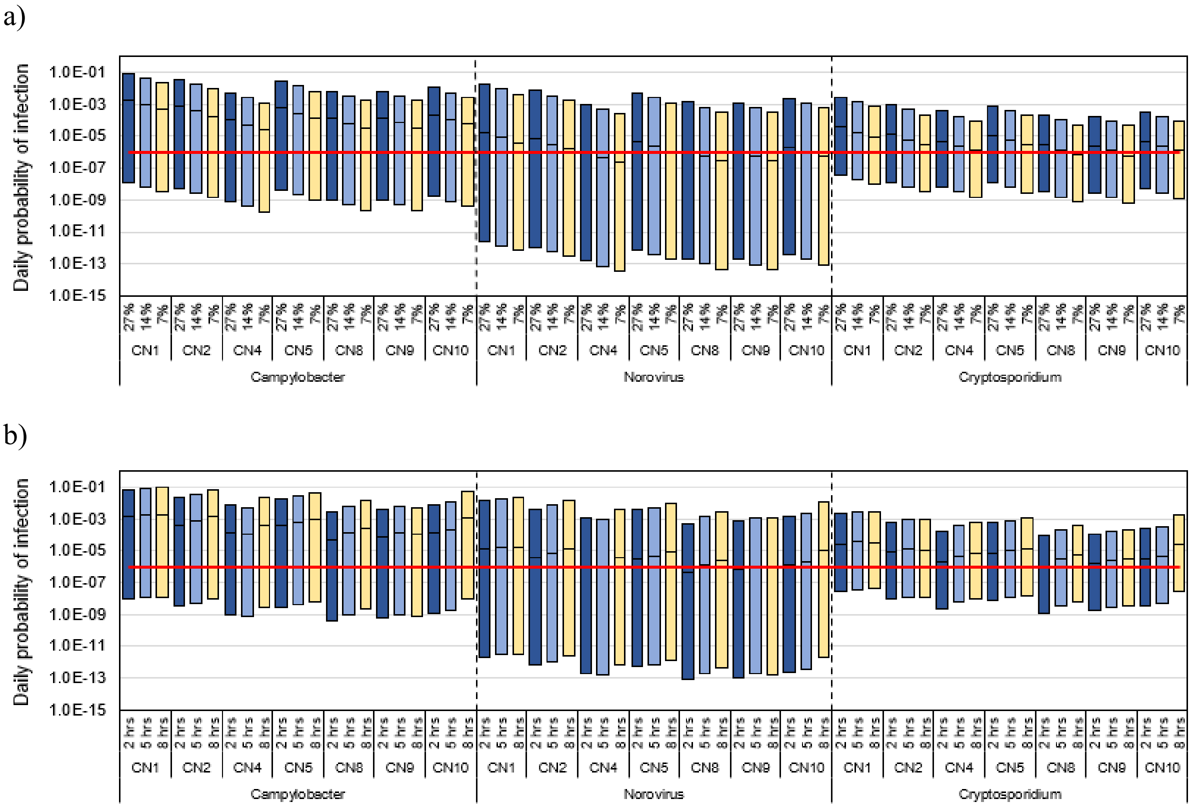

Figure 11. Box plot for daily probability of infection considering a) different leakage rates (27% – dark blue, 14% – light blue, 7% – yellow) and b) different event times (2 h – dark blue, 5 h – light blue, 8 h – yellow) for selected critical nodes (CN1–10) for the pathogens *Campylobacter*, Norovirus, and *Cryptosporidium* for scenario 2 (2 $\log_{10}$ units reduction/dilution factor between the pathogen concentrations in wastewater and in soil water). The target limit for the daily probability of infection (10^{-6}) is represented by the red line, while the 25th percentile is represented by the bottom line of the box, 50th percentile is the line in the box, and 75th percentile is the top line of the box. Taken from **Paper III**.

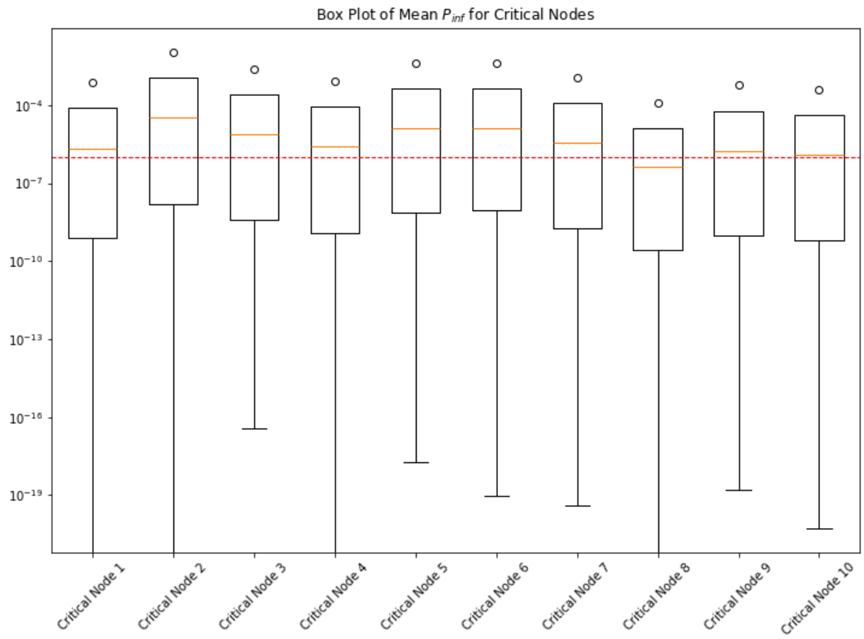

Figure 12. Box plots of P_{inf} from *Campylobacter* for stochastic simulations of intrusion events using faecal indicator/metagenomics results from the sampling campaign. Taken from **Paper IV**.

The stochastic intrusion model was enhanced through the incorporation of a logistic regression model, which was utilised to calculate the probability of pipe failures and to supplement the health risk calculations in **Paper V**. The predictions from the aforementioned model are illustrated in Figure 13. Integration of the raw probability of infection estimates calculated using the WNTR modules developed in **Paper IV** with these probabilities of failures resulted in the generation of a daily risk of infection map (Figure 14). The WDN segments marked in red represent those that on average exceed the acceptable daily risk target during the entirety of the simulation. It is worth noting that the pipe break studied in **Paper III** - Scenario 1 does not seem to have elevated infection risks, due to the low probability of pipe failure. The highest risks are concentrated in the central section of the WDN and in some of the branches.

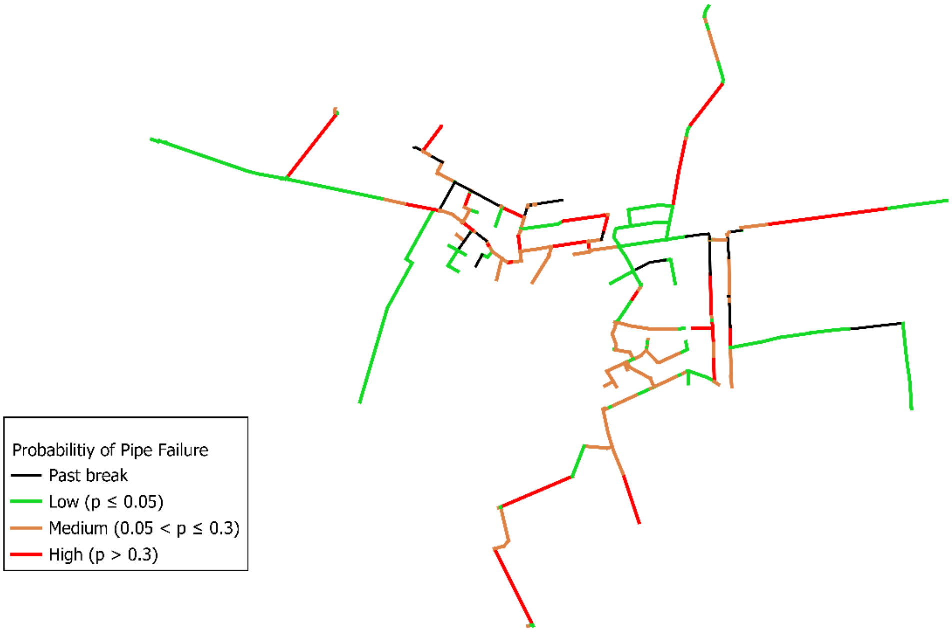

Figure 13. Map of the WDN in Scania County showing the probability of a pipe break calculated with the logistic regression model from **Paper V**. Pipes with low probability of failure (≤ 0.05) are shown in green, pipes with medium probability (0.05 < p ≤ 0.3) are shown in orange and pipes with high probability of failure (p > 0.3) are shown in red.

Figure 14. Map of the WDN in Scania County showing pipe segments with a daily risk of infection above 10^{-6} in red, while pipe segments under 10^{-6} are shown in green. The daily risk of infection was calculated by using the raw probability of infection values and combining them with the probability of failure. Taken from **Paper V**.

4.2 Effect of measures and stakeholder preferences

Table 9 illustrates the impact of varying specific inputs within the model, which may be regarded as analogous to the implementation of risk-reducing measures. For instance, an increase in the $\log_{10}$ reduction of contamination in the vicinity of the pipes may be indicative of an increase in wastewater network renewal, whereby leakages from wastewater pipes are reduced. Another potential measure is the implementation of improved physical separation between drinking water and wastewater pipes. The largest reductions in median P_{inf} were observed for these two scenarios. An increase in the log reduction from 2 to 4 $\log_{10}$ resulted in a notable reduction in the median P_{inf} for all critical nodes (CN) and all pathogens. It is worth noting that the reduction in P_{inf} was proportional to the magnitude of the $\log_{10}$ reduction, i.e. an additional $\log_{10}$ reduction in pathogen concentration resulted in a 1 $\log_{10}$ reduction in P_{inf}. This would be in-line with an epidemiological study conducted during WDN repairs in Sweden, where it was found that the presence of sewage lines at the same level as the drinking water pipe was associated with an increased risk of GI (Säve-Söderbergh et al., 2017).

Two further scenarios were also evaluated, in which the leakage rate was reduced to 14% and 7%. The reduction of leakage rates within the network may be achieved through the implementation of enhanced pipe renewal programmes and the introduction of more robust leakage detection strategies. A reduction in the leakage rate of approximately 50% and 75% resulted in comparable reductions in P_{inf} across all nodes and for all pathogens. Similarly, the reduction in leakage rates resulted in proportional reductions in median P_{inf} for all nodes and all reference pathogens, in a manner analogous to the log reduction scenarios.

Two additional scenarios were evaluated to ascertain the impact of varying repair times, encompassing both the detection period and the duration of repair. In one scenario, the overall repair time was reduced by a period of three hours. In the alternative scenario, the overall repair time was increased by an equivalent duration. A reduction in the overall repair time may be indicative of a combination of measures that could be employed in the future. Such measures could include the implementation of an early warning system to detect a break and promptly alert the relevant personnel from the water supplier. Conversely, an extended repair time could be indicative of prospective scenarios in which the detection and repair of pipes is more challenging due to unforeseen circumstances, necessitating the implementation of alternative strategies.

The results from varying repair times did not corroborate the linear trends that had been anticipated based on previous scenarios. While a reduction in repair time resulted in a reduction in the median P_{inf} for the majority of nodes, certain nodes, notably CN6, exhibited a significant increase (> 2000%) in P_{inf} for *Campylobacter* and *Cryptosporidium*. Conversely, an increase in repair time resulted in an overall increase in the median P_{inf} for the network, accompanied by a slight reduction in P_{inf} for some reference pathogens in nodes CN1, CN2, and CN9. One potential explanation for this is the dynamic nature of water distribution networks, whereby the time component affects the hydraulics, and thus the transport of contaminants within the network.

It can be posited that the duration of the repair is equivalent to the duration of the low-pressure event, which another intrusion model found to be one of the most relevant factors influencing the probability of virus infection (Teunis et al., 2010b). However, epidemiological studies in the Nordic countries have yielded mixed results; one study found an association between the duration of the water shutdown and a higher risk of GI (Nygard et al., 2007), while another study found no association (Säve-Söderbergh et al., 2017).

The experts' input resulted in a clear prioritization of strategies (see Figure 15 & Table 10). The highest ranked risk reduction measure was pressure management (score = 0.71), followed by Pipe Repair and Rehabilitation (score = 0.65), Increase Metering (score = 0.62), and Active Leakage Control (score = 0.61). The lowest ranked measure was public awareness and engagement (score =

0.41). The most preferred criterion was risk reduction (weight = 0.2663) and the least preferred criterion was social concern (weight = 0.1576), while executability and reliability were equally preferred (weight = 0.2011). The survey findings suggest that investing in physical infrastructure and operational efficiency, particularly through pressure management and pipe repair, would be preferred for risk reduction.

Table 9. Summary of variation in percentage (%) of median (50th percentile) daily probability of infection for measures tested compared to baseline Scenario 2 (2 $\log_{10}$ unit reduction, 27% leakage/intrusion rate, and 5 hours event period) for varying $\log_{10}$ reduction, leakage, and repair time (**Paper III**). For varying $\log_{10}$ reduction and leakage, an average of the variation for the three reference pathogens is shown due to the similarity of the values. Positive values indicate an increase in risk while negative values indicate a reduction in risk ("Campy" – *Campylobacter*, "Crypto" – *Cryptosporidium*).

Critical Node (CN)	Median P_{inf} for Scenario 2*			% Variation of Pinf with measures (negative – reduction, positive – increase)									
	Campy	Norovirus	Crypto	3 $\log_{10}$ reduction	4 $\log_{10}$ reduction	Varying leakage/intrusion		2-hour repair time			8-hour repair time		
						14%	7%	Campy	Norovirus	Crypto	Campy	Norovirus	Crypto
1	7.1E-05	6.6E-07	1.4E-06	-92.33	-99.24	-49.94	-82.99	-24.42	-23.35	-23.28	-4.89	11.68	-4.23
2	6.3E-05	4.8E-07	2.8E-06	-92.43	-99.24	-49.85	-83.04	-41.02	-43.19	-29.2	94.69	82.99	-4.21
3	5.4E-05	7.2E-07	2.2E-06	-92.41	-99.24	-49.66	-82.96	-32.89	-32.66	-29.74	144.8	140.4	11.34
4	2.1E-05	1.9E-07	5.3E-07	-92.24	-99.22	-49.59	-82.5	42.16	14.65	-58.78	292.5	275.0	55.82
5	5.8E-05	6.2E-07	1.3E-06	-92.32	-99.23	-49.55	-82.87	-34.65	-26.57	-30.66	54.42	78.31	26.84
6	8.5E-05	7.6E-07	1.6E-06	-92.41	-99.24	-49.02	-82.93	2020	-17.83	2011	27.18	110.36	4.69
7	7.8E-05	7.1E-07	2.5E-06	-92.38	-99.24	-49.73	-82.98	-43.24	-46.62	-24.42	728.3	757.9	774.2
8	6.6E-05	5.9E-07	1.4E-06	-92.41	-99.24	-49.51	-82.93	-60.82	-61.12	-58.07	114.2	116.4	113.9
9	7.2E-05	4.1E-07	2.0E-06	-92.43	-99.24	-49.09	-82.92	-41.74	-41.88	-35.61	-21.12	-10.39	26.12
10	6.8E-05	6.2E-07	1.7E-06	-92.35	-99.24	-49.72	-82.94	-32.24	-37.38	-30.93	445.5	409.8	417.3
Average	6.3E-05	5.8E-07	1.7E-06	-92.37	-99.24	-49.57	-82.91	175.1	-31.60	169.0	187.6	197.2	142.2

*2 $\log_{10}$ unit reduction, 27% leakage/intrusion rate, and 5-hour repair time

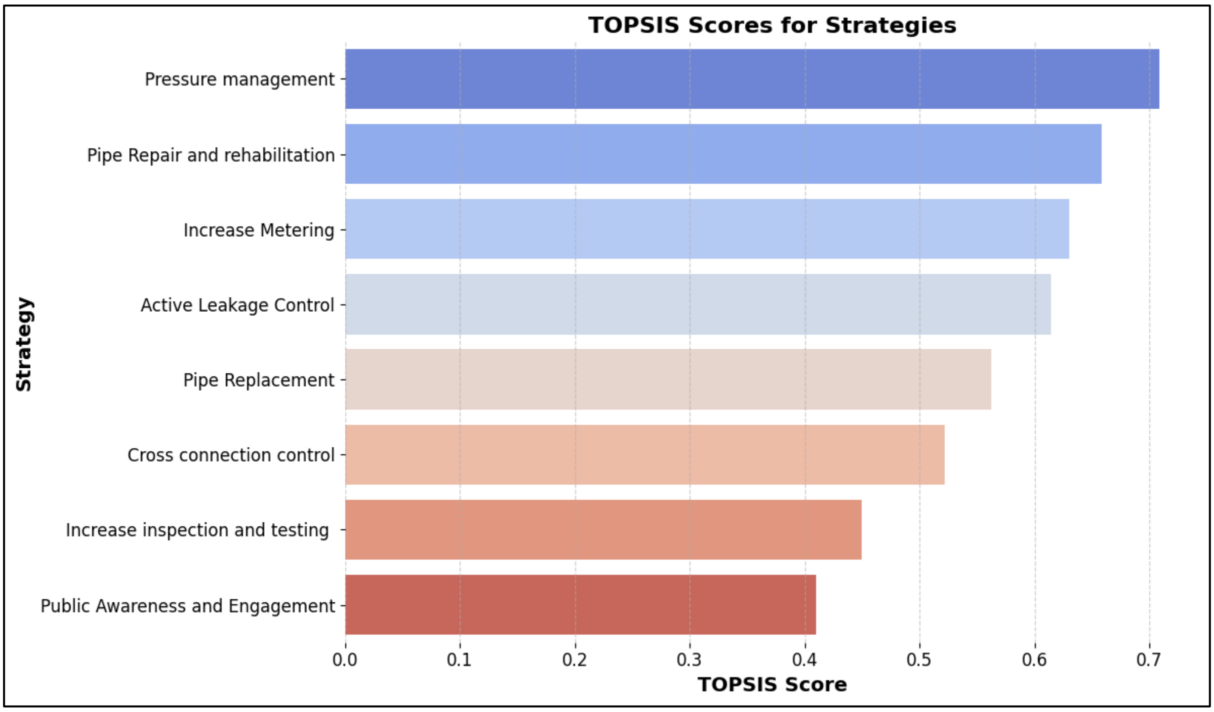

Figure 15. Ranking of strategies according to their TOPSIS score (n = 10 participants).

Table 10. Normalised weights based on expert judgement.

Criteria	Ranking	Weight
Cost	4	0.1739
Executability	2	0.2011
Risk Reduction	1	0.2663
Social Concern	5	0.1576
Reliability	2	0.2011

4.3 Validation of risk estimates

Based on the results for scenario C1 from **Paper II**, the number of symptomatic cases was quantified (Figure 9). The median number of symptomatic cases, as estimated using hydraulic modelling, varied between 97 and 148, depending on the assumptions about the duration of the contamination event (9 or 10 days) and the number of people per household (3 or 4). The most affected neighbourhood was Z2, with between 41 and 61 symptomatic cases, depending on underlying assumptions. The number of symptomatic cases for the outbreak case study was also estimated using the results from the national FTA. The number of symptomatic cases, as estimated using the infection risk calculated with FTA for 1 h duration, was 87. The epidemiological survey carried out during the outbreak recorded 179 cases of illness.

In the outbreak case study, the results of the epidemiological survey conducted in the area were found to be in close alignment with the outcomes for scenario C1, as simulated using the hydraulic model. This suggests that the assumptions made for scenario C1 were the most accurate in terms of representing the actual conditions that led to the outbreak, specifically in regard to contamination entering the network and the duration of the event. One potential limitation was the absence of a local consumption pattern, which would have been relevant in QMRA studies (Blokker et al., 2018; Säve-Söderbergh et al., 2018; Westrell et al., 2006). It is plausible that the peak contamination occurred during a time of day when most users were utilising the water for non-potable purposes. This could result in an overestimation of the risk to consumers, given that not all water used by the household would be for drinking purposes. Nevertheless, it seems unlikely that this was a contributing factor in our simulation, as the estimated cases were consistent with the findings of the outbreak survey.

A further potential source of uncertainty was secondary person-to-person transmission, which was not considered in the outbreak simulation. Outbreak investigations have shown that secondary person-to-person transmission may represent between 7% to 84% of the norovirus cases during an outbreak (Tsang et al., 2018). As illustrated in Figure 17, this could be a potential source for the underestimation of the model, given that a number of households were not included in the analysis due to falling outside of the identified impacted zones.

In order to validate the results obtained in **Paper III**, a comparison was made between the infection risks associated with norovirus and the estimated burden of acute gastrointestinal illness (AGI) in Sweden. The predicted incidence of AGI cases due to norovirus using the aggregated median daily infection risks for all scenarios was **0.07 cases per person-year** (min-max: 0.02 - 4.4) [calculations not shown]. The incidence of AGI in Sweden has been estimated at **0.36** and **0.43 cases per person per year** (Edelstein et al., 2016; Säve-Söderbergh et al., 2019). Furthermore, epidemiological studies in Scandinavia have estimated that the proportion of AGI cases attributable to water distribution network incidents is 37–38% (equivalent to **0.10–0.12 cases per person per year**) (Säve-Söderbergh et al., 2017). Other studies have demonstrated that the number of additional cases resulting from tap water consumption may range from 14 to 40% (Payment et al., 1997). Although the available estimates of AGI incidence do not differentiate between the various pathogens, these estimates provide a theoretical upper limit to the extent of the risk that may be reasonably predicted. It can be posited that the developed model can produce reasonable estimates of risk when benchmarked against the available epidemiological studies. Nevertheless, it is imperative to exercise caution, as the modelling may potentially underestimate or overestimate by a factor of one order of magnitude.

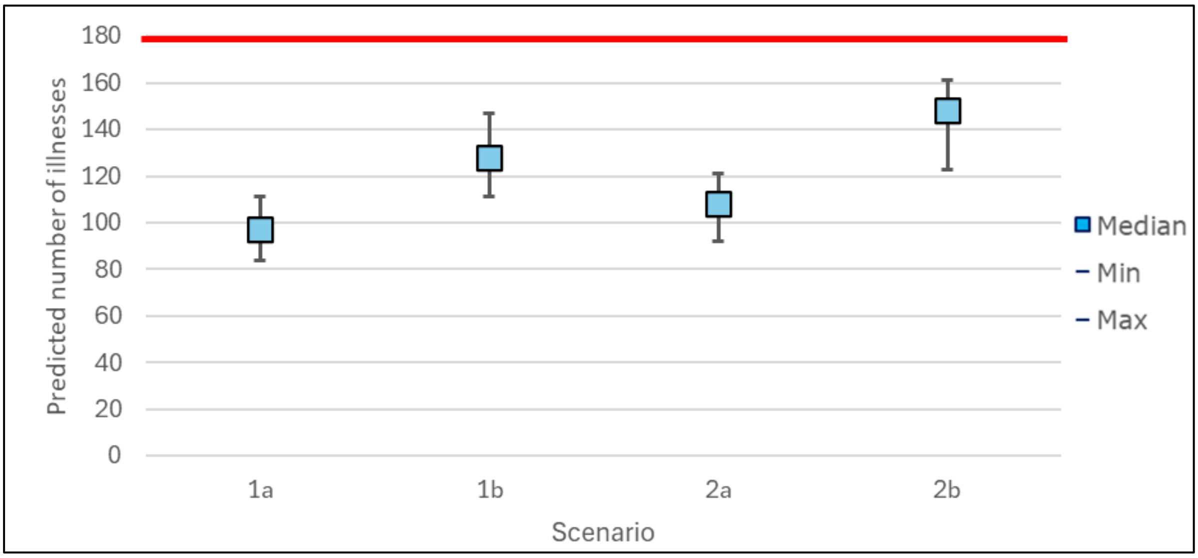

Figure 16. Predicted infected consumers (symptomatic infections) for scenario C1 under different assumptions for the outbreak case study. Total number of persons becoming ill (n = 179) documented in an outbreak survey is shown as a red line. Scenario 1a – 3 persons per household and 9 days cross-connection and backflow duration; Scenario 1b - 4 persons per household and 9 days duration; Scenario 2a - 3 persons per household and 10 days duration; Scenario 2b - 4 persons per and 10 days duration. Adapted from Table 4 in **Paper II**.

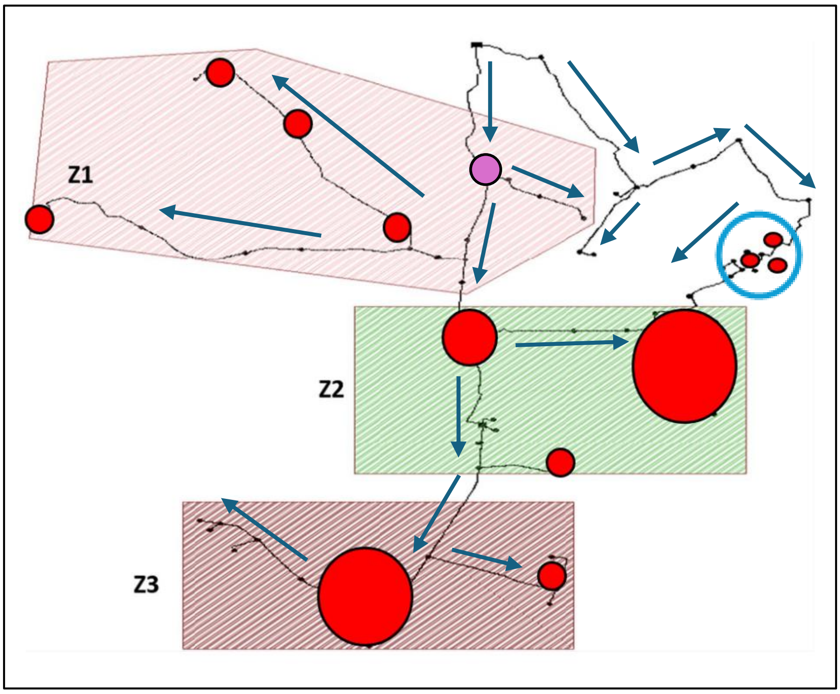

Figure 17. Geographical distribution of reported cases (red circles) according to the epidemiological survey. The direction of flow is indicated by blue arrows, with the pink circle marking the contamination site. A small number of households (enclosed in the blue circle) were situated outside of the identified impacted zones by the hydraulic model. Adapted from **Paper II**.

4.4 Contamination levels and microbial communities

Based on the sampling results obtained to use as input for **Papers IV & V**, it seems reasonable to conclude that the predominant microorganisms in the water surrounding the drinking water pipes have adapted to the local environmental conditions in the soil and are therefore unlikely to be indicative of faecal contamination. Notwithstanding the absence of *E. coli* in the metagenomic data, other faecal bacteria were observed. The dominant genera from the human gut were Bacteroides, Phocaeicola, Faecalibacterium, Prevotella, Alistipes, and Blautia A (Liu et al., 2021). Figure 18 illustrates the relative abundance of these genera in the samples. It is noteworthy that samples with low *E. coli* counts (less than 100 per L) exhibited a complete absence of gut-associated genera, whereas samples with high *E. coli* counts demonstrated the presence of at least five of these genera. Moreover, the samples with elevated *E. coli* levels also exhibit a high degree of diversity (Figure 19). The microbial communities present in sewage are markedly distinct from those that are adapted to the soil environment. It can thus be postulated that the intrusion of sewage into the soil may be a potential explanation for the increased diversity, the observations of gut-associated bacteria, and the elevated *E. coli* counts.

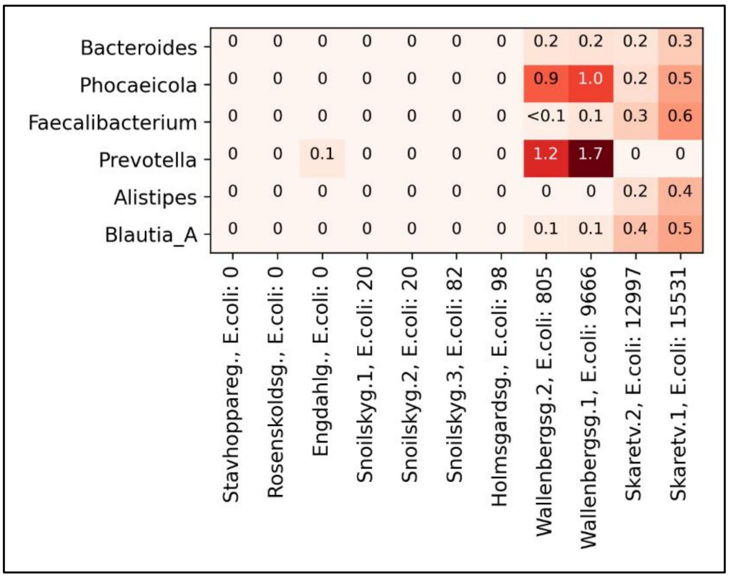

Figure 18. Relative abundance (%) of gut-associated genera. The samples are ordered from left to right in ascending order based on *E. coli* counts.

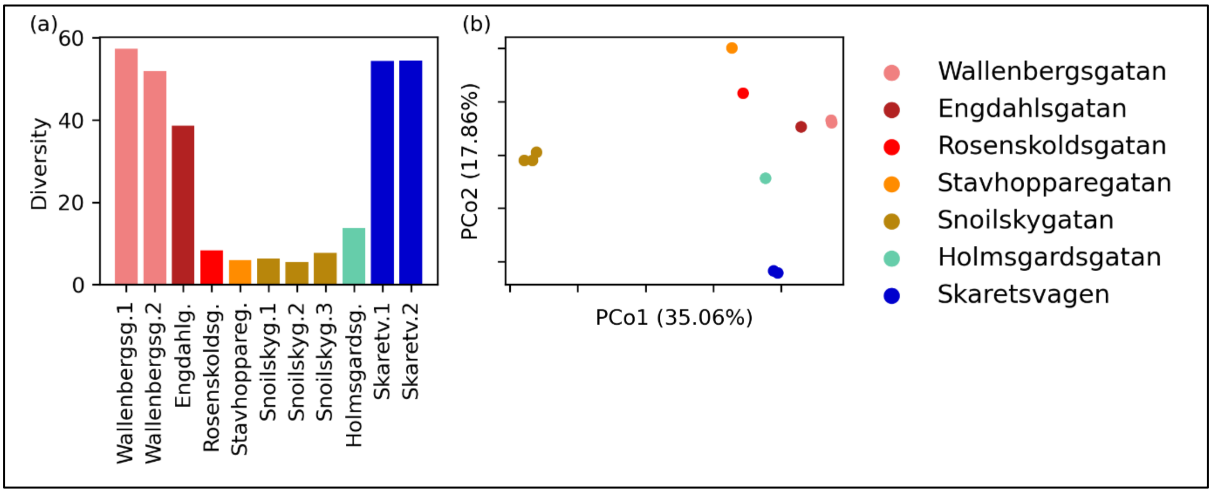

Figure 19. Diversity of the sample (a) and similarity in microbial community composition shown in a principal coordinate analysis (b).

5 Conclusions

This research aimed to establish a comprehensive microbial risk management framework for water distribution networks. The study addressed the following research questions:

RQ1: What are the current methodologies and tools for evaluating microbial risks in water distribution networks?

Several methodologies and tools were employed to evaluate microbial risks, including Fault Tree Analysis (FTA), Quantitative Microbial Risk Assessment (QMRA), hydraulic modelling, scenario generation techniques, Monte Carlo simulations, logistic regression for calculating the probability of pipe failure, and the combination of metagenomics with indicator analysis. These methods provided a robust foundation for assessing microbial risks.

RQ2: How can these methodologies be improved to provide a more accurate and comprehensive assessment of health risks in water distribution systems?

Improvements to the methodologies were identified, such as using local data instead of national aggregate data for FTA to enhance accuracy. The combination of metagenomics with indicator analysis showed significant potential for better detection of faecal contamination. Additionally, the use of stochastic models allowed for a more nuanced understanding of pipe failures and contamination levels, providing a more comprehensive assessment of health risks.

RQ3: What are the most influential factors affecting the risk of infection in water distribution networks?

The study identified key factors influencing the risk of infection, including the concentration of pathogens in intruding soil water, the duration of low-pressure events, and the temporal variation in consumption and hydraulics. These factors were found to have a significant impact on the risk of infection and should be considered in risk assessments.

RQ4: How can the developed models and methodologies be integrated into a comprehensive microbial risk management framework for water distribution networks?

The developed models and methodologies were integrated into a comprehensive framework by executing simulations within the Python environment, streamlining the process and eliminating the need for proficiency in multiple software programs. The combination of different methods (FTA, QMRA, hydraulic modelling, logistic regression) and the implementation of measures to reduce failure events and contamination levels were key components of the framework. Additionally, Multi-Criteria Decision Analysis (MCDA) was used to rank and highlight stakeholder preferences for risk reduction measures. Practical implications include the potential for water suppliers to use these integrated models to enhance microbial risk management and improve public health outcomes.

5.1 Possibilities for future work

The findings of this research open several avenues for future work, which can further enhance the understanding and management of microbial risks in water distribution networks. The following points outline potential areas for future research and practical applications.

- ***Legionella* and premise plumbing.**

The latest European Drinking Water Directive (Directive (EU) 2020/2184) [3] introduced more stringent measures to control *Legionella* in premise plumbing systems. *Legionella*, the causative agent of Legionnaires' disease, poses a significant health burden, with increasing incidence rates, particularly among vulnerable populations (Collier et al., 2021; Moffa et al., 2023). Future research endeavours could focus on incorporating the risk posed by opportunistic pathogens into the overarching microbial risk framework for WDNs. However, in order to achieve this, the following challenges must be addressed: limitations on actionable strategies from the water supplier side due to legislative constraints, as premise plumbing is frequently under the jurisdiction of other entities; and the absence of actionable information (e.g. plumbing material, physical condition, etc.) for the aforementioned reason. The integration of these legislative requirements into existing microbial risk management frameworks will be pivotal in ensuring a comprehensive water safety strategy.

- **Machine learning/deep learning techniques for predicting failure/contamination events**

In recent years, advanced machine learning (ML) and deep learning (DL) algorithms have been used to predict water quality and identify potential contamination events (Feng et al., 2025; Li et al., 2024). These techniques can analyze large amounts of data to find patterns and make accurate predictions. For example, they can help predict when and where contamination might occur in a water distribution network. However, these methods require high-quality data to be effective. Future research should focus on combining ML and DL with other technologies like data mining and cloud computing to improve prediction accuracy. Collaboration between water professionals and AI experts will be essential to develop and refine these predictive models.

- **Complementary methods to traditional microbiological water analyses as input data for QMRA.**

Traditional water quality assessment methods, such as membrane filtration or Colilert, have limitations in detecting a broad range of microbial contaminants and providing real-time data. Alternative techniques like flow cytometry and shotgun sequencing offer significant benefits. For example, flow cytometry has been shown to detect rapid changes in bacterial concentration, composition, and viability in real-time, making it a valuable tool for early-warning systems (Safford & Bischel, 2019). Flow cytometry could be used to detect intrusion and cross-connection contamination. Shotgun sequencing allows for the identification of entire microbial communities, offering insights into the presence of pathogens and microbial diversity (Taylor et al., 2024). These methods can enhance the accuracy and comprehensiveness of water quality assessments, supporting more effective microbial risk management.

- **Combination of water quality sampling, hydraulic modelling and syndromic surveillance.**

The potential exists for enhancement of the detection and validation of contamination events through combination of microbial water quality sampling with hydraulic modelling and syndromic surveillance. Despite an absence of conclusive evidence on the effectiveness of syndromic surveillance for early detection of waterborne outbreaks, the combination of multiple sources of signals with spatial information has the capacity to improve timeliness and reduce false alarms

[3] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020L2184

(Hyllestad et al., 2021). Consequently, future research should concentrate on integrating these methods to potentially validate model predictions and test various contamination scenarios. The integration of these approaches holds considerable promise in offering a more comprehensive understanding of contamination dynamics and enhancing the accuracy of risk assessments.

6 References

Abrahamsson, J. L., Ansker, J., Heinicke, G. . (2009). *MRA – a model for Swedish water works (in Swedish)*. Retrieved from Stockholm: https://www.svensktvatten.se/globalassets/dricksvatten/riskanalys-och-provtagning/svu-2009-05_mra---ett-modellverktyg-for-svenska-vattenverk.pdf

Angles, M. L., Chandy, J. P., Cox, P. T., Fisher, I. H., & Warnecke, M. R. (2007). Implications of biofilm-associated waterborne Cryptosporidium oocysts for the water industry. *Trends in Parasitology, 23*(8), 352-356. doi:10.1016/j.pt.2007.06.001

Atmar, R. L., Opekun, A. R., Gilger, M. A., Estes, M. K., Crawford, S. E., Neill, F. H., & Graham, D. Y. (2008). Norwalk virus shedding after experimental human infection. *Emerg Infect Dis, 14*(10), 1553-1557. doi:10.3201/eid1410.080117

Aven, T. (2010). On how to define, understand and describe risk. *Reliability Engineering & System Safety, 95*(6), 623-631. doi:https://doi.org/10.1016/j.ress.2010.01.011

Berry, D., Xi, C., & Raskin, L. (2006). Microbial ecology of drinking water distribution systems. *Current Opinion in Biotechnology, 17*(3), 297-302. doi:10.1016/j.copbio.2006.05.007

Besner, M. C., Broséus, R., Lavoie, J., Giovanni, G. D., Payment, P., & Prévost, M. (2010). Pressure monitoring and characterization of external sources of contamination at the site of the payment drinking water epidemiological studies. *Environmental Science & Technology, 44*(1), 269-277. doi:10.1021/es901988y

Besner, M. C., Prévost, M., & Regli, S. (2011). Assessing the public health risk of microbial intrusion events in distribution systems: Conceptual model, available data, and challenges. *Water Research, 45*(3), 961-979.

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., Berry, S. E., Spector, T. D., Franzosa, E. A., Pasolli, E., Asnicar, F., Huttenhower, C., & Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*. doi:10.1038/s41587-023-01688-w

Blokker, M., Smeets, P., & Medema, G. (2018). Quantitative microbial risk assessment of repairs of the drinking water distribution system. *Microbial Risk Analysis, 8*, 22-31. doi:10.1016/j.mran.2017.12.002

Boulos, P. F., Karney, B. W., Wood, D. J., & Lingireddy, S. (2005). Hydraulic Transient Guidelines for Protecting Water Distribution Systems. *Journal AWWA, 97*(5), 111-124. doi:https://doi.org/10.1002/j.1551-8833.2005.tb10892.x

Bouwknegt, M., Knol, A. B., van der Sluijs, J. P., & Evers, E. G. (2014). Uncertainty of population risk estimates for pathogens based on QMRA or epidemiology: a case study of Campylobacter in the Netherlands. *Risk Analysis, 34*(5), 847-864.

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying Species Diversity, Phylogenetic Diversity, Functional Diversity, and Related Similarity and Differentiation Measures Through Hill Numbers. *Annual Review of Ecology, Evolution, and Systematics, 45*(1), 297-324. doi:10.1146/annurev-ecolsys-120213-091540

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), i884-i890. doi:10.1093/bioinformatics/bty560

Clark, R. M., Abdesaken, F., Boulos, P. F., & Mau, R. E. (1996). Mixing in distribution system storage tanks: Its effect on water quality. *Journal of environmental engineering, 122*(9), 814-821.

Colford, J. M., Jr., Wade, T. J., Sandhu, S. K., Wright, C. C., Lee, S., Shaw, S., Fox, K., Burns, S., Benker, A., Brookhart, M. A., van der Laan, M., & Levy, D. A. (2005). A randomized, controlled trial of in-home drinking water intervention to reduce gastrointestinal illness. *Am J Epidemiol, 161*(5), 472-482. doi:10.1093/aje/kwi067

Collier, S., Deng, L., Adam, E., Benedict, K., Beshearse, E., Blackstock, A., Bruce, B., Derado, G., Edens, C., Fullerton, K., Gargano, J., Geissler, A., Hall, A., Havelaar, A., Hill, V., Hoekstra, R., Reddy, S., Scallan, E., Stokes, E., Yoder, J., & Beach, M. (2021). Estimate of Burden and Direct Healthcare Cost of Infectious Waterborne Disease in the United States. *Emerging Infectious Disease journal, 27*(1), 140. doi:10.3201/eid2701.190676

Craun, G. F. (2012). The importance of waterborne disease outbreak surveillance in the United States. *Annali dell'Istituto superiore di sanità, 48*, 447-459.

Cummings, J. H., Bingham, S. A., Heaton, K. W., & Eastwood, M. A. (1992). Fecal weight, colon cancer risk, and dietary intake of nonstarch polysaccharides (dietary fiber). *Gastroenterology, 103*(6), 1783-1789. doi:10.1016/0016-5085(92)91435-7

De Feo, G., Angelakis, A. N., Antoniou, G. P., El-Gohary, F., Haut, B., Passchier, C. W., & Zheng, X. Y. (2013). Historical and Technical Notes on Aqueducts from Prehistoric to Medieval Times. *Water (Switzerland), 5*(4), 1996-2025. Retrieved from https://www.mdpi.com/2073-4441/5/4/1996

Edelstein, M., Merk, H., Deogan, C., Carnahan, A., & Wallensten, A. (2016). Quantifying the incidence and cost of acute gastrointestinal illness in Sweden, 2013–2014. *Epidemiology and Infection, 144*(13), 2831-2839. doi:10.1017/S0950268816000467

Eftim, S. E., Hong, T., Soller, J., Boehm, A., Warren, I., Ichida, A., & Nappier, S. P. (2017). Occurrence of norovirus in raw sewage - A systematic literature review and meta-analysis. *Water Res, 111*, 366-374. doi:10.1016/j.watres.2017.01.017

Ercumen, A., Gruber, J. S., & Colford, J. M., Jr. (2014). Water distribution system deficiencies and gastrointestinal illness: a systematic review and meta-analysis. *Environmental Health Perspectives, 122*(7), 651-660. doi:10.1289/ehp.1306912

Falco, R., & Williams, S. I. (2009). *Waterborne Salmonella Outbreak in Alamosa, Colorado March and April 2008: Outbreak Identification, Response, and Investigation*. Retrieved from http://deq.mt.gov/wqinfo/pws/security/incident/AlamosaInvestRpt.pdf

Fan, X., Wang, X., Zhang, X., & Yu, X. (2022). Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. *Reliability Engineering & System Safety, 219*, 108185. doi:10.1016/j.ress.2021.108185

Farkas, A., Dragan-Bularda, M., Ciataras, D., Bocos, B., & Tigan, S. (2012). Opportunistic pathogens and faecal indicators in drinking water associated biofilms in Cluj, Romania. *J Water Health, 10*(3), 471-483. doi:10.2166/wh.2012.148

Feng, F., Zhang, Y., Chen, Z., Ni, J., Feng, Y., Xie, Y., & Zhang, C. (2025). Machine learning to assess and support safe drinking water supply: a systematic review. *Journal of Environmental Chemical Engineering, 13*(1), 114481. doi:https://doi.org/10.1016/j.jece.2024.114481

Forero-Ortiz, E., Martinez-Gomariz, E., Sanchez-Juny, M., Cardus Gonzalez, J., Cucchietti, F., Baque Viader, F., & Sarrias Monton, M. (2023). Models and explanatory variables in modelling failure for drinking water pipes to support asset management: a mixed literature review. *Applied Water Science, 13*(11), 210. doi:10.1007/s13201-023-02013-1

Gasso, G. (2019). Logistic regression. *INSA Rouen-ASI Departement Laboratory, 1-30*.

Guzman-Herrador, B., Carlander, A., Ethelberg, S., De Blasio, B. F., Kuusi, M., Lund, V., Löfdahl, M., MacDonald, E., Nichols, G., Schönning, C., Sudre, B., Trönnberg, L., Vold, L., Semenza, J. C., & Nygård, K. (2015). Waterborne outbreaks in the Nordic countries, 1998 to 2012. *Eurosurveillance, 20*(24), 1-10.

Haas, C. N., Rose, J. B., & Gerba, C. P. (2014). *Quantitative microbial risk assessment*: John Wiley & Sons.

Hellard, M. E., Sinclair, M. I., Forbes, A. B., & Fairley, C. K. (2001). A randomized, blinded, controlled trial investigating the gastrointestinal health effects of drinking water quality. *Environmental Health Perspectives, 109*(8), 773-778.

Henze, M., Harremoes, P., Jansen, J. C., & Arvin, E. . (2001). Wastewater Components. In *Wastewater Treatment: Biological and Chemical Processes* (pp. 27-37): Springer Berlin Heidelberg.

Hewitt, J., Leonard, M., Greening, G. E., & Lewis, G. D. (2011). Influence of wastewater treatment process and the population size on human virus profiles in wastewater. *Water Research, 45*(18), 6267-6276. doi:https://doi.org/10.1016/j.watres.2011.09.029

Hrudey, S. E., & Hrudey, E. J. (2004). *Safe drinking water : lessons from recent outbreaks in affluent nations*. London, UK: IWA Pub.

Hrudey, S. E., & Hrudey, E. J. (2007). Published case studies of waterborne disease outbreaks--evidence of a recurrent threat. *Water Environ Res, 79*(3), 233-245. doi:10.2175/106143006x95483

Hyllestad, S., Amato, E., Nygård, K., Vold, L., & Aavitsland, P. (2021). The effectiveness of syndromic surveillance for the early detection of waterborne outbreaks: a systematic review. *BMC Infectious Diseases, 21*(1), 696. doi:10.1186/s12879-021-06387-y

Hynds, P. D., Misstear, B. D., & Gill, L. W. (2012). Investigation of water consumption patterns among Irish adults for waterborne quantitative microbial risk assessment (QMRA). *Journal of Water and Health, 10*(3), 453-464. doi:10.2166/wh.2012.025

Institute of Medicine. (2000). The National Academies Collection: Reports funded by National Institutes of Health. In J. R. Davis & J. Lederberg (Eds.), *Public Health Systems and Emerging Infections: Assessing the Capabilities of the Public and Private Sectors: Workshop Summary*. Washington (DC): National Academies Press (US).

Islam, N., Farahat, A., Al-Zahrani, M. A. M., Rodriguez, M. J., & Sadiq, R. (2015). Contaminant intrusion in water distribution networks: review and proposal of an integrated model for decision making. *Environmental Reviews, 23*(3), 337-352. doi:10.1139/er-2014-0069

Jamal, R., Mubarak, S., Sahulka, S. Q., Kori, J. A., Tajammul, A., Ahmed, J., Mahar, R. B., Olsen, M. S., Goel, R., & Weidhaas, J. (2020). Informing water distribution line rehabilitation through quantitative microbial risk assessment. *Science of the Total Environment, 739*. doi:10.1016/j.scitotenv.2020.140021

James, G. (2013). An introduction to statistical learning. In: springer.

Jonsson, P., Lindgren, D., Asplund, M., Stavklint, H., Magounakis, M., Mokhlesi, S., Pettersson, T., Eriksson, M., Winquist, F., Jonasson, C., Ilver, D., & Strömbeck, N. (2018). *Electronic tongue and other online sensors for detection of contaminants in the drinking water network - an evaluation (in Swedish)*. Retrieved from Stockholm: https://vattenbokhandeln.svensktvatten.se/produkt/elektronisk-tunga-och-andra-onlinesensorer-for-detektion-av-fororeningar-i-dricksvattennatet/

Kanakoudis, V., & Tsitsifli, S. (2019). Water Networks Management: New Perspectives. *Water (Switzerland), 11*(2), 239. Retrieved from https://www.mdpi.com/2073-4441/11/2/239

Karim, M. R., Abbaszadegan, M., & Lechevallier, M. (2003). Potential for pathogen intrusion during pressure transients. *Journal American Water Works Association, 95*(5), 134-146. Retrieved from <Go to ISI>://WOS:000182978700013

Kauppinen, A., Ikonen, J., Pursiainen, A., Pitkanen, T., & Miettinen, I. T. (2012). Decontamination of a drinking water pipeline system contaminated with adenovirus and Escherichia coli utilizing peracetic acid and chlorine. *Journal of Water and Health, 10*(3), 406-418. doi:10.2166/wh.2012.003

Khanal, N., Buchberger, S. G., & McKenna, S. A. (2006). Distribution system contamination events: Exposure, influence, and sensitivity. *Journal of Water Resources Planning and Management, 132*(4), 283-292.

Kimutai, E., Betrie, G., Brander, R., Sadiq, R., & Tesfamariam, S. (2015). Comparison of Statistical Models for Predicting Pipe Failures: Illustrative Example with the City of Calgary Water Main Failure. *Journal of Pipeline Systems Engineering and Practice, 6*(4), 04015005. doi:10.1061/(ASCE)PS.1949-1204.0000196

Kirmeyer, G. J., Thomure, T. M., Rahman, R., Marie, J. L., LeChevallier, M. W., Yang, J., Hughes, D. M., & Schneider, O. (2014). *Effective microbial control strategies for main breaks and depressurization*. Retrieved from Denver:

Kleiner, Y., & Rajani, B. (2001). Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water, 3*(3), 131-150. doi:10.1016/S1462-0758(01)00033-4

Klise, K., Hart, D., Bynum, M., Hogge, J., Haxton, T., Murray, R., & Burkhardt, J. (2020). *Water Network Tool for Resilience (WNTR). User Manual, Version 0.2. 3*. Retrieved from

Kristianstads kommun. (2015). *Contaminated water in Everöd, Östra Sönnarslöv and Huaröd: Summer 2015 (in Swedish)*. Retrieved from

Laine, J., Huovinen, E., Virtanen, M. J., Snellman, M., Lumio, J., Ruutu, P., Kujansuu, E., Vuento, R., Pitkanen, T., Miettinen, I., Herrala, J., Lepisto, O., Antonen, J., Helenius, J., Hanninen, M. L., Maunula, L., Mustonen, J., Kuusi, M., & Pirkanmaa Waterborne Outbreak Study, G. (2011).

An extensive gastroenteritis outbreak after drinking-water contamination by sewage effluent, Finland. *Epidemiol Infect, 139*(7), 1105-1113. doi:10.1017/S0950268810002141

Larock, B. E., Jeppson, R. W., & Watters, G. Z. (1999). *Hydraulics of Pipeline Systems*: CRC Press.

Li, W., Zhao, Y., Zhu, Y., Dong, Z., Wang, F., & Huang, F. (2024). Research progress in water quality prediction based on deep learning technology: a review. *Environmental Science and Pollution Research, 31*(18), 26415-26431. doi:10.1007/s11356-024-33058-7

Lindhe, A. (2010). *Risk assessment and decision support for managing drinking water systems*: Chalmers University of Technology.

Liu, C., Li, Z., Ding, J., Zhen, H., Fang, M., & Nie, C. (2021). Species-Level Analysis of the Human Gut Microbiome Shows Antibiotic Resistance Genes Associated With Colorectal Cancer. *Frontiers in Microbiology, 12*. doi:10.3389/fmicb.2021.765291

Magana-Arachchi, D. N., & Wanigatunge, R. P. (2020). Chapter 2 - Ubiquitous waterborne pathogens. In M. N. Vara Prasad & A. Grobelak (Eds.), *Waterborne Pathogens* (pp. 15-42): Butterworth-Heinemann.

Makar, J. M. (2000). A preliminary analysis of failures in grey cast iron water pipes. *Engineering Failure Analysis, 7*(1), 43-53. doi:10.1016/S1350-6307(99)00005-9

Maliva, R., & Missimer, T. (2012). *Arid Lands Water Evaluation and Management*. Berlin: Springer Berlin.

Malm, A., Axelsson, G., Barregard, L., Ljungqvist, J., Forsberg, B., Bergstedt, O., & Pettersson, T. J. R. (2013). The association of drinking water treatment and distribution network disturbances with Health Call Centre contacts for gastrointestinal illness symptoms. *Water Research, 47*(13), 4474-4484. doi:10.1016/j.watres.2013.05.003

Malm, A., Pettersson, T. J. R., & Bergstedt, O. (2010). *Health effects of quality disturbances in Swedish water distribution networks (in Swedish)*. Paper presented at the Proceedings of the 7th Nordic Drinking Water Conference, Copenhagen.

Mansour-Rezaei, S., & Naser, G. (2013). Contaminant intrusion in water distribution systems: An ingress model. *American Water Works Association, 105*(1), E29-E39. Retrieved from http://www.jstor.org/stable/jamewatworass.105.1.e29

Mays, L., Antoniou, G. P., & Angelakis, A. N. (2013). History of Water Cisterns: Legacies and Lessons. *Water (Switzerland), 5*(4), 1916-1940. Retrieved from https://www.mdpi.com/2073-4441/5/4/1916

Mays, L. W. (2008). A very brief history of hydraulic technology during antiquity. *Environmental Fluid Mechanics, 8*(5), 471-484. doi:10.1007/s10652-008-9095-2

Mays, L. W., Koutsoyiannis, D., & Angelakis, A. N. (2007). A brief history of urban water supply in antiquity. *Water Supply, 7*(1), 1-12. doi:10.2166/ws.2007.001

McInnis, D. (2004). A relative-risk framework for evaluating transient pathogen intrusion in distribution systems. *Urban Water Journal, 1*(2), 113-127. doi:10.1080/15730620412331290010

Medema, G. J., Teunis, P. F., Havelaar, A. H., & Haas, C. N. (1996). Assessment of the dose-response relationship of Campylobacter jejuni. *Int J Food Microbiol, 30*(1-2), 101-111. doi:10.1016/0168-1605(96)00994-4

Messner, M., Shaw, S., Regli, S., Rotert, K., Blank, V., & Soller, J. (2006). An approach for developing a national estimate of waterborne disease due to drinking water and a national estimate model application. *J Water Health, 4 Suppl 2*, 201-240. doi:10.2166/wh.2006.024

Messner, M. J., Berger, P., & Nappier, S. P. (2014). Fractional poisson--a simple dose-response model for human norovirus. *Risk Anal, 34*(10), 1820-1829. doi:10.1111/risa.12207

Metcalf & Eddy, Inc. (2003). *Wastewater Engineering: Treatment and Reuse*: McGraw-Hill Education.

Miura, F., Matsuyama, R., & Nishiura, H. (2018). Estimating the Asymptomatic Ratio of Norovirus Infection During Foodborne Outbreaks With Laboratory Testing in Japan. *Journal of epidemiology, 28*(9), 382-387. doi:10.2188/jea.JE20170040

Modin, O., Liébana, R., Saheb-Alam, S., Wilén, B.-M., Suarez, C., Hermansson, M., & Persson, F. (2020). Hill-based dissimilarity indices and null models for analysis of microbial community assembly. *Microbiome, 8*(1), 132. doi:10.1186/s40168-020-00909-7

Moffa, M. A., Rock, C., Galiatsatos, P., Gamage, S. D., Schwab, K. J., & Exum, N. G. (2023). Legionellosis on the rise: A scoping review of sporadic, community-acquired incidence in the United States. *Epidemiol Infect, 151*, e133. doi:10.1017/s0950268823001206

Mora-Rodríguez, J., Delgado-Galván, X., Ramos, H. M., & López-Jiménez, P. A. (2014). An overview of leaks and intrusion for different pipe materials and failures. *Urban Water Journal, 11*(1), 1-10. doi:10.1080/1573062X.2012.739630

Murphy, H., Thomas, M., Medeiros, D., McFadyen, S., & Pintar, K. (2016). Estimating the number of cases of acute gastrointestinal illness (AGI) associated with Canadian municipal drinking water systems. *Epidemiology & Infection, 144*(7), 1371-1385.

National Research Council. (2006). *Drinking water distribution systems : assessing and reducing risks*. Washington, DC: National Academies Press.

Ngubane, Z., Bergion, V., Dzwairo, B., Stenström, T. A., & Sokolova, E. (2024). Multi-criteria decision analysis framework for engaging stakeholders in river pollution risk management. *Scientific Reports, 14*(1). doi:10.1038/s41598-024-57739-y

Nilsson, K. A., Buchberger, S. G., & Clark, R. M. (2005). Simulating Exposures to Deliberate Intrusions into Water Distribution Systems. *Journal of Water Resources Planning and Management, 131*(3), 228-236. doi:doi:10.1061/(ASCE)0733-9496(2005)131:3(228)

Nocker, A., Burr, M., & Camper, A. (2014). Chapter One - Pathogens in Water and Biofilms. In S. L. Percival, M. V. Yates, D. W. Williams, R. M. Chalmers, & N. F. Gray (Eds.), *Microbiology of Waterborne Diseases* (2nd Edition ed., pp. 3-32). London: Academic Press.

Nygard, K., Wahl, E., Krogh, T., Tveit, O. A., Bohleng, E., Tverdal, A., & Aavitsland, P. (2007). Breaks and maintenance work in the water distribution systems and gastrointestinal illness: a cohort study. *International Journal of Epidemiology, 36*(4), 873-880. doi:10.1093/ije/dym029

Nygård, K., Wahl, E., Krogh, T., Tveit, O. A., Bøhleng, E., Tverdal, A., & Aavitsland, P. (2007). Breaks and maintenance work in the water distribution systems and gastrointestinal illness: a cohort study. *Int J Epidemiol, 36*(4), 873-880. doi:10.1093/ije/dym029

Ottoson, J., Hansen, A., Westrell, T., Johansen, K., Norder, H., & Stenström, T. A. (2006). Removal of noro- and enteroviruses, Giardia cysts, Cryptosporidium oocysts, and fecal indicators at four secondary wastewater treatment plants in Sweden. *Water Environ Res, 78*(8), 828-834. doi:10.2175/106143006x101719

Park, H., Kim, K., Hyung, J., Kim, T., & Koo, J. (2023). Decision-making for the hazard ranking of water distribution networks using the TOPSIS method. *Water Supply, 23*(2), 715-726. doi:10.2166/ws.2023.013

Payment, P., Richardson, L., Siemiatycki, J., Dewar, R., Edwardes, M., & Franco, E. (1991). A randomized trial to evaluate the risk of gastrointestinal disease due to consumption of drinking water meeting current microbiological standards. *American journal of public health, 81*(6), 703-708.

Payment, P., Siemiatycki, J., Richardson, L., Renaud, G., Franco, E., & Prévost, M. (1997). A prospective epidemiological study of gastrointestinal health effects due to the consumption of drinking water. *International Journal of Environmental Health Research, 7*(1), 5-31. doi:10.1080/09603129773977

Petterson, S. R., & Ashbolt, N. J. (2016). QMRA and water safety management: review of application in drinking water systems. *J Water Health, 14*(4), 571-589. doi:10.2166/wh.2016.262

Pryor, M., Springthorpe, S., Riffard, S., Brooks, T., Huo, Y., Davis, G., & Sattar, S. A. (2004). Investigation of opportunistic pathogens in municipal drinking water under different supply and treatment regimes. *Water Science and Technology, 50*(1), 83-90.

Ramírez-Castillo, F. Y., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F. J., Harel, J., & Guerrero-Barrera, A. L. (2015). Waterborne pathogens: detection methods and challenges. *Pathogens, 4*(2), 307-334. doi:10.3390/pathogens4020307

Risebro, H. L., Doria, M. F., Andersson, Y., Medema, G., Osborn, K., Schlosser, O., & Hunter, P. R. (2007). Fault tree analysis of the causes of waterborne outbreaks. *J Water Health, 5 Suppl 1*, 1-18. doi:10.2166/wh.2007.136

Roche, S., Jones, A., Majowicz, S., McEwen, S., & Pintar, K. (2012). Drinking water consumption patterns in Canadian communities (2001–2007). *Journal of Water and Health, 10*(1), 69-86.

Safford, H. R., & Bischel, H. N. (2019). Flow cytometry applications in water treatment, distribution, and reuse: A review. *Water Research, 151*, 110-133. doi:https://doi.org/10.1016/j.watres.2018.12.016

Seitz, S. R., Leon, J. S., Schwab, K. J., Lyon, G. M., Dowd, M., McDaniels, M., Abdulhafid, G., Fernandez, M. L., Lindesmith, L. C., Baric, R. S., & Moe, C. L. (2011). Norovirus infectivity in humans and persistence in water. *Appl Environ Microbiol, 77*(19), 6884-6888. doi:10.1128/aem.05806-11

September, S. M., Els, F. A., Venter, S. N., & Brözel, V. S. (2007). Prevalence of bacterial pathogens in biofilms of drinking water distribution systems. *Journal of Water and Health, 5*(2), 219-227. doi:10.2166/wh.2007.004

Seyoum, A. G., & Tanyimboh, T. T. (2014). *Pressure-dependent network water quality modelling.* Paper presented at the Proceedings of the Institution of Civil Engineers-Water Management.

Signor, R. S., & Ashbolt, N. J. (2009). Comparing probabilistic microbial risk assessments for drinking water against daily rather than annualised infection probability targets. *J Water Health, 7*(4), 535-543. doi:10.2166/wh.2009.101

Silman, A. J., Macfarlane, G. J., & Macfarlane, T. (2018). *Epidemiological Studies: A Practical Guide*: Oxford University Press.

Skraber, S., Schijven, J., Gantzer, C., & De Roda Husman, A. M. (2005). Pathogenic viruses in drinking-water biofilms: a public health risk? *Biofilms, 2*(2), 105-117.

Soller, J. A., Schoen, M. E., Bartrand, T., Ravenscroft, J. E., & Ashbolt, N. J. (2010). Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Research, 44*(16), 4674-4691. doi:https://doi.org/10.1016/j.watres.2010.06.049

Storey, M. V., & Ashbolt, N. J. (2003). Enteric virions and microbial biofilms - a secondary source of public health concern? *Water Science and Technology, 48*(3), 97-104. Retrieved from <Go to ISI>://WOS:000185579600014

Svenskt Vatten. (2020). *P114 Drinking water supply (in Swedish)*. Retrieved from Stockholm:

Swamee, P. K., & Sharma, A. K. (2008). *Design of water supply pipe networks*: John Wiley & Sons.

Szumilas, M. (2010). Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry, 19*(3), 227-229.

Säve-Söderbergh, M., Bylund, J., Malm, A., Simonsson, M., & Toljander, J. (2017). Gastrointestinal illness linked to incidents in drinking water distribution networks in Sweden. *Water Research, 122*, 503-511. doi:https://doi.org/10.1016/j.watres.2017.06.013

Säve-Söderbergh, M., Toljander, J., Bylund, J., & Simonsson, M. (2019). Burden of gastrointestinal illness in Sweden-SMS as a tool for collecting self-reported gastrointestinal illness. *Epidemiology and Infection, 147*, e322. doi:10.1017/S0950268819002103

Säve-Söderbergh, M., Toljander, J., Mattisson, I., Åkesson, A., & Simonsson, M. (2018). Drinking water consumption patterns among adults—SMS as a novel tool for collection of repeated self-reported water consumption. *Journal of Exposure Science & Environmental Epidemiology, 28*(2), 131-139. doi:10.1038/jes.2017.8

Taherdoost, H., & Madanchian, M. (2023). Multi-Criteria Decision Making (MCDM) Methods and Concepts. *Encyclopedia, 3*(1), 77-87. doi:10.3390/encyclopedia3010006

Taylor, W., Devane, M. L., Russell, K., Lin, S., Roxburgh, C., Williamson, J., & Gilpin, B. J. (2024). Metagenomic evaluation of bacteria in drinking water using full-length 16S rRNA amplicons. *Journal of Water and Health, 22*(8), 1429-1443. doi:10.2166/wh.2024.090

Teunis, P., Schijven, J., & Rutjes, S. (2016). A generalized dose-response relationship for adenovirus infection and illness by exposure pathway. *Epidemiol Infect, 144*(16), 3461-3473. doi:10.1017/s0950268816001862

Teunis, P., Takumi, K., & Shinagawa, K. (2004). Dose response for infection by Escherichia coli O157:H7 from outbreak data. *Risk Anal, 24*(2), 401-407. doi:10.1111/j.0272-4332.2004.00441.x

Teunis, P., Van den Brandhof, W., Nauta, M., Wagenaar, J., Van den Kerkhof, H., & Van Pelt, W. (2005). A reconsideration of the Campylobacter dose-response relation. *Epidemiol Infect, 133*(4), 583-592. doi:10.1017/s0950268805003912

Teunis, P. F., Kasuga, F., Fazil, A., Ogden, I. D., Rotariu, O., & Strachan, N. J. (2010a). Dose-response modeling of Salmonella using outbreak data. *Int J Food Microbiol, 144*(2), 243-249. doi:10.1016/j.ijfoodmicro.2010.09.026

Teunis, P. F., Nagelkerke, N. J., & Haas, C. N. (1999). Dose response models for infectious gastroenteritis. *Risk Anal, 19*(6), 1251-1260. doi:10.1023/a:1007055316559

Teunis, P. F., Ogden, I. D., & Strachan, N. J. (2008a). Hierarchical dose response of E. coli O157:H7 from human outbreaks incorporating heterogeneity in exposure. *Epidemiol Infect, 136*(6), 761-770. doi:10.1017/s0950268807008771

Teunis, P. F. M., Chappell, C. L., & Okhuysen, P. C. (2002). Cryptosporidium Dose-Response Studies: Variation Between Hosts. *Risk Analysis, 22*(3), 475-485. doi:https://doi.org/10.1111/0272-4332.00046

Teunis, P. F. M., Moe, C. L., Liu, P., E. Miller, S., Lindesmith, L., Baric, R. S., Le Pendu, J., & Calderon, R. L. (2008b). Norwalk virus: How infectious is it? *Journal of Medical Virology, 80*(8), 1468-1476. doi:https://doi.org/10.1002/jmv.21237

Teunis, P. F. M., Xu, M., Fleming, K. K., Yang, J., Moe, C. L., & LeChevallier, M. W. (2010b). Enteric Virus Infection Risk from Intrusion of Sewage into a Drinking Water Distribution Network. *Environmental Science & Technology, 44*(22), 8561-8566. doi:10.1021/es101266k

The Swedish Water & Wastewater Association. (2000). *Facts on water supply and sanitation in Sweden*. Retrieved from Stockholm, Sweden: www.svensktvatten.se

Todini, E., & Pilati, S. (1988). A gradient algorithm for the analysis of pipe networks. In *Computer applications in water supply: vol. 1---systems analysis and simulation* (pp. 1–20): Research Studies Press Ltd.

Torres, J. M., Brumbelow, K., & Guikema, S. D. (2009). Risk classification and uncertainty propagation for virtual water distribution systems. *Reliability Engineering & System Safety, 94*(8), 1259-1273.

Tsakiris, G., & Spiliotis, M. (2014). A Newton–Raphson analysis of urban water systems based on nodal head-driven outflow. *European Journal of Environmental and Civil Engineering, 18*(8), 882-896. doi:10.1080/19648189.2014.909746

Tsang, T. K., Chen, T. M., Longini, I. M., Jr., Halloran, M. E., Wu, Y., & Yang, Y. (2018). Transmissibility of Norovirus in Urban Versus Rural Households in a Large Community Outbreak in China. *Epidemiology, 29*(5), 675-683. doi:10.1097/ede.0000000000000855

USEPA. (2002). *Potential Contamination Due to Cross-Connections and Backflow and the Associated Health Risks* Retrieved from Washington DC, USA: http://www.epa.gov/safewater/disinfection/tcr/pdfs/issuepaper_tcr_crossconnection-backflow.pdf

Vairavamoorthy, K., Yan, J. M., Galgale, H. M., & Gorantiwar, S. D. (2007). IRA-WDS: A GIS-based risk analysis tool for water distribution systems. *Environmental Modelling & Software, 22*(7), 951-965. doi:10.1016/j.envsoft.2006.05.027

Warad, A. A. M., Wassif, K., & Darwish, N. R. (2024). An ensemble learning model for forecasting water-pipe leakage. *Scientific Reports, 14*(1), 10683. doi:10.1038/s41598-024-60840-x

Werner de Vargas, V., Schneider Aranda, J. A., dos Santos Costa, R., da Silva Pereira, P. R., & Victória Barbosa, J. L. (2023). Imbalanced data preprocessing techniques for machine learning: a systematic mapping study. *Knowledge and Information Systems, 65*(1), 31-57. doi:10.1007/s10115-022-01772-8

Westrell, T., Andersson, Y., & Stenström, T. A. (2006). Drinking water consumption patterns in Sweden. *J Water Health, 4*(4), 511-522.

WHO. (2014). *Water safety in distribution systems* (W. H. Organization Ed.). Geneva: World Health Organization.

WHO. (2016). *Quantitative microbial risk assessment: application for water safety management*. Retrieved from Geneva:

WHO. (2022). *Guidelines for drinking-water quality: fourth edition incorporating the first and second addenda* (4th ed + 2st add ed.). Geneva: World Health Organization.

WHO/UNICEF JMP. (2023). *Progress on household drinking water, sanitation and hygiene 2000-2022: Special focus on gender*. Retrieved from New York:

Wilson, D., Filion, Y., & Moore, I. (2017). State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water Journal, 14*(2), 173-184. doi:10.1080/1573062X.2015.1080848

Wingender, J., & Flemming, H. C. (2011). Biofilms in drinking water and their role as reservoir for pathogens. *International Journal of Hygiene and Environmental Health, 214*(6), 417-423. doi:10.1016/j.ijheh.2011.05.009

Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., & Tscheikner-Gratl, F. (2018). Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering, 14*(10), 1402-1411. doi:10.1080/15732479.2018.1443145

Wood, D. J., & Charles, C. O. A. (1972). Hydraulic Network Analysis Using Linear Theory. *Journal of the Hydraulics Division, 98*(7), 1157-1170. doi:doi:10.1061/JYCEAJ.0003348

Yamijala, S., Guikema, S. D., & Brumbelow, K. (2009). Statistical models for the analysis of water distribution system pipe break data. *Reliability Engineering & System Safety, 94*(2), 282-293. doi:10.1016/j.ress.2008.03.011

Yang, J., LeChevallier, M. W., Teunis, P. F. M., & Xu, M. (2011). Managing risks from virus intrusion into water distribution systems due to pressure transients. *Journal of Water and Health, 9*(2), 291-305. doi:10.2166/wh.2011.102

Yang, J., Schneider, O. D., Jjemba, P. K., & Lechevallier, M. W. (2015). Microbial Risk Modeling for Main Breaks. *Journal American Water Works Association, 107*(2), E97-E108. doi:10.5942/jawwa.2015.107.0010

Zmirou-Navier, D., Gofti-Laroche, L., & Hartemann, P. (2006). Waterborne microbial risk assessment: a population-based dose-response function for Giardia spp. (E.MI.R.A study). *BMC Public Health, 6*, 122. doi:10.1186/1471-2458-6-122

Zolghadr-Asli, B., Bozorg-Haddad, O., & van Cauwenbergh, N. (2021). Multi-attribute Decision-Making: A View of the World of Decision-Making. In O. Bozorg-Haddad (Ed.), *Essential Tools for Water Resources Analysis, Planning, and Management* (pp. 305-322). Singapore: Springer.

Åström, J., Lindhe, A., Bergvall, M., Rosén, L., & Lång, L.-O. (2016). *Microbial risk assessment of groundwater sources – development and implementation of a QMRA tool (in Swedish)*. Retrieved from Stockholm, Sweden: