

NARROWING THE UNCERTAINTY ASSOCIATED WITH PATHOGEN PERSISTENCE IN  
SURFACE WATERS FOR APPLICATIONS IN QUANTITATIVE MICROBIAL RISK  
ASSESSMENT

By

Kara Jane Dean

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Biosystems Engineering – Doctor of Philosophy  
Environmental Science and Policy – Doctor of Philosophy

2022

## ABSTRACT

Surface waters are used for recreation, irrigation water, and as source water for drinking water treatment plants. These uses can be associated with human health risks when fecal contamination from point and non-point sources introduces pathogens into surface waters capable of causing waterborne disease. The quality of surface waters is typically monitored with indicator organisms, and it has commonly been assumed in literature, quantitative microbial risk assessments (QMRA), and surface water management decision-making that indicator and pathogen persistence are similar, and that indicators and pathogens decay at a constant rate in the environment.

To address these assumptions, this dissertation presents i) a systematic literature review that collated, compared, and analyzed the available persistence data for indicators and pathogens in surface waters; ii) a meta-analysis that fit exponential decay and alternative models to the database of over 600 experiments, identified a model that best fit the data most frequently, and statistically evaluated the relationships between frequently documented water quality factors and observed persistence dynamics; iii) a general model developed with Bayesian hierarchical modeling that quantifies the uncertainty between indicator and pathogen persistence; and iv) a QMRA case study that evaluates the impact of persistence knowledge for decision-making pertaining to a recreational waterbody impacted by a sewage spill event.

The systematic literature review (Chapter 2) found that the 61 selected studies predominantly evaluated FIB, freshwater matrices, and culture-based methods of detection. Comparing the methods and results across the studies qualitatively suggested potential interactions between sunlight, water type, and method of detection, and between predation, water type, and temperature. Within the subsequent meta-analysis (Chapter 3), the Juneja and Marks 2 (JM2) model, based on the logistic probability distribution, provided the best fit to the data most

frequently. First-order decay kinetics provided the best fit to less than 20% of the analyzed data. Random forest methods identified temperature, water type, and predation as the most important factors influencing persistence, and the protozoa target type differed the most from FIB.

A general model was developed using the comprehensive database of persistence experiments, the JM2 model, temperature, predation, and water type data, and Bayesian hierarchical modeling techniques (Chapter 4). A varying-intercept model with target-specific intercepts and population-level coefficients for temperature, predation, and water type was the optimal evaluated model form. The general model indicated that protozoa persistence more commonly has initial periods of minimal decay and virus decay typically tapers off the most quickly over time. Median uncertainty factors quantified with the general model ranged from 1 to 3.4 for bacteria, bacteriophage, virus, and protozoa persistence behaviors compared to FIB. The application of the uncertainty factors was demonstrated within a QMRA case study in which the JM2 model was fit to culturable enterococci (cENT), enterohemorrhagic *Escherichia coli* (EHEC), and adenovirus (HAV) data to characterize the persistence of the targets after the containment of a sewage spill (Chapter 5). Applying temperature-specific uncertainty factors to the cENT data ensured the risk of illness associated with EHEC and HAV ingestion fell below the Recreational Water Quality Criteria limit of 36 in 1,000 swimmers.

The work presented herein indicates that broadly applying first-order decay kinetics to persistence data may lead to erroneous decision making in the fields of water management and protection, and that a general model for persistence can add value to the indicator-pathogen paradigm.

## ACKNOWLEDGEMENTS

I would first like to acknowledge my advisor, Jade Mitchell, for her guidance, support, and mentorship for the past eight years. Joining the Mitchell Lab in 2014 inspired me to pursue an unexpected career path and provided me with a support system that I relied upon throughout my educational pursuits. I am incredibly grateful for the opportunities I have had as Dr. Mitchell's student to contribute to meaningful projects and broaden my skillsets; it has been an honor to be her student. This work would not have been possible without her, and I truly could not have asked for a better advisor and mentor.

I would also like to thank the members of my committee: Drs. Pouyan Nejadhashemi, Joan Rose, and Erin Dreelin. Their feedback and guidance were a pivotal part of my development as an independent and successful researcher, and their direction provided significant value to this work. I appreciate each of them immensely and am thankful for their counsel throughout my graduate career.

My time at Michigan State University has also been enriched with several friendships that provided me with much needed laughter, advice, and support. I am grateful for my fellow graduate students, lab mates, and friends.

Dr. Raul Gonzalez and Kyle Curtis from Hampton Roads Sanitation District provided data and important insights that facilitated the completion of the quantitative microbial risk assessment case study in Chapter 5. It was a pleasure working with both of them.

I am also highly appreciative of the fellowship program that supported this work. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-1849739). Any opinions, findings, and conclusions or recommendations expressed in this material are my own and do not necessarily reflect the views of the National Science Foundation.

Finally, I would like to acknowledge my mother, father, sister, and fiancé. My family has always believed in my ability to achieve my goals, and I cannot thank them enough for their support during my graduate studies. They motivated and reassured me throughout this process and were gracious enough to attend the numerous practice conference and dissertation presentations I gave in the living room over the past few years. I am eternally grateful for their love and support.

## PREFACE

Chapter 2 of this document, “Identifying water quality and environmental factors that influence indicator and pathogen decay in natural surface waters”, has been published in *Water Research* and is reprinted with permission from *Water Research* 2022, 211, 1, 118051.

Copyright 2022 Elsevier Ltd.

Chapter 3 of this document, “Meta-analysis addressing the implications of model uncertainty in understanding the persistence of indicators and pathogens in natural surface waters”, has been published in *Environmental Science & Technology* and is reprinted with permission from *Environ. Sci. Technol.* 2022, 56, 17, 12106–12115. Copyright 2022 American Chemical Society.

## TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION .....	1
1.1 Problem Statement .....	1
1.2 Relevant Background .....	3
1.3 Research Objectives .....	8
REFERENCES .....	10
CHAPTER 2: IDENTIFYING WATER QUALITY AND ENVIRONMENTAL FACTORS THAT INFLUENCE INDICATOR AND PATHOGEN DECAY IN NATURAL SURFACE WATERS .....	15
2.1 Introduction .....	15
2.2 Selection Criteria .....	17
2.3 Results .....	18
2.4 Discussion .....	40
2.5 Conclusions .....	45
REFERENCES .....	50
APPENDIX .....	57
CHAPTER 3: META-ANALYSIS ADDRESSING THE IMPLICATIONS OF MODEL UNCERTAINTY IN UNDERSTANDING THE PERSISTENCE OF INDICATORS AND PATHOGENS IN NATURAL SURFACE WATERS .....	59
3.1 Introduction .....	59
3.2 Methods .....	61
3.3 Results .....	68
3.4 Discussion .....	78
REFERENCES .....	86
APPENDIX A: GOODNESS OF FIT METRIC SENSITIVITY ASSESSMENT .....	92
APPENDIX B: FACTOR ANALYSIS METHODS .....	94
APPENDIX C: MODEL FITTING DETAILS .....	95
APPENDIX D: ADDITIONAL FACTOR ANALYSIS DETAILS AND RESULTS .....	96
CHAPTER 4: TESTING A GENERAL MODEL FOR PATHOGEN PERSISTENCE IN SURFACE WATERS .....	106
4.1 Introduction .....	106
4.2 Methods .....	108
4.3 Results .....	114
4.4 Discussion .....	126
REFERENCES .....	130
APPENDIX A: PRIOR DISTRIBUTION SELECTION AND SENSITIVITY ANALYSIS .....	132
APPENDIX B: EVALUATION OF PREDICTIVE POWER .....	136
CHAPTER 5: APPLYING PERSISTENCE KNOWLEDGE WITHIN A QMRA CASE STUDY OF A SEWAGE SPILL EVENT .....	137
5.1 Introduction .....	137
5.2 Methods .....	138

5.3 Results .....	147
5.4 Discussion .....	155
REFERENCES.....	160
APPENDIX.....	162
CHAPTER 6: CONCLUSIONS .....	163
REFERENCES.....	166
CHAPTER 7: FUTURE WORK .....	167



## CHAPTER 1: INTRODUCTION

### 1.1 Problem Statement

Pathogens are disease-causing microorganisms that can pose human health risks when present in the environment. Pathogen contamination is the most commonly reported cause of water pollution in the United States (EPA, 2012), and is one of the leading causes of stream and river impairments (EPA, 2017). It has been estimated that 7.15 million waterborne illnesses occur in the U.S. annually, and that these illnesses can incur approximately \$3.33 billion in direct healthcare costs (Collier et al., 2021). Exposure to waterborne pathogens can occur through ingestion, inhalation, and dermal contact pathways, and otitis externa, norovirus infection, giardiasis, cryptosporidiosis, and campylobacteriosis are the top five waterborne diseases acquired domestically (Collier et al., 2021).

When considering surface water-related exposures, ingestion is typically the primary pathway of concern as surface waters can be used for recreation, as source water for drinking water treatment plants, and for irrigation in agriculture. The ingestion of pathogen-contaminated water is most frequently associated with gastrointestinal illness, and it is estimated that globally diarrheal diseases are one of the main causes of child mortality (Prüss-Ustün et al., 2016; WHO, 2019). Fecal-oral pathogens associated with inadequate drinking water (unprotected wells, springs, or surface waters) are responsible for approximately 35% of diarrhea-related deaths (Prüss-Ustün et al., 2016; 2019). Thus, it is highly important that decision makers be well-informed about the sources of pathogen contamination, possible risks associated with exposure, and available options for mitigation.

Risks associated with pathogens in surface and drinking waters can be characterized and predicted with quantitative microbial risk assessments (QMRA) (WHO, 2017). QMRA provides

a framework for identifying and characterizing risks associated with pathogens in environmental matrices and is a valuable tool for informing water treatment and management (Haas et al., 2014). QMRA consists of five key components: i) Hazard Identification, ii) Dose Response Assessment, iii) Exposure Assessment, iv) Risk Characterization, and v) Risk Management. The exposure and dose response assessment components of a QMRA involve modeling: the modeling of an exposure pathway to determine possible exposure doses and the modeling of the relationship between exposure doses and probabilities of adverse outcomes. The exact inputs into these models are often not known, and the uncertainty and variability associated with each parameter are commonly represented with probability distributions. Including uncertainty in the parameterization of exposure and dose response assessments facilitates the output of risk distributions which can be used to explore risk management opportunities and inform decision makers (Haas et al., 2014).

Narrowing the uncertainty associated with persistence in exposure assessments may yield more accurate risk characterizations for pathogens of concern in surface waters. The sources of uncertainty associated with pathogen persistence in surface waters include the use of indicator organisms and a reliance on the application of first-order decay kinetic models. In terms of model form uncertainty, there has been substantial evidence of biphasic decay patterns in natural and bench-scale experiments (Medema et al. 1997; Easton et al., 1999; Easton et al. 2005; Park et al., 2016; Mitchell & Akram, 2017). The log-linear model traditionally applied to decay data assumes a constant rate of decay and does not account for tailing or shoulder behaviors that have been observed in the literature (de Brauwere et al., 2014; Pachepsky et al., 2006; Crane & Moore, 1986). The application of a more accurate model form capable of capturing the observed behaviors may reduce the uncertainty associated with pathogen populations present at the point

of exposure. As pathogen contamination is a common cause of impairments capable of causing a significant number of illnesses and associated healthcare costs, improving the characterization of surface-water related risks is an important next step towards further protecting human health.

## 1.2 Relevant Background

### 1.2.1 Legislative Context

Pathogens can be introduced to surface waters via nonpoint and point source pollution. Surface waters are protected from pollution in the U.S. under the Clean Water Act (CWA) of 1972 by considering pollution “articles of commerce”. In its original form, the CWA focused on point source pollution by providing federal assistance for the construction of municipal sewage treatment plants and putting regulatory requirements on industrial and municipal discharges (33 U.S.C § 1251 *et seq.*). Some of the key programs addressing point source pollution include the requirement for states to set water quality standards, the use of the National Pollutant Discharge Elimination System (NPDES) and the calculation of Total Maximum Daily Loads (TMDLs). In brief, the CWA requires that states set water quality standards for their navigable surface waters. There are three components to a standard: a designated use, water quality criteria, and anti-degradation. In terms of human exposure to pathogens in surface waters, some of the designated uses of concern include primary contact (immersion via swimming), source of drinking water, and agriculture uses. Based on the designated use for the surface water, water quality criteria are selected, and water quality monitoring is conducted to ensure the water body is meeting the state’s standards. If the standards are not met, the water is listed as impaired and a TMDL is developed. A TMDL is the total amount of pollutant that a waterbody can receive, and it helps inform the permits given through the NPDES to dischargers in the watershed. The permits distinguish between lawful and unlawful discharges through the requirement of certain types of

technology or the specification of effluent limitations. Monitoring is done to confirm compliance and evaluate the status of the water body as impaired.

The 1987 amendment to the CWA established Section 319 which focused on protections against nonpoint source pollution. Section 319 requires states to develop and implement Nonpoint Source Pollution Management programs and provides federal funding for nonpoint source pollution projects (33 U.S.C §§ 1251 *et seq.*). A state's management program has to identify best management practices to reduce the introduction of contamination, present a plan for implementing the identified measures or practices, and budget the use of federal and state funds. Section 319 also funds the National Monitoring Program (NMP) which provides additional funds for intense monitoring and evaluation of a subset of state watershed projects (Lombardo et al., 2000). If pathogen contamination has been identified as a source of impairment for a watershed, monitoring of indicator levels could be a part of the management plan.

The Safe Drinking Water Act (SDWA) of 1974 aims to protect the public from manmade and natural contaminants in drinking water. The SDWA has requirements for the treatment of surface waters that serve as sources for drinking water within the series of Surface Water Treatment Rules (42 U.S.C. § 300f *et seq.*). For systems that do not provide filtration for *Cryptosporidium* spp. or *Giardia lamblia*, there are various additional requirements to verify source water quality. The Long Term 2 Enhanced SWTR also requires that raw source water be monitored for *Cryptosporidium* or *E. coli* based on system size and treatment levels. The SDWA, however, primarily gives authority for the protection and improvement of drinking water. The protection or improvement of surface waters predominantly relies on local volunteer efforts or the regulatory programs under the CWA described previously (42 U.S.C. § 300f *et seq.*; EPA, 2014).

### 1.2.2 Surface Water Monitoring Practices

The water quality monitoring mandated by the CWA and SDWA, and any voluntary monitoring of beaches and waters conducted by local authorities, rely on the use of indicator organisms. Indicator organisms are nonpathogenic species that provide evidence for the presence or absence of other pathogenic organisms capable of surviving the given conditions. The presence, absence, or population density of indicators is assumed to be related to the risk of illness for water users (Benham et al., 2006). *An ideal organism should be easily detectable, specific to a source of pollution, present in concentrations that correlate well with pathogens, and have similar fate and transport characteristics to pathogens (NRC, 2004).* It is often not feasible to monitor for pathogens specifically in surface waters, due to time, technology, and cost limitations.

Total coliforms, fecal coliforms, *Escherichia coli* and enterococci are fecal indicator bacteria (FIB) that have been frequently used to monitor surface waters. The EPA’s Recreational Water Quality Criteria recommend the use of enterococci for marine or freshwaters and *E. coli* for freshwaters, as both have been shown to perform well as indicators for illness in sewage-contaminated waters (EPA, 2012). The use of indicators in state water quality standards is demonstrated in Table 1.1 with a brief summary of Michigan’s water quality standards that pertain to microorganisms.

Table 1.1: Michigan Water Quality Standards (EGLE, 2006)

Water Monitored	Specification	Value	Sample Requirements
Recreational Waters	Full Body Contact	130 <i>E. coli</i> /100mL	30-day geometric mean (five or more sampling events, three samples or more per event)
		300 <i>E. coli</i> /100 mL	1-day geometric mean (single sampling event, three or more samples)
	Partial Body Contact	1000 <i>E. coli</i> /100 mL	1-day geometric mean (single sampling event, three or more samples)

Table 1.1 (cont'd)

Wastewater Discharge	Discharge Containing Treated or Untreated Human Sewage	200 fecal coliform/100 mL	30-day geometric mean (5 or more samples)
		400 fecal coliform/100 mL	7-day geometric mean (3 or more samples taken during single discharge event)

The reliance on indicators has been noted as a potential source of uncertainty and error in the monitoring of surface waters (Harwood et al., 2005; Korajkic et al., 2018). Differences in shedding rates and fate and transport behaviors in secondary habitats have been noted as obstacles in the indicator-pathogen paradigm (Korajkic et al., 2018). For example, protozoa have been shown to survive longer in the environment than FIB, making FIB a possibly unreliable indicator for important human health-related targets like *Cryptosporidium* and *Giardia* (Craun et al., 1997). FIB and viruses have also been observed to respond differently to water treatment and environmental degradation processes, and as such the EPA is currently developing coliphage-based water quality criteria to improve protections against viral pathogens (EPA, 2015). Qualitative and quantitative differences between indicator and pathogen persistence behaviors are important considerations for the advancement of surface water monitoring protocols.

### 1.2.3 Persistence Modeling

The reliance on first-order decay kinetics is another source of uncertainty pertaining to the assessment, prediction, and management of pathogen contamination in surface waters. The fate of indicators and pathogens has typically been assumed to occur at a constant, linear rate (de Brauwere et al., 2014; Pachepsky et al., 2006; Crane & Moore, 1986). The exponential, or log-linear, model has been commonly applied to microorganism persistence data because it has a single parameter that is simple to estimate (Crane & Moore, 1986). Bench scale and in-situ experiments available in the literature, however, challenge the assumption of first-order decay

(Easton et al., 1999; Benham et al., 2006; Pachpesky et al., 2006; Blaustien et al., 2013; Park et al., 2016). Two stages of first-order decay (biphasic decay), initial periods of minimal decay, and decay rates tapering off with time are common dynamics observed for targets of interest in water-related matrices through experimentation and meta-analyses within the Global Water Pathogens Project (Easton et al., 1999; Hellweger et al., 2009; Brouwer et al., 2017; Mitchell & Akram, 2017). Possible drivers of these non-linear behaviors may be related to population heterogeneity, responses to population density (Brouwer et al., 2017; Easton et al., 2005), or the complex conditions and stressors found in environmental matrices.

The application of first-order decay has been used to model the persistence of targets in surface waters over time, predict metrics of interest such as the time for a 1-log reduction (T<sub>90</sub>), and generate dependent variables for analyses of the conditions influencing decay (Ahmed et al., 2019; Avery et al., 2008; Espinosa et al., 2008; Korajkic et al., 2013, 2014, 2019; Levin-Edens et al., 2011; Liang et al., 2017; Tiwari et al., 2019; Wanjugi et al., 2016; Boehm et al., 2018). As previous analyses of water-related matrices (Mitchell & Akram, 2017; Dean et al., 2020) have indicated that non-linear model forms may be better able to describe target persistence, it is possible that the application of first-order decay kinetics in prior studies has i) introduced uncertainty into the characterization of indicator and pathogen persistence for quantitative microbial risk assessments and surface water decision making applications, and ii) limited our interpretation of the experimental factors influencing persistence.

A suite of 17 models with two and three parameters, predominantly from food science and medicine, have been fit to a variety of persistence data encompassing enteric markers, viruses, and other pathogens on fomites, in sewage, and in other water matrices (Mitchell & Akram, 2017; Dean et al., 2020; Enger et al., 2018; Brooks et al., 2015; Tamrakar et al., 2016).

The exponential-damped, Juneja and Marks 1, Juneja and Marks 2, and double exponential models were found to best fit the data most frequently in previous studies and were selected as possible candidates for a model form better able to describe the persistence of indicators and pathogens in surface waters than the exponential model (Dean et al., 2020; Mitchell & Akram, 2017; Tamrakar et al., 2016; Enger et al., 2018). Reducing the model uncertainty associated with pathogen persistence is an important goal as it will improve the prediction of persistence values of interest for surface waters and progress our understanding of the water quality factors affecting persistence. Furthermore, the identification of an optimal model form for persistence may facilitate the development of a general model for characterizing and predicting the persistence of indicators and pathogens in varied surface water conditions in lieu of site or target-specific monitoring data. Quantifying the uncertainty between indicator and pathogen persistence within this general model form will add knowledge to the indicator-pathogen paradigm, improving indicator-reliant decision-making and policymaking.

### 1.3 Research Objectives

The goal of this work was to reduce the uncertainty associated with persistence modeling in exposure assessments within the QMRA framework to better inform decision makers involved in the management and treatment of surface waters. To achieve this goal, four objectives were pursued:

1. Develop a comprehensive database of experiments that reflects the current i) quantity of data available for pathogen and indicator organism persistence modeling, and ii) state of knowledge with regards to the environmental and water quality factors that influence persistence behaviors.



2. Systematically evaluate the available data in order to identify a model (or models) best able to describe the persistence of pathogens and indicators in surface waters.
3. Elucidate the most important factors influencing persistence behaviors through rigorous analysis of the data and models to produce a generalizable model.
4. Demonstrate the importance of reducing persistence modeling uncertainty for the characterization of risks associated with surface water exposures in order to highlight the relevance of this work for water management.

These four objectives were addressed in the form of four separate manuscripts entitled “Identifying water quality and environmental factors that influence indicator and pathogen decay in natural surface waters” (Dean & Mitchell, 2022a), “A meta-analysis addressing the implications of model uncertainty in understanding the persistence of indicators and pathogens in natural surface waters” (Dean & Mitchell, 2022b), “Testing a general model for pathogen persistence in surface waters”, and “Applying persistence knowledge within a QMRA case study of a sewage spill event”, that are Chapters 2, 3, 4, and 5 herein.

## REFERENCES

- Abraham, G., Debray, E., Candau, Y., & Piar, G. (1990). Mathematical Model of Thermal Destruction of *Bacillus stearothermophilus* Spores. *Applied and Environmental Microbiology*, 56(10), 3073–3080. <https://doi.org/10.1128/aem.56.10.3073-3080.1990>
- Atwood, K. C., & Norman, A. (1949). On the Interpretation of Multi-Hit Survival Curves. *Proceedings of the National Academy of Sciences*, 35(12), 696–709. <https://doi.org/10.1073/pnas.35.12.696>
- Benham, B.L., Baggaut, C., Zeckoski, R.W., Mankin, K.R., Pachepsky, Y.A., Sadeghi, A.M., Brannan, K.M., Soupir, M.L., & Habersack, M.J. 2006. Modeling bacteria fate and transport in watersheds to support TMDLs. *Transactions of the ASABE*, 49(4): 987-1002. doi: 10.13031/2013.21739
- Blaustein, R. A., Pachepsky, Y., Hill, R. L., Shelton, D. R., & Whelan, G. (2013). Escherichia coli survival in waters: Temperature dependence. *Water Research*, 47(2), 569–578. <https://doi.org/10.1016/j.watres.2012.10.027>
- Boehm, A. B., Graham, K. E., & Jennings, W. C. (2018). Can We Swim Yet? Systematic Review, Meta-Analysis, and Risk Assessment of Aging Sewage in Surface Waters. *Environmental Science & Technology*, 52(17), 9634–9645. <https://doi.org/10.1021/acs.est.8b01948>
- Brooks, Y., Aslan, A., Tamrakar, S., Murali, B., Mitchell, J., & Rose, J. B. (2015). Analysis of the persistence of enteric markers in sewage polluted water on a solid matrix and in liquid suspension. *Water Research*, 76, 201–212. <https://doi.org/10.1016/j.watres.2015.02.039>
- Brouwer, A. F., Eisenberg, M. C., Remais, J. V., Collender, P. A., Meza, R., & Eisenberg, J. N. S. (2017). Modeling Biphasic Environmental Decay of Pathogens and Implications for Risk Analysis. *Environmental Science & Technology*, 51(4), 2186–2196. <https://doi.org/10.1021/acs.est.6b04030>
- Carlier, V., Augustin, J.C., & Rozier, J. (1996). Heat resistance of *Listeria monocytogenes* (Phagovar 2389/2425/3274/2671/47/108/340): D- and z-values in ham. *Journal of Food Protection*, 59(6): 588-591.
- Chick, H. (1908). An Investigation of the Laws of Disinfection. *Journal of Hygiene*, 8(1), 92–158. <https://doi.org/10.1017/S0022172400006987>
- Clean Water Act: 33 U.S.C § § 1251 *et seq.*
- Collier, S. A., Deng, L., Adam, E. A., Benedict, K. M., Beshearse, E. M., Blackstock, A. J., Bruce, B. B., Derado, G., Edens, C., Fullerton, K. E., Gargano, J. W., Geissler, A. L., Hall, A. J., Havelaar, A. H., Hill, V. R., Hoekstra, R. M., Reddy, S. C., Scallan, E., Stokes, E. K., ... Beach, M. J. (2021). Estimate of Burden and Direct Healthcare Cost of Infectious Waterborne Disease in the United States. *Emerging Infectious Diseases*, 27(1), 140–149. <https://doi.org/10.3201/eid2701.190676>

- Crane, S. R., & Moore, J. A. (1986). Modeling enteric bacterial die-off: A review. *Water, Air, & Soil Pollution*, 27(3–4), 411–439. <https://doi.org/10.1007/BF00649422>
- Craun, G. F., Berger, P. S., & Calderon, R. L. (1997). Coliform bacteria and waterborne disease outbreaks. *Journal - American Water Works Association*, 89(3), 96–104. <https://doi.org/10.1002/j.1551-8833.1997.tb08197.x>
- Dean, K., Wissler, A., Hernandez-Suarez, J.S., Nejadhashemi, A.P., & Mitchell, J. (2020). Modeling the Persistence of Viruses in Untreated Groundwater. *Science of the Total Environment*, 717(15). <https://doi.org/10.1016/j.scitotenv.2019.134599>
- Dean, K. & Mitchell, J. (2022a). Identifying water quality and environmental factors that influence indicator and pathogen decay in natural surface waters. *Water Research*, in press. Available at: <https://doi.org/10.1016/j.watres.2022.118051>
- Dean, K. & Mitchell, J.(2022b). A meta-analysis addressing the implications of model uncertainty in understanding the persistence of indicators and pathogens in natural surface waters. *Environmental Science & Technology*, 56(17): 12106-12115.
- de Brauwere, A., Ouattara, N. K., & Servais, P. (2014). Modeling Fecal Indicator Bacteria Concentrations in Natural Surface Waters: A Review. *Critical Reviews in Environmental Science and Technology*, 44(21), 2380–2453. <https://doi.org/10.1080/10643389.2013.829978>
- Easton, J.H., Lalor, M., Gauthier, J. J., Pitt, R., Newman, D. E., & Meyland, S. (1999). *Determination of Survival Rates for Selected Bacterial and Protozoan Pathogens from Wet Weather Discharges*. Water Environment Federation 72<sup>nd</sup> Annual Conference & Exposition. New Orleans, October 1999.
- Easton, J. H., Gauthier, J. J., Lalor, M. M., & Pitt, R. E. (2005). DIE-OFF OF PATHOGENIC E. COLI O157:H7 IN SEWAGE CONTAMINATED WATERS. *Journal of the American Water Resources Association*, 41(5), 1187–1193. <https://doi.org/10.1111/j.1752-1688.2005.tb03793.x>
- EGLE. (2006). Part 4: Water Quality Standards. Water Bureau, Water Resources Protection. Retrieved from: <https://www.michigan.gov/egle/about/organization/Water-Resources/assessment-michigan-waters/water-quality-standards>
- Enger, K. S., Mitchell, J., Murali, B., Birdsell, D. N., Keim, P., Gurian, P. L., & Wagner, D. M. (2018). Evaluating the long-term persistence of *Bacillus* spores on common surfaces. *Microbial Biotechnology*, 11(6), 1048–1059. <https://doi.org/10.1111/1751-7915.13267>
- EPA. (2012). Recreational Water Quality Criteria. Office of Water 820-F-12-058. Retrieved from: <https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf>
- EPA. (2014). Opportunities to protect drinking water sources and advance watershed goals through the Clean Water Act: A toolkit for state, interstate, tribal, and federal water program managers. Retrieved from:

<https://www.gwpc.org/topics/source-water-protection/cwa-sdwa-coordination-toolkit/>

- EPA. (2015). Review of coliphages as possible indicators of fecal contamination for ambient water quality. Retrieved from:  
[https://www.epa.gov/sites/default/files/201607/documents/review\\_of\\_coliphages\\_as\\_possible\\_indicators\\_of\\_fecal\\_contamination\\_for\\_ambient\\_water\\_quality.pdf](https://www.epa.gov/sites/default/files/201607/documents/review_of_coliphages_as_possible_indicators_of_fecal_contamination_for_ambient_water_quality.pdf)
- EPA. (2017). National Water Quality Inventory: Report to Congress. Retrieved from:  
[https://www.epa.gov/sites/default/files/2017-12/documents/305brtc\\_finalowow\\_08302017.pdf](https://www.epa.gov/sites/default/files/2017-12/documents/305brtc_finalowow_08302017.pdf)
- Haas, C.N., Rose, J.B., Gerba, C.P. (2014). *Quantitative Microbial Risk Assessment*. 2<sup>nd</sup> ed. New York: Wiley.
- Harwood, V. J., Levine, A. D., Scott, T. M., Chivukula, V., Lukasik, J., Farrah, S. R., & Rose, J. B. (2005). Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. *Applied and Environmental Microbiology*, 71(6), 3163–3170. <https://doi.org/10.1128/AEM.71.6.3163-3170.2005>
- Hellweger, F. L., & Masopust, P. (2008). Investigating the Fate and Transport of *Escherichia coli* in the Charles River, Boston, Using High-Resolution Observation and Modeling <sup>1</sup>. *JAWRA Journal of the American Water Resources Association*, 44(2), 509–522. <https://doi.org/10.1111/j.1752-1688.2008.00179.x>
- Juneja, V. K., & Marks, H. M. (2003). Mathematical description of non-linear survival curves of *Listeria monocytogenes* as determined in a beef gravy model system at 57.5 to 65 °C. *Innovative Food Science & Emerging Technologies*, 4(3), 307–317. [https://doi.org/10.1016/S1466-8564\(03\)00025-0](https://doi.org/10.1016/S1466-8564(03)00025-0)
- Juneja, V. K., Marks, H. M., & Mohr, T. (2003). Predictive Thermal Inactivation Model for Effects of Temperature, Sodium Lactate, NaCl, and Sodium Pyrophosphate on *Salmonella* Serotypes in Ground Beef. *Applied and Environmental Microbiology*, 69(9), 5138–5156. <https://doi.org/10.1128/AEM.69.9.5138-5156.2003>
- Korajkic, A., McMinn, B., & Harwood, V. (2018). Relationships between Microbial Indicators and Pathogens in Recreational Water Settings. *International Journal of Environmental Research and Public Health*, 15(12), 2842. <https://doi.org/10.3390/ijerph15122842>
- Little, J.B. (1968). Cellular Effects of Ionizing Radiation. *The New England Journal of Medicine*, 272(7): 369-376.
- Lombardo, L.A., G.L. Grabow, J. Spooner, D.E. Line, D.L. Osmond, and G.D. Jennings. (2000). Section 319 Nonpoint Source National Monitoring Program Successes and Recommendations. NCSU Water Quality Group, Biological and Agricultural Engineering Department, NC State University, Raleigh, North Carolina
- McKellar, R., C., & Lue, X. (2004). Primary Models. In *Modeling Microbial Responses in Food*. CRC Press.

- Medema, G.J., Bahara, M., & Schets, F.M. (1997). Survival of *Cryptosporidium parvum*, *Escherichia coli*, faecal enterococci, and *Clostridium perfringens* in river water: influence of temperature and autochthonous microorganisms. *Water Science and Technology*, 35(11/12):249-252. [https://doi.org/10.1016/S0273-1223\(97\)00267-9](https://doi.org/10.1016/S0273-1223(97)00267-9)
- Mitchell, J. & Akram, S. 2017. "Pathogen Specific Persistence Modeling Data." In: J.B. Rose and B. Jiménez-Cisneros, (eds) *Global Water Pathogens Project*, <http://www.waterpathogens.org> (M. Yates (eds) *Part 4 Management of Risk from Excreta and Wastewater*) Accessible at: <http://www.waterpathogens.org/book/pathogen-specific-persistence-modeling-data>. Michigan State University, E. Lansing, MI, UNESCO.
- National Research Council (NRC). (2004). *Indicators for Waterborne Pathogens*. Washington, D.C.: The National Academies Press. <https://doi.org/10.17226/11010>
- Nomiya, T. (2013). Discussions on target theory: Past and present. *Journal of Radiation Research*, 54(6), 1161–1163. <https://doi.org/10.1093/jrr/rrt075>
- Pachepsky, Y. A., Sadeghi, A. M., Bradford, S. A., Shelton, D. R., Guber, A. K., & Dao, T. (2006). Transport and fate of manure-borne pathogens: Modeling perspective. *Agricultural Water Management*, 86(1–2), 81–92. <https://doi.org/10.1016/j.agwat.2006.06.010>
- Park, Y., Pachepsky, Y., Shelton, D., Jeong, J., & Whelan, G. (2016). Survival of Manure-borne *Escherichia coli* and Fecal Coliforms in Soil: Temperature Dependence as Affected by Site-Specific Factors. *Journal of Environmental Quality*, 45(3), 949–957. <https://doi.org/10.2134/jeq2015.08.0427>
- Prüss-Ustün, A., Wolf, J., Corvalán, C., Bos, R., & Neira, M. (2016). *Preventing Disease Through Healthy Environments: A global assessment of the burden of disease from environmental risks*. World Health Organization. <https://apps.who.int/iris/handle/10665/345241>
- Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M. C., Gordon, B., Hunter, P. R., Medlicott, K., & Johnston, R. (2019). Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene and Environmental Health*, 222(5), 765–777. <https://doi.org/10.1016/j.ijheh.2019.05.004>
- Safe Drinking Water Act: 42 U.S.C. § 300f et seq.
- Shull, J. J., Cargo, G. T., & Ernst, R. R. (1963). Kinetics of Heat Activation and of Thermal Death of Bacterial Spores. *Applied Microbiology*.
- Tamrakar, S. B., Henley, J., Gurian, P. L., Gerba, C. P., Mitchell, J., Enger, K., & Rose, J. B. (2017). Persistence analysis of poliovirus on three different types of fomites. *Journal of Applied Microbiology*, 122(2), 522–530. <https://doi.org/10.1111/jam.13299>

Watson, H. E. (1908). A Note on the Variation of the Rate of Disinfection with Change in the Concentration of the Disinfectant. *Epidemiology and Infection*, 8(4), 536–542.  
<https://doi.org/10.1017/S0022172400015928>

Whiting, R.C. & Buchanan, R.L. 2001. Predictive Modeling and Risk Assessment in *Food Microbiology: Fundamentals and Frontiers (2<sup>nd</sup> Ed)*. ASM Press, Washington, D.C.

World Health Organization. (2017). *Guidelines for Drinking Water* (Fourth). Geneva: World Health Organization. <https://doi.org/10.4060/cb7678en>

World Health Organization. (2019). Safer water, better health (2019 update). Geneva: World Health Organization. <https://www.who.int/publications/i/item/9789241516891>

## **CHAPTER 2: IDENTIFYING WATER QUALITY AND ENVIRONMENTAL FACTORS THAT INFLUENCE INDICATOR AND PATHOGEN DECAY IN NATURAL SURFACE WATERS**

This chapter has been published in *Water Research* and is reprinted with permission from *Water Research* 2022, 211, 1, 118051. Copyright 2022 Elsevier Ltd.

### 2.1 Introduction

Indicator and pathogen decay in surface waters has typically been assumed to follow first-order decay kinetics (de Brauwere et al., 2014; Pachepsky et al., 2006; Crane & Moore, 1986). The simplicity of models such as the exponential or log-linear model and the ease of their application continue to motivate the use of first-order decay kinetics for modeling pathogen fate (Crane & Moore, 1986). However, there is significant evidence in the literature that challenges the assumption of first-order decay (Easton et al., 1999; Benham et al., 2006; Pachepsky et al., 2006; Blaustien et al., 2013; Park et al., 2016). Biphasic decay, in which there are two stages of decay described by two different first-order rates, has been observed for a variety of targets (Easton et al., 1999; Hellweger et al., 2009; Brouwer et al., 2017), and the fitting of alternative model forms in previous analyses within the Global Water Pathogens Project suggest models that capture shouldering and tailing decay dynamics better describe target persistence in various water matrices (Mitchell & Akram, 2017). Assumptions of first order decay kinetics can lead to potential underestimations of residual pathogens.

Deviations from classic linear decay have been hypothesized to be due to population heterogeneity, resistant states, viable but not culturable states, and possibly quorum sensing effects (Brouwer et al., 2017; Easton et al., 2005). Additionally, there are a wide range of environmental, water quality, and species-related factors that influence a target's persistence in a matrix of concern. It is possible that the application of first-order decay kinetics for data analysis in prior studies, has limited our interpretation of the experimental factors influencing persistence.

This is perhaps reflected in distributions of optimized first-order decay rates that have been shown to be highly variable, spanning orders of magnitude within pathogen groups (Boehm et al., 2018).

Simplifications in pathogen fate modeling can possibly lead to erroneous decision making in the fields of water management and protection, and as such, reducing model uncertainty by applying a more appropriate model form to indicator and pathogen decay is an important goal. As natural surface waters are impacted by a number of natural stressors that impact decay, it is hypothesized that two- or three-parameter persistence models will be more representative of indicator and pathogen persistence in surface waters in general than traditional first-order kinetics. Reducing this uncertainty may yield more accurate decay-metrics that can be used to further explore the relationships between indicator and pathogen persistence and typical natural stressors, such as temperature, sunlight, predation, or water composition.

To facilitate the future testing of these hypotheses, this review aims to evaluate the quantity and quality of indicator and pathogen persistence experiments in surface waters available in the literature that could be used to explore the implications of first-order decay kinetic assumptions. This is the qualitative summary of the identified literature and serves to summarize the current state of the knowledge of the water quality and environmental factors affecting indicator and pathogen persistence in surface waters. The relevant methods, modeling techniques, and results from each study described herein will aid future researchers in experimental design and data analysis. This compilation of experiments and methodologies will facilitate the identification of potential interactions that can be further explored to reduce uncertainties associated with pathogen behavior in natural water matrices.



## 2.2 Selection Criteria

For a study to be included in this review, the researchers needed to sample natural surface waters or conduct the experiments *in situ*. The primary objective of this review was the identification of experiments with data for persistence modeling, and as such, the experiments found in the literature needed to document target concentrations or log-reductions values for at least four time points to be selected, as a minimum of four time points would facilitate the fitting of alternative persistence models with more than one parameter. Additionally, the following factors were documented from each study in this review: target type (FIB, bacteriophages, bacteria, viruses, protozoa), water type (fresh, marine, brackish), temperature, sunlight (presence/absence), predation (presence/absence), and method of detection (culture-based/molecular-based). If a light source was not mentioned in the study, it was assumed the experiments were conducted in the dark. If the sampled surface water was filtered or autoclaved in any way, the dataset was identified as predation absent. Studies that included experiments with targets not of focus in this review (MST markers, animal/fish pathogens) were not extracted and included in the review. This literature review describes the results of each study qualitatively; in some cases, not all of the experiments included in a study had enough documented timepoints to facilitate the fitting of models other than those representing first-order decay. Although all the experiments are described qualitatively herein, Table 2.1 only reflects the characteristics of the experiments that have the aforementioned criteria that could be leveraged in future analyses of alternative persistence modeling methods.

## 2.3 Results

The systematic literature review was conducted following PRISMA guidelines using the methods outlined in Appendix A. Sixty-one studies of the 3,949 identified (Fig A.2.1) were included in this review and are summarized in Table 2.1 based on the factors of concern relevant to the experiments within each study that included the necessary detail for modeling.

Experiments were categorized by water type (fresh, marine, brackish), sunlight (presence/absence), predation (presence/absence), and method of detection (culture-based/molecular-based). The following qualitative summaries are divided into target groups: FIB, bacteria, viruses, bacteriophages, and protozoa.

### 2.3.1 Fecal Indicator Bacteria

The majority of the identified literature focused on the persistence or decay of fecal indicator bacteria (FIB) targets, most commonly *Escherichia coli* and *Enterococcus* spp. Of the 39 studies that addressed FIB persistence, 24 assessed FIB persistence independently and 15 compared FIB persistence to that of a pathogen of concern, as shown in Table 2.1. The majority of the identified studies focused on freshwater (27/39) and the inclusion of a light source was a common experimental factor; some or all of the experiments were exposed to light in 26 of the 39 studies. Predation was a less commonly studied manipulated factor and the methods used in the 39 studies were predominantly culture-based.

#### 2.3.1.1 Water Type

Water type did not consistently affect decay in the identified literature. Ahmed et al. (2019) explored the relationship between the decay of qPCR-based targets for *E. coli* and *Enterococcus* spp. in fresh and marine mesocosms and although FIB decay was faster in water than sediment, differences in FIB decay between water types were not noted. Turbidity and

temperature were found to be significantly and positively associated with FIB decay. Sunlight and pH did not have statistically significant correlations. The conclusions about water type echoed those of a previous study that used culture-based methods (Ahmed et al., 2014), in which nonfiltered marine and fresh water were inoculated with raw sewage and the T90s of *E. coli* and enterococci in the two water types were not significantly different.

Water type was a significant factor in two studies employing outdoor mesocosms (Korajkic et al., 2013; Korajkic et al., 2019). In general, Korajkic et al. (2013) observed significantly more decay when indigenous microbiota were present in the mesocosms, and greater decay in marine water experiments than freshwater experiments regardless of indigenous microbiota presence. Water type accounted for 40.1% of the variation in *E. coli* decay, the presence/absence of indigenous microbiota accounted for 49.2% of the variation in decay, and interactions between the two factors (water type and predation) accounted for 9.74% of the variation (Korajkic et al., 2013). Korajkic et al. (2019) monitored both molecular and culturable targets for FIB originating from cattle manure in mesocosms while manipulating water type, sunlight, and predation. Korajkic et al. (2019) concluded that water type was the most influential factor affecting decay, with most targets decaying faster in marine water than fresh. Additionally, there were clear differences between culture and qPCR target decay patterns in freshwater but not in marine (Korajkic et al. 2019). Fujioka et al. (1982) studied fecal coliform and fecal streptococci survival in fresh and marine water and also observed greater persistence in freshwater than marine.

Jeanneau et al. (2012) observed target-specific water type differences in microcosms that simulated contamination from a wastewater input. The first log-reduction (T90) of *E. coli* was 3.4 times lower than in marine water than freshwater, while the T90s for enterococci in the two

water types were not significantly different (Jeanneau et al., 2012). Water type also had conflicting effects in a study conducted by Liang et al. (2017). Although increasing sunlight intensity resulted in increasing decay rates, the effects of salinity on decay were less obvious; increased salinity was associated with increased decay in *E. coli* and decreased decay of enterococci (Liang et al., 2017). Target-specific salinity effects were also observed in a study by Okabe & Shimazu (2007), in which increasing salinity increased the decay of total coliforms in non-filtered river water, but not fecal coliforms (Okabe & Shimazu, 2007).

#### 2.3.1.2 Sunlight

Sunlight was a commonly evaluated environmental factor in both bench-scale and in situ experiments. Korajkic et al. (2014) evaluated how biotic interactions and sunlight influenced FIB persistence in an *in situ* mesocosm located near a recreational beach. The culturable FIB decayed the fastest- there was a strong correlation between the MST markers and the molecular FIB targets, but not the culturable. The effect of sunlight was more pronounced during the initial stages of decay, and over the course of time biotic interactions had a greater influence on decay than sunlight. Both sunlight and predation significantly impacted culturable enterococci, but only sunlight was an influential factor in the decay of culturable *E. coli* (Korajkic et al., 2014). Sunlight was also found to be a primary driver for the inactivation of culturable bacterial indicators in pond water experiments conducted by Greaves et al. (2021) and for the freshwater experiments conducted by Bailey et al. (2018). Temperature and mixing speed, however, were not found to significantly impact FIB decay (Bailey et al., 2018).

Sunlight did not have a pronounced effect on the relative decay of MST markers and *E. coli* in freshwater microcosms where sunlight, temperature, predation and sediment were manipulated, but the high turbidity value of the source water (~102 NTU) and the strength of the

artificial light source were noted as potential causes (Dick et al., 2010). Gutierrez-Cacciabue et al. (2015) studied the effects of both turbidity and sunlight on FIB decay in experiments using culture and qPCR methods. *E. coli* and *E. faecalis* were inoculated into the environmental water with turbidity values of >900 NTU or <3 NTU. There was quite a bit of variability depending on the experiment conditions, but in general sunlight presence caused faster decay, as did the presence of solid particles in the water. The persistence of *E. faecalis* DNA was greater than that of the culturable cells (Gutierrez-Cacciabue et al., 2015). Sunlight was only found to significantly affect the decay rates of enterococci as measured with culture-based methods (as opposed to molecular-based) in the light and dark seawater microcosms constructed by Walters, Yamahara, & Boehm (2009), and the effects of sunlight were found to be target-dependent in the study conducted by Walters & Fields (2009). In mesocosms exploring the decay of MST markers and human and cattle-associated FIB, light influenced the rate of survival more so for enterococci than *E. coli* (Walters & Fields, 2009).

In two of the studies that focused on estuarine water only, sunlight was a significant factor affecting FIB decay (Bordalo, Onrassami, & Dechsakulwatana, 2002; Chandran & Hatha, 2005). Bordalo et al. (2002) noted that the effect of sunlight was more prevalent in higher salinity experiments, suggesting a possible sunlight-water type interaction. Chandran & Hatha (2005) observed significant inactivation of *E. coli* and *S. typhimurium* in estuarine water due to predation and sunlight but deemed sunlight the most important inactivating factor (Chandran & Hatha, 2005).

The effects of sunlight on FIB decay in marine water was more frequently studied than estuarine water. Yukselen et al. (2003) evaluated the effects of temperature and solar radiation on coliform bacteria die-off rates and concluded that sunlight exposure was the most significant

factor affecting decay, and the effect of increasing temperatures was more pronounced in the dark. Mattioli et al. (2017) looked at FIB persistence at different water depths in marine water using molecular and culture-based methods. Both enterococci and *E. coli* decayed faster at shallower depths, where more photoinactivation could occur, although there were some seasonal differences in decay. Noble et al. (2004) addressed the inactivation of various indicators in seawater; temperature and sunlight significantly affected decay, but there were no significant interactions with nutrients, TSS, or initial concentrations. Maracinni et al. (2016) explored the inactivation of enterococci and *E. coli* at different depths (5, 18 and 99 cm) in all three water types- fresh, brackish, and marine. Decay rates were faster in high light conditions than in low light conditions, suggesting endogenous photoinactivation is a major pathway for bacterial decay (Maracinni et al., 2016).

#### 2.3.1.3 Predation

The effects of predation were also commonly assessed in the studies identified in the literature. In the experiments already discussed, predation was associated with significant inactivation of *E. coli* in fresh, marine, and estuarine waters (Dick et al., 2010; Korajkic et al., 2013; Chandra & Hatha, 2005) and culturable enterococci in freshwater (Korajkic et al., 2014). Chandran et al. (2011) completed microcosm studies with *E. coli*, *S. paratyphi*, and *V. parahaemolyticus* in freshwater and sediments. The decay of *E. coli* in sterile water was slower than that in nonfiltered water, highlighting a significant effect from the presence of biological factors. Wanjugi, Fox, and Harwood (2016) used *E. coli* to explore interactions between predation, nutrient levels, and competition on target survival in river water microcosms set-up in an open greenhouse. Overall, predation and competition had negative effects on survival, while

nutrient addition increased survival. Specifically, predation accounted for the greatest amount of decay variation (40%), followed by nutrients (25%) and competition (15%).

Solecki et al. (2011) compared FIB persistence to the persistence of various microbial and chemical pig manure markers in dark fresh and marine microcosms and the observed biphasic decay was hypothesized to be due to predation, as the waters used were unfiltered. Both generic and pathogenic *E. coli* were studied in agricultural surface water microcosms constructed by Topalcengiz & Danyluk (2019) and the most rapid decay was observed in the non-sterile waters. Medema, Bahar, & Schets (1997) investigated the persistence of *E. coli*, *Clostridium perfringens*, and enterococci in autoclaved and natural river water microcosms that were maintained in the dark. Die-off of *E. coli* and enterococci were faster in the natural river water at both tested temperatures, with possible multiplication of both FIB observed in the autoclaved water maintained at 15°C (Medema et al., 1997).

Ahmed et al. (2021) explored the persistence of various targets, including FIB with culture-based techniques. The mesocosms were constructed with water sampled from two lakes, either filtered or left natural, and kept at 15°C or 25°C for the experiments. There was not a pronounced effect from predation, however this was hypothesized to be likely due to the sewage inoculum which could have introduced predators to the filtered water experiments. For the *E. coli* trials, decay was significantly faster at 25°C than 15°C in three of the four microcosms (Ahmed et al., 2021).

#### *2.3.1.4 Other Factors*

Although the majority of the studies identified focused on water type, sunlight or predation as the influencing factors, a few of the studies assessed site-specific FIB decay (Irankhah et al., 2016), or modified variables such as temperature, mixing speeds, initial FIB

loads, and sediment or vegetation inclusion. Increasing temperatures were associated with increased decay of *E. coli* in chlorinated and untreated lake water (Lund, 1996), fresh water (Terzieva & McFeters, 1991; Easton et al., 2005; Nasser et al., 2003), brackish water and marine water (Nasser et al., 2003). Mixing speed of fresh surface water samples did not significantly affect the decay of indigenous microbiota (Bailey et al., 2018), and the initial seed concentration of *E. coli* was shown to affect decay rates in fresh and brackish water studies (Beckinghausen et al., 2014; Gronewold et al., 2011). When comparing the decay of human and bovine *E. coli* in freshwater microcosms, Liang et al. (2012) concluded that human *E. coli* decayed faster than bovine. Zhang, He and Yan (2015) monitored the decay of FIB in seawater and beach sand microcosms. Enterococci and *C. perfringens* decayed more slowly in beach sand than water, but *E. coli* decayed similarly in water and sand matrices. Mezrioui et al. (1995) looked at *E. coli* and *S. typhimurium* decay in brackish water and sewage mixtures, mimicking slow or rapid marine stress. Survival was greater when predators were removed in the fall and summer experiments, but not in the winter experiments (Mezrioui et al., 1995).

Four of the identified studies addressed the influence of vegetation or natural aquatic plant life on the persistence of FIB. Submerged aquatic vegetation indirectly facilitated the persistence of enterococci in freshwater microcosms (Badgley et al., 2010), algae presence in freshwater microcosms led to greater persistence under UV conditions of *E. coli* and *Salmonella enterica* serovar Typhimurium (Beckinghausen et al., 2014) and the presence of wrack was associated with increased levels of FIB in water and sediments (Imamura et al., 2010). Tiwari, Kauppinen, & Pitkanen (2019) compared decay of indicators and *Vibrio* spp. in a brackish beach mesocosm with and without the presence of an aquatic plant, *Myriophyllum sibiricum*. The



molecular *Enterococcus* spp. markers decayed more slowly than the culturable enterococci, and biphasic decay was observed for both culturable and molecular targets in water.

#### 2.3.1.5 Summary

In summary, some studies noted differential decay of FIB based on water type: greater decay of FIB was observed in marine waters for fecal coliforms, streptococci, and *E. coli* (Fujioka et al., 1982; Korajkic et al., 2013; Korajkic et al., 2019; Liang et al., 2017; Jeanneau et al., 2012). The literature also suggests the possibility of interactions between water type and sunlight presence and water type and method; one study found that the effects of sunlight were more prevalent in higher salinity waters (Bordalo et al., 2002) and another identified differences in decay rates between molecular and culture-based targets only in freshwater matrices (Korajkic et al., 2019). The majority of the studies found that sunlight significantly accelerated decay (Bailey et al. 2018; Bordalo et al., 2002; Chandran & Hatha, 2005; Fujioka et al., 1982; Korajkic et al., 2014; Gutierrez-Cacciabue et al., 2013; Greaves et al., 2021; Liang et al., 2017; Maracinni et al., 2016; Mattioli et al., 2017; Noble et al., 2004; Yukselen et al., 2004), however there were a few studies that determined it was not a significant factor or found its impact to be variable (Ahmed et al., 2019; Dick et al., 2010; Korajkic et al., 2019; Walters & Fields 2009; Walters et al., 2009). This was hypothesized to be due to the water having a high turbidity or a potential shading issue (Dick et al., 2010; Korajkic et al., 2019).

In general, increasing temperature was correlated with increasing decay (Ahmed et al., 2019; Lund et al., 1996; Easton et al., 2005; Nasser et al., 2003; Noble et al., 2004; Terzieva & McFeters, 1991; Yukselen et al., 2003; Medema et al., 1997), however a few studies did not observe this typical pattern (Bailey et al., 2018; Okabe & Shimanzu, 2007). Predation was also commonly found to be a significant factor affecting decay (Chandran & Hatha, 2005; Chandran

et al., 2011; Dick et al., 2010; Korajkic et al., 2013; Topalcengiz & Danyluk, 2019; Wanjugi et al., 2016; Medema et al., 1997), but in some cases the effects of predation were unclear or FIB-dependent (Korajkic et al., 2019; Korajkic et al., 2014). When molecular-based methods were used, they were typically associated with slower decay (Gutierrez-Cacciabue et al., 2015; Frankhah et al., 2016; Korajkic et al., 2019; Walters et al., 2009), but Korajkic et al. (2019) only found this to be the case in freshwater matrices. Turbidity was a significant water quality parameter affecting decay (Ahmed et al., 2019; Gutierrez-Cacciabue et al., 2015) and more generally, the presence of vegetation or algae in the water was also associated with greater persistence (Badgley et al., 2010; Beckinghausen et al., 2014; Imamura et al., 2010).

### 2.3.2 Bacteria

Pathogenic bacteria were the next target type most commonly studied in the identified literature. Twenty studies focused on the persistence of bacteria, with targets such as pathogenic *E. coli* strains, *Salmonella* spp., *Vibrio* spp., *Listeria* spp., *Campylobacter* spp., *Staphylococcus* spp., and *Yersinia enterocolitica*. Freshwater was the water type studied most frequently (15/20) and the majority of the studies conducted experiments in the dark (12/20). Varying predation status was slightly more variable across the studies and the method type was the least variable documented factor; 18 of the 20 identified studies quantified the pathogenic bacteria concentrations with culture-based methods.

#### 2.3.2.1 Water Type

The two studies that compared bacteria decay in different water types observed some potential differences (Boehm et al., 2012; Levin-Edens et al., 2011). Levin-Edens et al. (2011) evaluated the persistence of methicillin-resistant *Staphylococcus aureus* in sterilized marine and freshwater and both temperature and salinity were found to significantly influence decay. Decay

rates were higher at the higher temperature and in freshwater. Boehm et al. (2012) addressed the inactivation of *Salmonella* serovars in filter-sterilized fresh and marine water matrices, with sunlight as a stressor. Exponential decay was observed for the serovars studied in seawater under light conditions, but little to no decay was observed for the serovars studied in seawater under dark conditions. A similar effect from sunlight was observed in freshwater, however, the freshwater experiments exposed to sunlight exhibited a clear shoulder effect not observed in the marine water. The shoulder period was only 1-1.5 hour in length, and the observed T90s were less than 1 day in length, indicating that the potential sunlight-water type interaction may not greatly impact decay observations (Boehm et al. 2012).

#### 2.3.2.2 Sunlight

Beckinghausen et al. (2014) also observed rapid decay for *Salmonella. S. typhimurium* had T90 values of 1 day or less in freshwater microcosms exposed to natural sunlight regardless of initial seed concentration, and results indicated that the pathogen persists longer than the studied indicator. As highlighted in section 2.3.1.4, the presence of algae in the microcosm, the main factor studied by Beckinghausen et al. (2014), led to greater persistence for both *E. coli* and *S. typhimurium*. The T90s for *S. typhimurium* in raw estuarine water maintained in the dark at 20 and 30°C were less than 2 days and less than 1 day, respectively (Chandran & Hatha, 2005). The presence of biological factors (raw water) was deemed an important influence on persistence; however sunlight caused the greatest inactivation in the experiments overall (Chandran & Hatha, 2005).

Rodriguez & Araujo (2011) specifically looked at *Campylobacter* persistence in river water microcosms in situ and in the laboratory. Temperature and sunlight were found to significantly affect decay in the *in-situ* experiments, but pH, oxygen concentration and water

conductivity did not. Silvester et al. (2021) explored the survival kinetics of five *Vibrio* species in a study that looked at the effect of biotic factors, protozoan grazing, temperature, salinity, sunlight, and chemical composition in estuarine water. The researchers observed better survival in sediment versus water, and the biological factors, chemical composition, and sunlight increased the removal of the *Vibrio* species.

#### 2.3.2.3 Predation

Avery et al. (2008) evaluated *E. coli* O157:H7 persistence in a variety of water sources including lake and river waters. Sampled waters were autoclaved or non-autoclaved, inoculated with *E. coli* O157:H7, and kept at 10°C. The authors concluded that sterilization was significantly correlated with decay (Avery et al., 2008). Topalcengiz & Danyluk (2019) examined the fate of generic and pathogenic *E. coli* in agricultural surface water microcosms and the bacteria decreased most rapidly in non-sterile waters. Wang & Doyle (1998) sampled water from a filtered and autoclaved municipal source, a reservoir and two recreational lakes, inoculated *E. coli* O157:H7, and monitored decay at 8, 15 and 25°C in the dark. For all water types, survival was greatest at 8°C and the greatest survival was observed in the filtered and autoclaved municipal water (Wang & Doyle, 1998).

The effects of predation on other pathogens, such as *S. paratyphi* and *V. parahaemolyticus*, were observed in a study by Chandran et al. (2011). The T90s for the pathogens were more rapid in the nonsterile waters, illustrating the effect of predation or other biological factors (Chandran et al., 2011). Silvester et al. (2021) also observed higher mortality rates of *Vibrio* spp. in raw sediments and waters compared to autoclaved sediments and waters. The experiments conducted by Lund et al. (1996) used untreated and autoclaved lake water that was kept at 4°C or 10°C in the dark. Survival was better at 4°C, and *Y. enterocolitica* survived

better at lower temperatures than *C. jejuni*. *Y. enterocolitica* also survived much longer in autoclaved water than untreated water. There was possibly some seasonal variation in the effects of predation as observed by Mezrioui et al. (1995), as the survival of *S. typhimurium* was greater when predation was absent in the fall and summer experiments, not in the winter.

Predation did not play a key role in the only study identified in this review that used molecular-based methods to quantify bacteria persistence. In addition to FIB and MST markers, Ahmed et al. (2021) explored *Campylobacter* spp. persistence with freshwater mesocosms exposed to artificial sunlight conditions. *Campylobacter* spp. decayed linearly, at faster rates at 25°C than at 15°C; and at similar rates between filtered and nonfiltered experiments, suggesting predation did not significantly impact decay. As noted in the FIB section, although the water was filtered for the predation ‘absent’ experiments, the sewage inoculum still had the potential to introduce predators, potentially masking the effect of predation in this study. Although it is hard to make any inferences given the limited number of studies, the use of the sewage inoculum, and the different targets between the studies (*Campylobacter* spp. vs *E. coli*, *S. paratyphi*, *Vibrio* spp. and *Y. enterocolitica*), predation not significantly affecting decay in Ahmed et al. (2021) in the context of this review suggests a potential interaction between predation and the method of detection (culture vs. molecular-based) for pathogenic bacteria.

#### 2.3.2.4 Other

Ibrahim et al. (2019) studied *E. coli* O157:H7, *S. Typhimurium*, human adenovirus serotype 2, and murine norovirus 1 survival while stored in sterilized river water microcosms stored at -20°, 4°, 24°C and 37°C in the dark. The highest T90 values for the bacterial pathogens were observed in the 4°C trials. Higher temperatures were associated with faster decay of *Listeria monocytogenes* in river water sampled near the outfall of a meat industry plant

(Budzinksa et al., 2012), *E. coli* O157:H7 in sampled creek water (Easton et al., 2005), and *E. coli*, *C. jejuni*, and *Y. enterocolitica* in sampled stream water (Terzieva & McFeters, 1991). *C. jejuni*, however, was less affected by temperature, and *Y. enterocolitica* had the greatest survival at both temperatures (Terzieva & McFeters, 1991).

Some of the identified studies addressed the persistence of a bacterial target without manipulating any water or environmental factor directly, and others studied alternative water quality factors- changing salinities, aquatic plant presence or turbidity. El Mejri et al. (2012) studied the survival of environmental and laboratory-adapted strains of *Salmonella enterica* serovar Typhimurium in marine water and the authors calculated an average T90 of 25-30 hours for all the strains. Notably, total cell count data suggested the bacteria were entering a viable but not culturable state (El Mejri et al., 2012). In brackish beach mesocosms, *Vibrio* spp. genetic markers exhibited biphasic decay, and there were positive correlations between the culturable *Enterococcus* targets and *V. cholerae* molecular targets only for the first few days of the experiment, suggesting a limitation in current monitoring practices (Tiwari et al., 2019). Except for *Vibrio* rRNA, the decay of the bacterial targets was greater in water than sediment and vegetation (Tiwari et al., 2019). Turbidity-related factors were significant in a study conducted by Czajkowska et al. (2005). *E. coli* serotype O157:H7 was inoculated into freshwater sampled from several different Polish lakes and rivers and maintained at 4 or 24°C. The researchers conducted experiments using water, “muddy” water, and bottom shore sediments. The bacteria survived slightly longer in muddy water experiments, and the survival times were shorter for the experiments maintained at 24°C. The authors did not detect a significant effect from the varying chemical oxygen demand or pH levels (Czajkowska et al., 2005).

#### 2.3.2.5 Summary

Looking at the selected literature as a whole for pathogenic bacteria, some expected relationships were consistently observed. In general, decay is higher at higher temperatures as expected (Levin-Edens et al., 2011; Rodriguez & Araujo, 2011; Wang & Doyle, 1998; Budzinska et al., 2012; Ahmed et al., 2021; Lund et al., 1996; Czajkowska et al., 2005; Terzieva & McFeters, 1991; Easton et al., 2005). The presence of a light source consistently accelerates decay (Beckinhausen et al., 2014; Chandran & Hatha, 2005; Rodriguez et al., 2011; Silvester et al., 2021), however one study identified some potential differences between the effect of sunlight in fresh and marine waters (Boehm et al., 2012). Predation was found to significantly increase decay in a variety of targets including pathogenic *E. coli* strains, *Salmonella* spp., *Vibrio* spp., and *Y. enterocolitica* (Avery et al., 2008; Chandran et al., 2011; Lund et al., 1996; Silvester et al., 2021; Topalcengiz & Danyluk, 2019; Wang & Doyle, 1998). The effects of predation were less obvious in a study by Mezriouia et al. (1995) in which *S. typhimurium* survival was only greater in sterile waters during autumn and summer seasons, and predation was not found to influence the decay of *Campylobacter* spp. at all in a freshwater mesocosm study (Ahmed et al., 2021). Although it was evaluated directly (and indirectly) less in the pathogenic bacteria literature than the FIB literature, turbidity seems to again be a factor potentially influencing decay (Czajkowska et al., 2005). Other water quality factors such as pH, dissolved oxygen, and water conductivity, were not found to significantly affect the inactivation of pathogenic bacterial targets (Rodriguez & Araujo, 2011; Czajkowska et al., 2005).

#### 2.3.3 Bacteriophages

Eight of the studies identified in the literature review assessed the persistence of bacteriophages in surface water matrices. Six of the studies used freshwater, marine water was

studied in two studies, and one study used brackish water. Sunlight was absent from most of the experiments, and predation was predominantly present. All eight studies used culture-based methods to quantify the bacteriophage concentrations over time.

Booncharoen et al. (2018) evaluated the persistence of human sewage-specific enterococcal bacteriophage persistence from the *Myoviridae* (A2), *Podoviridae* (S1) and *Siphoviridae* (A1, S4) families in highly and lowly populated fresh and marine water sources in Thailand. The highest decay rates of A1, A2, and S4 were in the highly polluted seawater matrix; the highest decay rate of S1 was in the highly polluted freshwater matrix. All bacteriophages exhibited slower decay in filtered samples and in lower pollution waters, suggesting predation does play an important role in bacteriophage persistence. Additional experiments at 5°C indicated that temperature significantly affects bacteriophage decay. The effects of salinity on decay varied with the pollution level of the water (Booncharoen et al., 2018). Marine water was only tested by one additional study, in which the marine bacteriophage H6 was inoculated into unfiltered marine water and maintained at 22°C (Olive et al., 2020). Minimal decay was observed over the 16-day experiments, suggesting that the existence of predators in the water affected the bacteriophage decay minimally (Olive et al., 2020).

Somatic coliphages from the *Myoviridae* and *Siphoviridae* family were also evaluated in another study identified in this review (Lee & Sobsey, 2011). Lee and Sobsey (2011) conducted experiments using reagent water and natural surface waters sampled from a freshwater body. Experiments were conducted at 4°C and 25°C. The analysis suggested that water type (reagent vs fresh), temperature, and incubation time were strong predictors of inactivation. For most of the coliphages, decay was slower at 4°C (Lee & Sobsey, 2011). Greaves et al. (2021) included a somatic coliphage in their large-scale freshwater mesocosm experiments. The mesocosms were



either exposed to sunlight or left covered, and somatic coliphage was detectable for two days in the uncovered experiments and five days in the covered experiments, indicating that sunlight significantly affects decay. Notably, the viral indicators in this study were more resistant to UV inactivation than the bacterial indicators studied (Greaves et al., 2021).

Long & Sobsey (2004) addressed the persistence of F+RNA and F+DNA coliphages in freshwater microcosms. Water was sampled from a surface drinking water source, spiked with an individual phage to a target concentration, and maintained at 4°C or 20°C in the dark. The authors noted tailing in the survival curves and for all coliphages, decay was slower at the lower temperature. The differences between the temperatures were statistically significant for the F+DNA coliphages and the Group II and Group IV F+RNA coliphages (Long & Sobsey, 2004). Ravva & Sarreal (2016) evaluated the persistence of F-specific RNA coliphages in the presence and absence of bacterial host in fresh surface waters. The experiments were kept in the dark and maintained at temperatures that simulated summer (25°C) or winter (10°C) conditions. The addition of the host resulted in greater phage persistence, with phages in the absence of host disappearing relatively rapidly. Temperature significantly affected persistence, as greater survival was observed at 10°C as compared to 25°C, and there were significant differences between environmental and prototype isolate survival. The chemical composition of the waters did not appear to affect decay; although QB was assayed in waters with higher suspended solids than the other strains, it decayed similarly (Ravva & Sarreal, 2016).

QB may not be representative of coliphage-suspended solids interactions as a whole, however, as it was an exception in a study conducted by Yang & Griffiths (2013) using river water experiments. The study manipulated temperatures, pH values, and some samples were autoclaved and filtered. Except for QB, all the other F+RNA phages persisted for longer in the

heat-treated water (predation absent) and were less stable when suspended solids were removed. The researchers concluded that temperature and pH were major factors that affected the phage survival (Yang & Griffiths, 2013). Tiwari, Kauppinen, and Pitkanen (2019) detected MS2 with culture-based methods in their study of decay patterns in brackish beach mesocosms. MS2 decayed log-linearly and had higher decay in the mesocosms without vegetation. MS2 counts was strongly correlated with culturable enterococci for the first 6-8 days (Tiwari et al., 2019).

#### *2.3.3.1 Summary*

Overall, the studies identified in the literature provided limited insight into the factors potentially affecting decay. Temperature was the most frequently addressed factor, and it was consistently determined to significantly affect bacteriophage decay (Booncharoen et al., 2018; Greaves et al., 2021; Lee & Sobsey, 2011; Long & Sobsey, 2004; Ravva et al., 2016; Yang et al., 2013). Sunlight was shown to significantly increase decay rates of bacteriophages, but bacteriophages may be more resistant to UV inactivation than FIB (Greaves et al., 2021). Predation significantly impacts bacteriophage decay, although the impact may be sub-type dependent (Booncharoen et al., 2018; Yang et al., 2013), and this review did not select any studies that suggested significant differences between bacteriophage decay rates in different water types (Booncharoen et al., 2018). As only culture-based methods were used in the described studies, no conclusions could be drawn about method-related decay differences. However, as opposed to the conclusions drawn in the FIB and Bacteria *Other Factors* sections (2.3.1.4 and 2.3.2.4), the bacteriophage literature suggests that pH does significantly affect decay (Yang & Griffiths, 2013).

#### 2.3.4 Viruses

There were 11 studies identified in the literature review that studied the persistence of viruses. Nine of the 11 studies used freshwater, six used marine water and only one used brackish water. The presence or absence of sunlight was fairly variable, with three studies exposing all experiments to sunlight, four studies keeping all experiments in the dark, and four studies manipulating the presence of sunlight. Predation was predominantly present in the experiments assessing virus persistence and the virus target group had the greatest variety of method type; four of the eleven studies used molecular-based methods.

Adenovirus was the virus-type most frequently studied in the identified literature. Ahmed et al. (2014) assessed adenovirus persistence and the T90s in fresh and marine waters were 13 and 9.4 days, respectively. Liang et al. (2017) assessed the effects of salinity and sunlight on adenovirus and their results contradicted those of Ahmed et al. (2014), in that increasing salinity was associated with greater adenovirus persistence. Increasing sunlight intensity increased adenovirus decay rates, and results suggested that the viral targets in this study were more susceptible to sunlight than their bacterial counterparts. Regardless of salinity or sunlight, intact cells decayed faster than total DNA (Liang et al., 2017).

Human adenovirus was also studied in freshwater mesocosms that manipulated predation status and temperatures (Ahmed et al., 2021). In one set of the experiments, the decay rates were similar between predation treatments, with higher decay rates observed at 25°C compared to 15°C. In the other set of experiments, the effects of the increased temperature were only prominent in the filtered treatments (Ahmed et al., 2021). The effect of temperature on virus persistence was also evaluated in a study using dark, sterilized river water microcosms stored at -20°C, 4°C, 24°C or 37°C (Ibrahim et al., 2019). Adenovirus had the slowest decay in -20°C with

a T90 of almost 200 days and the fastest decay at 37°C with a T90 of 27 hours. Norovirus was not affected by increasing temperatures in the typical manner with decay rates from fastest decay occurring in the 24°C trials and the slowest decay occurring in the 4°C (Ibrahim et al., 2019).

Carratalà et al. (2013) evaluated the persistence of adenoviruses in a range of water matrices, but the experiments conducted in seawater were the only ones that met the natural surface water criteria for the review. Human adenovirus type 2 was inoculated into the seawater reactors and dark conditions were tested at 7°C, 20°C and 37°C, and experiments were conducted with UVB and UVA light sources (both classified as sunlight present). For the dark experiments, no significant inactivation was observed at 7°C or 20°C, but higher inactivation was observed at 37°C. The authors concluded that biotic factors may be more relevant to virus inactivation than indirect photo-inactivation by UVA radiation, especially at higher temperatures (Carratalà et al., 2013).

Olive et al. (2020) investigated the microbial control of echovirus 11, adenovirus 2, and the bacteriophage H6 decay in different water types. Echovirus 11 was incubated in sterilized and non-sterilized lake water maintained at room temperature, and echovirus 11, adenovirus 2, and the bacteriophage H6 were inoculated into the eukaryotic fraction isolated from lake and ocean water. The eukaryotic fraction and bacteria fraction waters discussed in this study were merely classified as “predation present” for this review. The authors concluded that the microbial virus control was temperature dependent, with more obvious reductions at 22°C as compared to 16°C, dependent on the virus (Olive et al., 2020).

Bergstein et al. (1996) evaluated the persistence of poliovirus with *in situ* experiments in a freshwater lake located at different depths with variable light exposure, and during different seasons (winter/summer). Very little poliovirus decay was observed during the winter

experiments, regardless of light presence. In the summer experiment, light exposure resulted in a 1-log reduction within ~4 days, compared to minimal observed decay in the dark. Nasser et al. (2003) assessed the persistence of coxsackie A9 virus in fresh, marine, and brackish waters and virus persistence was found to be temperature-dependent, with faster decay observed at 30°C than 15°C. Die-off of coxsackie A9 virus at 15°C and 30°C was greatest in marine water (Nasser et al., 2003).

The survival of Rhesus rotavirus and human astrovirus was characterized in both groundwater and contaminated surface water in a study conducted by Espinosa et al. (2008). Only the surface water experiments met the “natural” waters criteria outlined in the Appendix. Results indicated that virus infectivity persisted longer in groundwater than surface water, and that rotavirus persistence was more stable than astrovirus. Good correlation between virus infectivity and genomic material detection was noted (Espinosa et al., 2008). The decay of enteroviruses using molecular-based methods was assessed in light and dark seawater microcosms in a study conducted by Walters, Yamahara, & Boehm (2009). The infectious enterovirus remained detectable longer in the dark microcosms than the light. No difference in genome decay rates was observed between the two treatments, suggesting a potential method-sunlight interaction for virus targets (Walters et al., 2009). de Oliveira et al. (2021) assessed the viability of SARS-CoV-2 in filtered and nonfiltered river water experiments kept at 4°C and 24°C. Nonfiltered and higher temperature river water experiments were associated with faster decay (de Oliveira et al., 2021).

#### *2.3.4.1 Summary*

Overall, the results of the studies that manipulated water type suggest that the decay of viruses may be faster in marine waters than in fresh (Ahmed et al., 2014; Nasser et al., 2003;

Walters et al., 2009), although Liang et al. (2017) observed greater persistence with increasing salinities. In general, sunlight increased viral decay (Carratalà et al., 2013; Bergstein et al., 1997; Liang et al., 2017), but there was no difference between genome decay rates in a study conducted by Walters et al. (2009). Similarly to the bacteriophage literature, the studies identified herein suggest that the relationship between virus persistence and predation may be complicated. de Oliveira et al. (2021) observed clear differences between SARS-CoV-2 persistence when predation was present versus absent, but the effects of predation on adenovirus persistence were unclear (Ahmed et al., 2021). Olive et al. (2020) concluded that the microbial control of virus populations may be temperature-dependent, as the differences between the predation status was less obvious at lower temperatures. The effects of temperature seem to be relatively consistent with the rest of the aforementioned targets, with the exception of norovirus. Most studies found that decay increased with increasing temperatures (Ahmed et al., 2014; Ahmed et al., 2021; Carratalà et al., 2013; Nasser et al., 2003; de Oliviera et al., 2021), however norovirus in one study had the highest decay rate at the second highest temperature (24°C) (Ibrahim et al., 2019).

#### 2.3.5 Protozoa

Protozoa were the least frequently studied target in the literature identified in this review. Only four studies assessed the persistence of protozoa, and as such, there was minimal water quality and environmental factor variation to facilitate comparisons.

Multiple water matrices (fresh, brackish, and marine) were used by Nasser et al. (2003) to test the persistence of *Cryptosporidium* at two different temperatures (15 or 30°C). The persistence of *Cryptosporidium* was not influenced by temperature in any of the water types. In general, little to no decay was observed for *Cryptosporidium* across all the experiments, and it was concluded that *E. coli* was not a suitable indicator for *Cryptosporidium*. Robertson et al.

(2006) studied *Cryptosporidium* oocyst and *Giardia* cyst persistence in a freshwater river. The river water temperature throughout the experiment fluctuated from 1.1°C to 7.3°C. River water was also sampled for a control experiment maintained in the refrigerator (4°C). *Cryptosporidium* oocysts were detected as viable up until ~20 weeks, and *Giardia* cysts were detected until about 1 month. Comparisons between control and river environments suggest that temperature changes, and other physical, chemical and biological factors do not significantly impact the decay of the studied parasites (Robertson et al., 2006).

Medema et al. (1997) evaluated the persistence of *Cryptosporidium parvum* in autoclaved and natural river water microcosms that were maintained in the dark. The effects of predation were only evident at 15°C, but not at 5°C. Die-off of *C. parvum* was faster at the higher temperature (Medema et al. 1997). The freshwater mesocosms constructed by Ahmed et al. (2021) also evaluated the persistence of *C. parvum*. *C. parvum* genome decay was best described by a biphasic model, and a greater reduction was observed at 25°C than 15°C for all trials. Vital dye assays were used to assess *C. parvum* viability, and the relationship with temperature and filtration status was found to be variable.

#### 2.3.5.1 Summary

The four studies addressing protozoa persistence described herein suggest that *Cryptosporidium* is relatively resistant to the factors that have been shown to increase decay rates of the other target types as expected. Two of the studies found no increase in decay rate with increased temperatures, and decay was similar between the sunlight present and absent studies conducted by Robertson et al. (2006). The study conducted by Medema et al. (1997) suggests that *C. parvum* may be affected by predation at some temperatures, again suggesting interactions between predation and temperature may be significantly affecting the decay of pathogens in

surface waters. However, the other study assessing predation suggested there was minimal impact from predation presence (Ahmed et al., 2021).

## 2.4 Discussion

### 2.4.1 Observed Decay Kinetics

The majority of the studies identified in this review used first-order kinetics to describe the observed decay. A number of the studies described herein, however, observed biphasic decay patterns (Ahmed et al., 2019; Dick et al., 2010; Solecki et al., 2011; Zhang et al., 2015; Mattioli et al., 2017; Tiwari et al., 2019; Carratalà et al., 2013; de Oliveira et al., 2021; Easton et al., 2005; Medema et al., 1997). Some of these studies noted biphasic decay but only fit log-linear models, while others fit alternative model forms. Of the 61 studies included in this review, less than 20% analyzed their data with biphasic or nonlinear model forms in addition to the traditionally assumed first-order decay kinetic profiles (Ahmed et al., 2019; Ahmed et al., 2021; Bailey et al., 2019; Carratalà et al., 2013; de Oliveira et al., 2021; Jeanneau et al., 2021; Lee & Sobsey, 2011; Mattioli et al., 2017; Solecki et al., 2011; Zhang et al., 2015). The most frequently applied model form in these studies was the so termed “biphasic decay model”, represented by Equations 1 and 2, where  $C(t)$  is the concentration at time  $t$ ,  $C_0$  is the concentration at time 0,  $t'$  is the time point where the second phase of decay begins, and  $k_1$  and  $k_2$  are the first-order decay constants for the two phases (Zhang et al., 2015; Ahmed et al., 2019; Ahmed et al., 2021). The biphasic decay model provided a good fit to cultured FIB in marine water (Zhang et al., 2015), molecular targets for FIB in fresh and marine water (Ahmed et al., 2019), and *C. parvum* molecular targets in freshwater (Ahmed et al., 2021).

$$\ln\left(\frac{C_t}{C_0}\right) = -k_1 t, \quad t \leq t' \quad \text{Eq. 1}$$

$$\ln\left(\frac{C_t}{C_0}\right) = -k_1 t - k_2(t - t'), \quad t > t' \quad \text{Eq. 2}$$



Solecki et al. (2011) also observed biphasic decay in their FIB experiments, however, the model applied to the data slightly differed from Equations 1-2. As shown in Equation 3, the two first-order decay rates were applied to different proportions of the initial population of  $C_0$ , designated with the parameter,  $f$ . The observed biphasic decay was hypothesized to be due to rapid die-off until a carrying capacity is reached, or the microorganisms using quorum sensing to regulate their numbers (Solecki et al., 2011). The same model was applied to the decay of FIB (Jeanneau et al., 2012), somatic coliphages (Lee & Sobsey, 2011) and adenovirus (Carratalà et al., 2013), albeit with  $f$  defined as a mixing parameter (designated as  $\omega$ ) in the study of adenovirus inactivation.

$$C(t) = C_0 \times (f \times e^{-k_1 t} + (1 - f) \times e^{-k_2 t}) \quad \text{Eq. 3}$$

Mattioli et al. (2017) fit the shoulder log-linear model to FIB persistence experiments and found that the shoulder log-linear model provided the best fit to enterococci molecular data. Instead of a second decay rate, as is the case for Equations 1-3, the shoulder log-linear model's second parameter,  $S$ , represents the shoulder or lag time where there is minimal inactivation (Eq. 4). The other studies that applied nonlinear model forms, such as the exponential biphasic, Weibull, and Gompertz models, found that the nonlinear models provided better fits to virus inactivation data under all tested conditions (de Oliveira et al., 2011), and bacteriophage inactivation data from the experiments evaluating the effect of their highest tested temperature (25°C vs 4°C) (Lee & Sobsey, 2011).

$$C(t) = C_0 e^{-kt} \left( \frac{e^{-kS}}{1 + (e^{kS} - 1)e^{-kt}} \right) \quad \text{Eq. 4}$$

Of the few studies discussed herein that analyzed the fit of both nonlinear and linear models to persistence data, the nonlinear models were found to provide a good fit to FIB, bacteriophage, virus, and protozoa data from experiments representing different water types,

temperatures, and methods of detection. Biphasic decay has also been described in other persistence studies regarding matrices with conditions that did not meet the requirements of the review herein (Easton et al., 1999; Park et al., 2016). These studies and the results of previous modeling studies (Mitchell & Akram, 2017; Dean et al., 2020) suggest that two or three-parameter models are better able to capture the dynamics of decay than the traditionally assumed first-order kinetics. One of the aims of this review was to determine the quantity of available pathogen and indicator persistence data for surface waters that could be explored in analyses that expand past the assumption of first-order decay kinetics. This review differed from some of the previous systematic literature reviews in that the studies had to include raw data that met the requirements for model fitting (more than three time points), as opposed to previous reviews that aimed to extract reported first-order decay rates (Boehm et al., 2018; Boehm et al., 2019). The 61 studies described herein represent over 600 experiments available in the literature for this purpose.

#### 2.4.2 Environment and Water Quality Factor Interactions

In addition to quantifying the available data in the literature for persistence modeling purposes, this review also aimed to assess the current state of the knowledge regarding the water quality and environmental factors that impact indicator and pathogen decay in surface waters. Previous factor analyses have used dependent variables derived from the assumed first-order decay kinetics (Boehm et al., 2018; Boehm et al., 2019). It is possible that the reduction in model uncertainty through the application of more accurate persistence models will also facilitate factor analyses capable of elucidating the finer relationships between environment and water quality factors affecting target persistence. Previous analyses that fit alternative persistence models to a

large database of experiments did not complete an adjoining factor analysis (Mitchell & Akram, 2017).

The studies discussed herein analyzed the effect of water quality and environmental variables with a variety of methods. Correlation coefficients and analyses (Ahmed et al., 2019; Tiwari et al., 2019; Korajkic et al., 2014; Espinosa et al., 2008), two-way and three-way ANOVAs (Korajkic et al., 2013; Korajkic et al., 2019; Wanjugi et al., 2016; Avery et al., 2008), multiple linear regression (Liang et al., 2017; Levin-Edens et al., 2011), and generalized linear mixed models (Wanjugi et al., 2016) were some of the methods implemented.

As illustrated in Figure 2.1, the factors most frequently addressed by the studies identified in this literature review were temperature, sunlight, and predation. Temperature consistently was associated with decay for FIB, pathogenic bacteria, bacteriophages, and viruses, but increasing temperatures had little effect on protozoan decay in the identified literature (Robertson et al., 2006; Nasser et al., 2003). The presence of artificial or natural sunlight was frequently one of the main drivers of inactivation for the studied target types (except for protozoa), however FIB may be more susceptible to UV inactivation than viral indicators (Greaves et al., 2021). The FIB and pathogenic bacteria literature suggest that there may be possible interactions between sunlight and water type (Boehm et al., 2012; Bordalo et al., 2002) as well as sunlight and method of detection (culture vs. molecular-based) (Walters et al., 2009; Korajkic et al., 2019). It is important to note however, that other factors such as turbidity, UV wavelength, or water depth that were not consistently documented may be impacting the interpretation of the effects of sunlight presence on target persistence. The presence of predation typically was found to increase decay for FIB, pathogenic bacteria, viruses, and bacteriophages, but the effects of predation were in some cases unclear (Korajkic et al., 2019; Ahmed et al.,

2021) and possible water type-predation interactions for FIB (Korajkic et al., 2013) and temperature-predation interactions for viral and protozoan targets were observed (Olive et al., 2020; Medema et al., 1997).

Fewer studies identified in the literature addressed the effects of water type and method of detection for the targets of concern. Decay differed by water type for some FIB with increasing salinities generally associated with faster decay (Fujioka et al., 1982; Korajkic et al., 2013; Liang et al., 2017; Jeanneau et al., 2012). Marine water was also associated with faster decay of viral targets (Ahmed et al., 2014; Nasser et al., 2003; Walters et al., 2009), which differs from conclusions in a previous review (Boehm et al., 2019). The FIB literature also suggested potential water type-method of detection interactions, as differences in FIB decay between culture-based and molecular-based methods were identified in the freshwater experiments but not the marine water (Korajkic et al., 2019). One study of pathogenic bacteria suggested decay may be faster in freshwater than marine (Levin-Edens et al., 2011), and the studies identified in this review did not have enough data to draw any conclusions about the effects of water type on decay. Turbidity was found to influence decay in the FIB and pathogenic bacteria literature, and pH was a potential factor of significance in the bacteriophage literature.

Notably, the majority of the studies focused on FIB and pathogenic bacteria and the factors affecting their persistence. Additional research manipulating water quality factors as they influence virus, bacteriophage, and protozoa decay are needed to better understand those relationships. The majority (75%) of the identified studies addressed persistence in freshwater matrices, and only 38% and 11% addressed persistence in marine and brackish matrices, respectively. In general, there was a lack of data for molecular-based methods. The lack of data on water type comparisons and persistence measured with molecular-based methods were

research gaps noted in a previous review (Boehm et al., 2018). It is important to note, however, that the qualitative synthesis of the studies discussed in this review reflect the water quality and environmental conditions assessed in experiments with shared in the peer-reviewed literature that also documented target concentrations for more than three independent time points.

## 2.5 Conclusions

This systematic review identified 61 studies that addressed the persistence of indicators or pathogens in natural surface water matrices, and these studies represent over 600 experiments in the literature that provided quantitative data to further explore persistence models that challenge traditional first-order decay kinetic assumptions. Strong relationships between sunlight, predation, temperature and persistence were consistently discussed for the majority of the target types. The effects of water type, method of detection, turbidity and pH were less consistent across target types and should be further explored in future analyses. This review highlighted several potential interactions between the water quality and experimental factors that could further complicate the relationships between factor and decay. Future experiments can be designed to test these potential interactions more thoroughly, and this review provides the foundation for the intentional inclusion of interaction terms in factor analyses, as opposed to exploratory analyses with forward or backwards selection, which is needed in the experimental design phase for such work. The methods frequently used to study persistence in surface waters (microcosms, mesocosms), models used to describe decay, and the techniques used to study factor relationships (correlation analyses, ANOVAs, regressions) were reviewed herein to aid future researchers in the endeavor to better understand pathogen threats and human health risks in surface waters.

Table 2.1: Studies Identified in the Literature Review and Relevant Characteristics

Source	Target(s)	Water Type	Predation	Sunlight	Method	Experimental Design*
Ahmed et al. 2014	FIB, Virus	Fresh, Marine	Present	Present	Culture-based, Molecular-based	Outdoor Microcosms
Ahmed et al. 2019	FIB	Fresh, Marine	Present	Present	Molecular-based	Outdoor Mesocosms
Ahmed et al. 2021	FIB, Bacteria, Virus, Protozoa	Fresh	Present/Absent	Present	Culture-based, Molecular-based	Laboratory Microcosms
Avery et al. 2008	Bacteria	Fresh	Present/Absent	Absent	Culture-based	Laboratory Microcosms
Badgley et al. 2010	FIB	Fresh	Present	Present	Culture-based	Outdoor Mesocosms
Bailey et al. 2019	FIB	Fresh	Present	Absent	Culture-based	Laboratory Experiments
Beckinghausen et al. 2014	FIB, Bacteria	Fresh	Absent	Present/Absent	Culture-based	Laboratory Microcosms
Bergstein et al. 1997	Virus	Fresh	Absent	Present/Absent	Culture-based	In-situ Experiments
Boehm et al. 2012	Bacteria	Fresh, Marine	Absent	Present/Absent	Culture-based	Laboratory Experiments
Booncharoen et al. 2018	Bacteriophage	Fresh, Marine	Present/Absent	Absent	Culture-based	Laboratory Microcosms
Bordalo et al. 2002	FIB	Brackish	Present	Present/Absent	Culture-based	Laboratory/Outdoor Microcosms
Budzinzka et al. 2012	Bacteria	Fresh	Present	Absent	Culture-based	Laboratory Experiments
Carratala et al. 2013	Virus	Marine	Present	Present/Absent	Culture-based	Laboratory Experiments
Chandran & Hatha 2005	FIB, Bacteria	Brackish	Present/Absent	Present/Absent	Culture-based	Laboratory/Outdoor Experiments
Chandran et al. 2011	FIB, Bacteria	Fresh	Present/Absent	Absent	Culture-based	Laboratory Microcosms
Czajkowska et al. 2005	Bacteria	Fresh	Present	Absent	Culture-based	Laboratory Experiments
de Oliveira et al. 2021	Virus	Fresh	Present/Absent	Absent	Culture-based	Laboratory Experiments
Dick et al. 2010	FIB	Fresh	Present/Absent	Present/Absent	Culture-based	Laboratory Microcosms

Table 2.1 (cont'd)

Easton et al. 2005	FIB, Bacteria	Fresh	Present	Present	Culture-based	Laboratory Experiments
El Mejri et al. 2012	Bacteria	Marine	Absent	Absent	Culture-based	Laboratory Microcosms
Espinosa et al. 2008	Virus	Fresh	Present	Present	Culture-based, Molecular-based	Laboratory Microcosms
Fujioka et al. 1982	FIB	Fresh, Marine	Present	Present/ Absent	Culture-based	Laboratory/ Outdoor Experiments
Greaves et al. 2021	FIB, Bacteriophage	Fresh	Present	Absent	Culture-based	Outdoor Mesocosms
Gronewald et al. 2011	FIB	Brackish	Present	Absent	Culture-based	Laboratory Experiments
Gutierrez-Cacciabue et al. 2016	FIB	Fresh	Present	Present/ Absent	Culture-based, Molecular-based	Outdoor Microcosms
Ibrahim et al. 2019	Bacteria, Virus	Fresh	Absent	Absent	Culture-based	Laboratory Experiments
Imamura et al. 2011	FIB	Marine	Present	Absent	Culture-based	Laboratory Microcosms
Irankhah et al. 2016	FIB	Marine	Present	Present	Culture-based	Laboratory Microcosms
Jeanneau et al. 2012	FIB	Fresh, Marine	Present	Absent	Culture-based	Laboratory Microcosms
Korajkic et al. 2013	FIB	Fresh, Marine	Present/ Absent	Present	Culture-based	Outdoor Mesocosms
Korajkic et al. 2014	FIB	Fresh	Present/ Absent	Present/ Absent	Culture-based, Molecular-based	Outdoor Mesocosms
Korajkic et al. 2019	FIB	Fresh, Marine	Present/Absent	Present/ Absent	Culture-based, Molecular-based	Outdoor Mesocosms
Lee & Sobsey 2011	Bacteriophage	Fresh	Present	Absent	Culture-based	Laboratory Experiments
Levin-Edens et al. 2011	Bacteria	Fresh, Marine	Absent	Absent	Culture-based	Laboratory Microcosms
Liang et al. 2012	FIB	Fresh	Present	Present	Culture-based	Outdoor Microcosms
Liang et al. 2017	FIB, Virus	Fresh, Marine	Present	Present/ Absent	Molecular-based	Laboratory Microcosms
Long & Sobsey 2004	Bacteriophages	Fresh	Present	Absent	Culture-based	Laboratory Microcosms

Table 2.1 (cont'd)

Lund et al. 1996	FIB, Bacteria	Fresh	Present/ Absent	Absent	Culture-based	Laboratory Experiments
Maracinni et al. 2016	FIB	Fresh, Marine, Brackish	Present	Present	Culture-based	In-situ Experiments
Mattioli et al. 2017	FIB	Marine	Present	Present	Culture-based, Molecular-based	In-situ Microcosms
Medema et al. 1997	FIB, Protozoa	Fresh	Present/ Absent	Absent	Culture-based	Laboratory Microcosms
Mezriouia et al. 1995	FIB and Bacteria	Brackish	Present/ Absent	Absent	Culture-based	Laboratory Microcosms
Nasser et al. 2003	FIB, Virus, Protozoa	Fresh, Marine, Brackish	Present	Absent	Culture-based	Laboratory Experiments
Noble et al. 2004	FIB, Bacteriophage	Marine	Present	Present/ Absent	Culture-based	Laboratory/Outdoor Experiments
Okabe et al. 2007	FIB	Fresh, Marine	Present	Absent	Culture-based	Laboratory Experiments
Olive et al. 2020	Bacteriophage, Virus	Fresh, Marine	Present/ Absent	Absent	Culture-based	Laboratory Experiments
Ravva et al. 2016	Bacteriophage, FIB	Fresh	Present	Absent	Culture-based	Laboratory Experiments
Robertson et al. 2006	Protozoa	Fresh	Present	Present/ Absent	Culture-based	Laboratory/In-situ Experiments
Rodriguez et al. 2012	Bacteria	Fresh	Present	Present/ Absent	Culture-based	Laboratory Microcosms/ In-situ Experiments
Silvester et al. 2021	Bacteria	Brackish	Present/ Absent	Present/ Absent	Culture-based	Laboratory Microcosms
Solecki et al. 2011	FIB	Fresh, Marine	Present	Absent	Culture-based	Laboratory Microcosms
Terzieva & McFeters 1991	FIB, Bacteria	Fresh	Present	Absent	Culture-based	Laboratory Experiments
Tiwari et al. 2019	FIB, Bacteria, Bacteriophage	Brackish	Present	Present	Culture-based, Molecular-based	Laboratory Mesocosms
Topalcengiz & Danyluk 2019	FIB, Bacteria	Fresh	Present/ Absent	Absent	Culture-based	Laboratory Experiments
Walters et al. 2009	FIB, Virus	Marine	Present	Present/ Absent	Culture-based, Molecular-based	Outdoor Microcosms
Walters & Fields 2009	FIB	Fresh	Present	Present/ Absent	Culture-based	Outdoor Microcosms



Table 2.1 (cont'd)

Wang & Doyle 1998	Bacteria	Fresh	Present/ Absent	Absent	Culture-based	Laboratory Experiments
Wanjugi et al. 2016	FIB	Fresh	Present/ Absent	Present	Culture-based	Outdoor Microcosms
Yang & Griffiths 2013	Bacteriophages	Fresh	Present	Absent	Culture-based	Laboratory Experiments
Yukselen et al. 2003	FIB	Marine	Present	Present/ Absent	Culture-based	Laboratory/ Outdoor Experiments
Zhang et al. 2015	FIB	Marine	Present	Absent	Culture-based	Laboratory Microcosms

\*If the researchers of each study did not use the terms microcosms or mesocosms in their methods, the experimental design is designated as experiments herein.

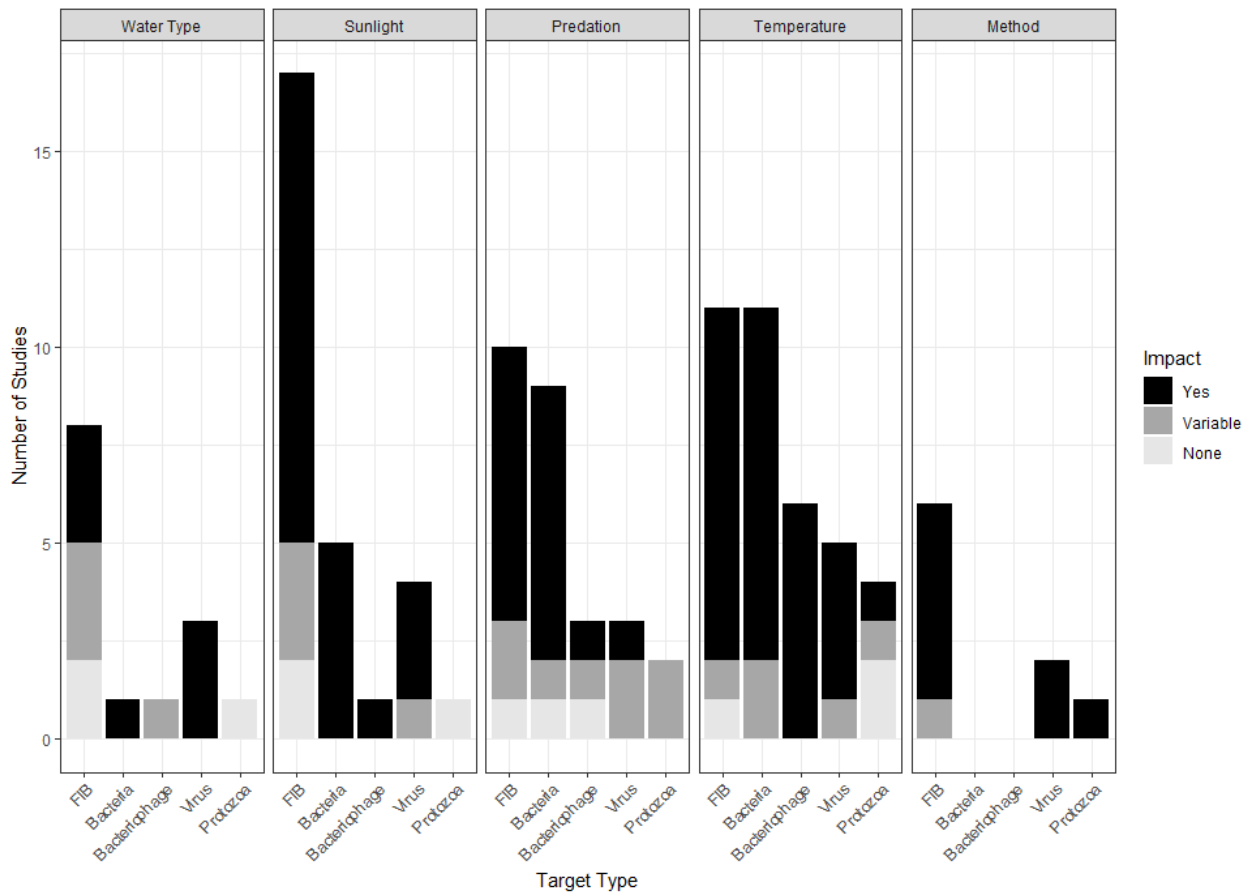


Figure 2.1: Number of studies identified in the review that made conclusions about a factor's influence on target persistence based on three tiers of impact: 1) **Yes**- the study identified a significant impact from the variable, 2) **Variable**- the impact of the factor on persistence was not consistent, or varied based on other factors, and 3) **None**- the factor was addressed in the study but no significant impact was identified.

## REFERENCES

- Ahmed, W., Gyawali, P., Sidhu, J. P. S., & Toze, S. (2014). Relative inactivation of faecal indicator bacteria and sewage markers in freshwater and seawater microcosms. *Letters in Applied Microbiology*, 59(3), 348–354. <https://doi.org/10.1111/lam.12285>
- Ahmed, W., Zhang, Q., Kozak, S., Beale, D., Gyawali, P., Sadowsky, M. J., & Simpson, S. (2019). Comparative decay of sewage-associated marker genes in beach water and sediment in a subtropical region. *Water Research*, 149, 511–521. <https://doi.org/10.1016/j.watres.2018.10.088>
- Ahmed, W., Toze, S., Veal, C., Fisher, P., Zhang, Q., Zhu, Z., Staley, C., & Sadowsky, M. J. (2021). Comparative decay of culturable faecal indicator bacteria, microbial source tracking marker genes, and enteric pathogens in laboratory microcosms that mimic a sub-tropical environment. *Science of The Total Environment*, 751, 141475. <https://doi.org/10.1016/j.scitotenv.2020.141475>
- Avery, L. M., Williams, A. P., Killham, K., & Jones, D. L. (2008). Survival of Escherichia coli O157:H7 in waters from lakes, rivers, puddles and animal-drinking troughs. *Science of The Total Environment*, 389(2–3), 378–385. <https://doi.org/10.1016/j.scitotenv.2007.08.049>
- Badgley, B. D., Thomas, F. I. M., & Harwood, V. J. (2010). The effects of submerged aquatic vegetation on the persistence of environmental populations of Enterococcus spp.: Enterococcus persistence in aquatic reservoirs. *Environmental Microbiology*, 12(5), 1271–1281. <https://doi.org/10.1111/j.1462-2920.2010.02169.x>
- Bailey, E. S., Casanova, L. M., & Sobsey, M. D. (2019). Effects of environmental storage conditions on survival of indicator organisms in a blend of surface water and dual disinfected reclaimed water. *Journal of Applied Microbiology*, 126(3), 985–994. <https://doi.org/10.1111/jam.14186>
- Beckinghausen, A., Martinez, A., Blersch, D., & Haznedaroglu, B. Z. (2014). Association of nuisance filamentous algae Cladophora spp. with E. coli and Salmonella in public beach waters: Impacts of UV protection on bacterial survival. *Environ. Sci.: Processes Impacts*, 16(6), 1267–1274. <https://doi.org/10.1039/C3EM00659J>
- Benham, B.L., Baggaut, C., Zeckoski, R.W., Mankin, K.R., Pachepsky, Y.A., Sadeghi, A.M., Brannan, K.M., Soupir, M.L., & Habersack, M.J. 2006. Modeling bacteria fate and transport in watersheds to support TMDLs. *Transactions of the ASABE*, 49(4): 987-1002. doi: 10.13031/2013.21739
- Bergstein-Ben Dan, T., Wynne, D., & Manor, Y. (1997). Survival of enteric bacteria and viruses in Lake Kinneret, Israel. *Water Research*, 31(11), 2755–2760. [https://doi.org/10.1016/S0043-1354\(97\)00135-8](https://doi.org/10.1016/S0043-1354(97)00135-8)
- Blaustein, R. A., Pachepsky, Y., Hill, R. L., Shelton, D. R., & Whelan, G. (2013). Escherichia coli survival in waters: Temperature dependence. *Water Research*, 47(2), 569–578. <https://doi.org/10.1016/j.watres.2012.10.027>

- Boehm, A. B., Soetjipto, C., & Wang, D. (2012). Solar inactivation of four Salmonella serovars in fresh and marine waters. *Journal of Water and Health*, 10(4), 504–510. <https://doi.org/10.2166/wh.2012.084>
- Boehm, A. B., Graham, K. E., & Jennings, W. C. (2018). Can We Swim Yet? Systematic Review, Meta-Analysis, and Risk Assessment of Aging Sewage in Surface Waters. *Environmental Science & Technology*, 52(17), 9634–9645. <https://doi.org/10.1021/acs.est.8b01948>
- Boehm, A. B., Silverman, A. I., Schriewer, A., & Goodwin, K. (2019). Systematic review and meta-analysis of decay rates of waterborne mammalian viruses and coliphages in surface waters. *Water Research*, 164, 114898. <https://doi.org/10.1016/j.watres.2019.114898>
- Booncharoen, N., Mongkolsuk, S., & Sirikanchana, K. (2018). Comparative persistence of human sewage-specific enterococcal bacteriophages in freshwater and seawater. *Applied Microbiology and Biotechnology*, 102(14), 6235–6246. <https://doi.org/10.1007/s00253-018-9079-1>
- Bordalo, A. A., Onrassami, R., & Dechsakulwatana, C. (2002). Survival of faecal indicator bacteria in tropical estuarine waters (Bangpakong River, Thailand). *Journal of Applied Microbiology*, 93(5), 864–871. <https://doi.org/10.1046/j.1365-2672.2002.01760.x>
- Brouwer, A. F., Eisenberg, M. C., Remais, J. V., Collender, P. A., Meza, R., & Eisenberg, J. N. S. (2017). Modeling Biphasic Environmental Decay of Pathogens and Implications for Risk Analysis. *Environmental Science & Technology*, 51(4), 2186–2196. <https://doi.org/10.1021/acs.est.6b04030>
- Budzińska, K., Wroński, G., & Szejniuk, B. (n.d.). *Survival Time of Bacteria Listeria monocytogenes in Water Environment and Sewage*. 8.
- Carratalà, A., Rusiñol, M., Rodriguez-Manzano, J., Guerrero-Latorre, L., Sommer, R., & Girones, R. (2013). Environmental Effectors on the Inactivation of Human Adenoviruses in Water. *Food and Environmental Virology*, 5(4), 203–214. <https://doi.org/10.1007/s12560-013-9123-3>
- Chandran, A., & Mohamed Hatha, A. A. (2005). Relative survival of Escherichia coli and Salmonella typhimurium in a tropical estuary. *Water Research*, 39(7), 1397–1403. <https://doi.org/10.1016/j.watres.2005.01.010>
- Chandran, A., Varghese, S., Kandeler, E., Thomas, A., Hatha, M., & Mazumder, A. (2011). An assessment of potential public health risk associated with the extended survival of indicator and pathogenic bacteria in freshwater lake sediments. *International Journal of Hygiene and Environmental Health*, 214(3), 258–264. <https://doi.org/10.1016/j.ijheh.2011.01.002>
- Crane, S.R. & Moore, J.A. 1986. Modeling Enteric Bacterial Die-Off: A Review. *Water, Air, and Soil Pollution*, 27: 411-439.

- Czajkowska, D., Witkowska-Gwiazdowska, A., Sikorska, I., Boszczyk-Maleszak, H., & Horoch, M. (n.d.). *Survival of Escherichia Coli Serotype O157:H7 in Water and in Bottom-Shore Sediments*. 9.
- Dean, K., Wissler, A., Hernandez-Suarez, J.S., Nejadhashemi, A.P., & Mitchell, J. 2020. Modeling the Persistence of Viruses in Untreated Groundwater. *Science of the Total Environment*, 717(15). <https://doi.org/10.1016/j.scitotenv.2019.134599>
- de Brauwere, A., Ouattara, N. K., & Servais, P. (2014). Modeling Fecal Indicator Bacteria Concentrations in Natural Surface Waters: A Review. *Critical Reviews in Environmental Science and Technology*, 44(21), 2380–2453. <https://doi.org/10.1080/10643389.2013.829978>
- de Oliveira, L. C., Torres-Franco, A. F., Lopes, B. C., Santos, B. S. Á. da S., Costa, E. A., Costa, M. S., Reis, M. T. P., Melo, M. C., Polizzi, R. B., Teixeira, M. M., & Mota, C. R. (2021). Viability of SARS-CoV-2 in river water and wastewater at different temperatures and solids content. *Water Research*, 195, 117002. <https://doi.org/10.1016/j.watres.2021.117002>
- Dick, L. K., Stelzer, E. A., Bertke, E. E., Fong, D. L., & Stoeckel, D. M. (2010). Relative Decay of *Bacteroidales* Microbial Source Tracking Markers and Cultivated *Escherichia coli* in Freshwater Microcosms. *Applied and Environmental Microbiology*, 76(10), 3255–3262. <https://doi.org/10.1128/AEM.02636-09>
- Easton, J. H., Gauthier, J. J., Lalor, M. M., & Pitt, R. E. (2005). Die-off of pathogenic *E. coli* O157:H7 in sewage contaminated waters. *Journal of the American Water Resources Association*, 41(5), 1187–1193. <https://doi.org/10.1111/j.1752-1688.2005.tb03793.x>
- El Mejri, S., El Bour, M., Boukef, I., Al Gallas, N., Mraouna, R., Got, P., Troussellier, M., Klena, J., & Boudabbous, A. (2012). INFLUENCE OF MARINE WATER CONDITIONS ON SALMONELLA ENTERICA SEROVAR TYPHIMURIUM SURVIVAL: SALMONELLA SPP. UNDER STRESS CONDITIONS. *Journal of Food Safety*, 32(3), 270–278. <https://doi.org/10.1111/j.1745-4565.2012.00377.x>
- Espinosa, A. C., Mazari-Hiriart, M., Espinosa, R., Maruri-Avidal, L., Méndez, E., & Arias, C. F. (2008). Infectivity and genome persistence of rotavirus and astrovirus in groundwater and surface water. *Water Research*, 42(10–11), 2618–2628. <https://doi.org/10.1016/j.watres.2008.01.018>
- Fujioka, R. S. (1981). Effect of Sunlight on Survival of Indicator Bacteria in Seawater. *APPL. ENVIRON. MICROBIOL.*, 41, 7.
- Greaves, J., Stone, D., Wu, Z., & Bibby, K. (2020). Persistence of emerging viral fecal indicators in large-scale freshwater mesocosms. *Water Research X*, 9, 100067. <https://doi.org/10.1016/j.wroa.2020.100067>
- Gronewold, A. D., Myers, L., Swall, J. L., & Noble, R. T. (2011). Addressing uncertainty in fecal indicator bacteria dark inactivation rates. *Water Research*, 45(2), 652–664. <https://doi.org/10.1016/j.watres.2010.08.029>

- Gutiérrez-Cacciabue, D., Cid, A. G., & Rajal, V. B. (2016). How long can culturable bacteria and total DNA persist in environmental waters? The role of sunlight and solid particles. *Science of The Total Environment*, 539, 494–502. <https://doi.org/10.1016/j.scitotenv.2015.07.138>
- Haas, C.N., Rose, J.B., Gerba, C.P. (2014). *Quantitative Microbial Risk Assessment*. 2<sup>nd</sup> ed. New York: Wiley.
- Hellweger, F. L., & Masopust, P. (2008). Investigating the Fate and Transport of *Escherichia coli* in the Charles River, Boston, Using High-Resolution Observation and Modeling <sup>1</sup>. *JAWRA Journal of the American Water Resources Association*, 44(2), 509–522. <https://doi.org/10.1111/j.1752-1688.2008.00179.x>
- Ibrahim, E. M. E., El-Liethy, M. A., Abia, A. L. K., Hemdan, B. A., & Shaheen, M. N. (2019). Survival of E. coli O157:H7, Salmonella Typhimurium, HAdV2 and MNV-1 in river water under dark conditions and varying storage temperatures. *Science of The Total Environment*, 648, 1297–1304. <https://doi.org/10.1016/j.scitotenv.2018.08.275>
- Imamura, G. J., Thompson, R. S., Boehm, A. B., & Jay, J. A. (2011). Wrack promotes the persistence of fecal indicator bacteria in marine sands and seawater: Beach wrack: FIB reservoir. *FEMS Microbiology Ecology*, 77(1), 40–49. <https://doi.org/10.1111/j.1574-6941.2011.01082.x>
- Frankhah, S., Soudi, M. R., & Gharavi, S. (n.d.). *Ex situ study of Enterococcus faecalis survival in the recreational waters of the southern coast of the Caspian Sea*. 7.
- Jeanneau, L., Solecki, O., Wéry, N., Jardé, E., Gourmelon, M., Communal, P.-Y., Jadas-Hécart, A., Caprais, M.-P., Gruau, G., & Pourcher, A.-M. (2012). Relative Decay of Fecal Indicator Bacteria and Human-Associated Markers: A Microcosm Study Simulating Wastewater Input into Seawater and Freshwater. *Environmental Science & Technology*, 46(4), 2375–2382. <https://doi.org/10.1021/es203019y>
- Korajkic, A., McMinn, B. R., Ashbolt, N. J., Sivaganesan, M., Harwood, V. J., & Shanks, O. C. (2019). Extended persistence of general and cattle-associated fecal indicators in marine and freshwater environment. *Science of The Total Environment*, 650, 1292–1302. <https://doi.org/10.1016/j.scitotenv.2018.09.108>
- Korajkic, A., McMinn, B. R., Shanks, O. C., Sivaganesan, M., Fout, G. S., & Ashbolt, N. J. (2014). Biotic Interactions and Sunlight Affect Persistence of Fecal Indicator Bacteria and Microbial Source Tracking Genetic Markers in the Upper Mississippi River. *Applied and Environmental Microbiology*, 80(13), 3952–3961. <https://doi.org/10.1128/AEM.00388-14>
- Korajkic, A., Wanjugi, P., & Harwood, V. J. (2013). Indigenous Microbiota and Habitat Influence *Escherichia coli* Survival More than Sunlight in Simulated Aquatic Environments. *Applied and Environmental Microbiology*, 79(17), 5329–5337. <https://doi.org/10.1128/AEM.01362-13>

- Lee, H. S., & Sobsey, M. D. (2011). Survival of prototype strains of somatic coliphage families in environmental waters and when exposed to UV low-pressure monochromatic radiation or heat. *Water Research*, 45(12), 3723–3734. <https://doi.org/10.1016/j.watres.2011.04.024>
- Levin-Edens, E., Bonilla, N., Meschke, J. S., & Roberts, M. C. (2011). Survival of environmental and clinical strains of methicillin-resistant *Staphylococcus aureus* [MRSA] in marine and fresh waters. *Water Research*, 45(17), 5681–5686. <https://doi.org/10.1016/j.watres.2011.08.037>
- Liang, Z., He, Z., Zhou, X., Powell, C. A., Yang, Y., Roberts, M. G., & Stoffella, P. J. (2012). High diversity and differential persistence of fecal Bacteroidales population spiked into freshwater microcosm. *Water Research*, 46(1), 247–257. <https://doi.org/10.1016/j.watres.2011.11.004>
- Liang, L., Goh, S. G., & Gin, K. Y. H. (2017). Decay kinetics of microbial source tracking (MST) markers and human adenovirus under the effects of sunlight and salinity. *Science of The Total Environment*, 574, 165–175. <https://doi.org/10.1016/j.scitotenv.2016.09.031>
- Long, S. C., & Sobsey, M. D. (2004). A comparison of the survival of F+RNA and F+DNA coliphages in lake water microcosms. *Journal of Water and Health*, 2(1), 15–22. <https://doi.org/10.2166/wh.2004.0002>
- Lund, Vidar. (1996). Evaluation of *E. coli* as an indicator for the presence of *Campylobacter jejuni* and *Yersinia enterocolitica* in chlorinated and untreated oligotrophic lake water. *Water Research*, 30(6), 1528–1534.
- Maraccini, P. A., Mattioli, M. C. M., Sassoubre, L. M., Cao, Y., Griffith, J. F., Ervin, J. S., Van De Werfhorst, L. C., & Boehm, A. B. (2016). Solar Inactivation of Enterococci and *Escherichia coli* in Natural Waters: Effects of Water Absorbance and Depth. *Environmental Science & Technology*, 50(10), 5068–5076. <https://doi.org/10.1021/acs.est.6b00505>
- Mattioli, M. C., Sassoubre, L. M., Russell, T. L., & Boehm, A. B. (2017). Decay of sewage-sourced microbial source tracking markers and fecal indicator bacteria in marine waters. *Water Research*, 108, 106–114. <https://doi.org/10.1016/j.watres.2016.10.066>
- Medema, G.J., Bahara, M., & Schets, F.M. (1997). Survival of *Cryptosporidium parvum*, *Escherichia coli*, faecal enterococci, and *Clostridium perfringens* in river water: influence of temperature and autochthonous microorganisms. *Water Science and Technology*, 35(11/12):249-252. [https://doi.org/10.1016/S0273-1223\(97\)00267-9](https://doi.org/10.1016/S0273-1223(97)00267-9)
- Mezrioui, N., Baleux, B., & Troussellier, M. (1995). A microcosm study of the survival of *Escherichia coli* and *Salmonella typhimurium* in brackish water. *Water Research*, 29(2), 459–465. [https://doi.org/10.1016/0043-1354\(94\)00188-D](https://doi.org/10.1016/0043-1354(94)00188-D)
- Mitchell, J. & Akram, S. 2017. “Pathogen Specific Persistence Modeling Data.” In: J.B. Rose and B. Jiménez-Cisneros, (eds) *Global Water Pathogens Project*, <http://www.waterpathogens.org> (M. Yates (eds) Part 4 Management of Risk from Excreta

and Wastewater) Accessible at: <http://www.waterpathogens.org/book/pathogen-specific-persistence-modeling-data>. Michigan State University, E. Lansing, MI, UNESCO.

- Nasser, A. M., Zaruk, N., Tenenbaum, L., & Netzan, Y. (2003). Comparative survival of Cryptosporidium, coxsackievirus A9 and Escherichia coli in stream, brackish and sea waters. *Water Science and Technology*, 47(3), 91–96. <https://doi.org/10.2166/wst.2003.0170>
- Noble, R. T., Lee, I. M., & Schiff, K. C. (2004). Inactivation of indicator micro-organisms from various sources of faecal contamination in seawater and freshwater. *Journal of Applied Microbiology*, 96(3), 464–472. <https://doi.org/10.1111/j.1365-2672.2004.02155.x>
- Okabe, S., & Shimazu, Y. (2007). Persistence of host-specific Bacteroides–Prevotella 16S rRNA genetic markers in environmental waters: Effects of temperature and salinity. *Applied Microbiology and Biotechnology*, 76(4), 935–944. <https://doi.org/10.1007/s00253-007-1048-z>
- Olive, M., Gan, C., Carratalà, A., & Kohn, T. (2020). Control of Waterborne Human Viruses by Indigenous Bacteria and Protists Is Influenced by Temperature, Virus Type, and Microbial Species. *Applied and Environmental Microbiology*, 86(3). <https://doi.org/10.1128/AEM.01992-19>
- Pachepsky, Y. A., Sadeghi, A. M., Bradford, S. A., Shelton, D. R., Guber, A. K., & Dao, T. (2006). Transport and fate of manure-borne pathogens: Modeling perspective. *Agricultural Water Management*, 86(1–2), 81–92. <https://doi.org/10.1016/j.agwat.2006.06.010>
- Park, Y., Pachepsky, Y., Shelton, D., Jeong, J., & Whelan, G. (2016). Survival of Manure-borne *Escherichia coli* and Fecal Coliforms in Soil: Temperature Dependence as Affected by Site-Specific Factors. *Journal of Environmental Quality*, 45(3), 949–957. <https://doi.org/10.2134/jeq2015.08.0427>
- Ravva, S. V., & Sarreal, C. Z. (2016). Persistence of F-Specific RNA Coliphages in Surface Waters from a Produce Production Region along the Central Coast of California. *PLOS ONE*, 11(1), e0146623. <https://doi.org/10.1371/journal.pone.0146623>
- Robertson, L. J., & Gjerde, B. K. (2006). Fate of Cryptosporidium Oocysts and Giardia Cysts in the Norwegian Aquatic Environment over Winter. *Microbial Ecology*, 52(4), 597–602. <https://doi.org/10.1007/s00248-006-9005-4>
- Rodríguez, S., & Araujo, R. (2012). Effect of environmental parameters on the inactivation of the waterborne pathogen Campylobacter in a Mediterranean river. *Journal of Water and Health*, 10(1), 100–107. <https://doi.org/10.2166/wh.2011.044>
- Silvester, R., Antony, A. C., Yousuf, J., Madhavan, A., Sooria, P. M., Kokkat, A., Harikrishnan, M., & Hatha, M. (2021). Survival Kinetics of Vibrio Species in a Tropical Estuary along the Southwest coast of India—As a function of selected Environmental Factors. *Fishery Technology*, 58: 40-47.

- Solecki, O., Jeanneau, L., Jardé, E., Gourmelon, M., Marin, C., & Pourcher, A. M. (2011). Persistence of microbial and chemical pig manure markers as compared to faecal indicator bacteria survival in freshwater and seawater microcosms. *Water Research*, 45(15), 4623–4633. <https://doi.org/10.1016/j.watres.2011.06.012>
- Terzieva, S. I., & McFeters, G. A. (1991). Survival and injury of *Escherichia coli*, *Campylobacter jejuni*, and *Yersinia enterocolitica* in stream water. *Canadian Journal of Microbiology*, 37(10), 785–790. <https://doi.org/10.1139/m91-135>
- Tiwari, A., Kauppinen, A., & Pitkänen, T. (2019). Decay of *Enterococcus faecalis*, *Vibrio cholerae* and MS2 Coliphage in a Laboratory Mesocosm Under Brackish Beach Conditions. *Frontiers in Public Health*, 7, 269. <https://doi.org/10.3389/fpubh.2019.00269>
- Topalcengiz, Z., & Danyluk, M. D. (2019). Fate of generic and Shiga toxin-producing *Escherichia coli* (STEC) in Central Florida surface waters and evaluation of EPA Worst Case water as standard medium. *Food Research International*, 120, 322–329. <https://doi.org/10.1016/j.foodres.2019.02.045>
- Walters, S. P., Yamahara, K. M., & Boehm, A. B. (2009). Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: Implications for their use in assessing risk in recreational waters. *Water Research*, 43(19), 4929–4939. <https://doi.org/10.1016/j.watres.2009.05.047>
- Walters, S. P., & Field, K. G. (2009). Survival and persistence of human and ruminant-specific faecal *Bacteroidales* in freshwater microcosms. *Environmental Microbiology*, 11(6), 1410–1421. <https://doi.org/10.1111/j.1462-2920.2009.01868.x>
- Wang, G., & Doyle, M. P. (1998). Survival of Enterohemorrhagic *Escherichia coli* 0157:H7 in Water. *J. Food Prot.*, 61(6): 662-667.
- Wanjugi, P., Fox, G. A., & Harwood, V. J. (2016). The Interplay Between Predation, Competition, and Nutrient Levels Influences the Survival of *Escherichia coli* in Aquatic Environments. *Microbial Ecology*, 72(3), 526–537. <https://doi.org/10.1007/s00248-016-0825-6>
- Yang, Y., & Griffiths, M. W. (2013). Comparative Persistence of Subgroups of F-Specific RNA Phages in River Water. *Applied and Environmental Microbiology*, 79(15), 4564–4567. <https://doi.org/10.1128/AEM.00612-13>
- Yukselen, M. A., Calli, B., Gokyay, O., & Saatci, A. (2003). Inactivation of coliform bacteria in Black Sea waters due to solar radiation. *Environment International*, 29(1), 45–50. [https://doi.org/10.1016/S0160-4120\(02\)00144-7](https://doi.org/10.1016/S0160-4120(02)00144-7)
- Zhang, Q., He, X., & Yan, T. (2015). Differential Decay of Wastewater Bacteria and Change of Microbial Communities in Beach Sand and Seawater Microcosms. *Environmental Science & Technology*, 49(14), 8531–8540. <https://doi.org/10.1021/acs.est.5b01879>



## APPENDIX

Web of Science was used during the first two weeks of July 2020 with the following keyword combinations: (survival OR persistence OR decay) AND (indicators OR pathogens OR pathogenic bacteria OR pathogenic viruses OR sewage OR sewage-associated) AND (surface waters OR environmental waters). Note, the term “NOT” was used to exclude the sources from prior searches. The titles and abstracts were scanned for each source, and if the study seemed relevant to the aforementioned criteria, it was selected for further evaluation. The review was updated using the same methodology in the last week of June 2021 to ensure the inclusion of any recently published relevant studies. In addition to the 20 resources carried over from prior analyses within the Global Water Pathogens Project, 3,929 sources were identified through the database searching. Two hundred and twenty-six unique resources were scanned for relevancy based off title and abstract, and 115 studies were eligible for further assessment. After assessing suitability of data and the inclusion of necessary details, 61 total studies were selected for inclusion in this review and the following analysis. An illustration of the PRISM Literature Review Process for this study is shown in Figure A2.1.

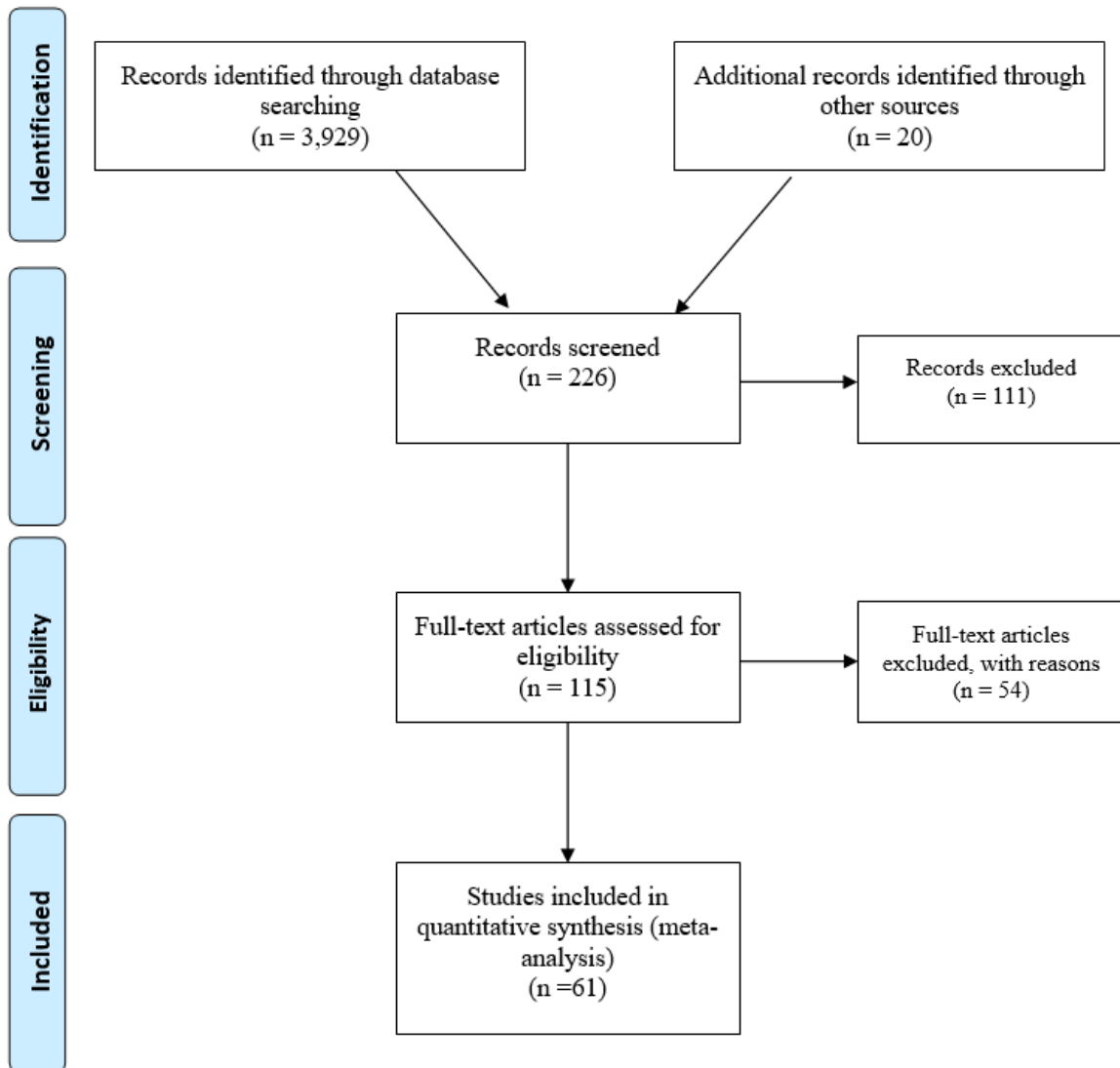


Figure A2.1: PRISMA 2009 Flow Diagram to Illustrate the Systematic Literature Review Process (Moher et al., 2009)

## CHAPTER 3: META-ANALYSIS ADDRESSING THE IMPLICATIONS OF MODEL UNCERTAINTY IN UNDERSTANDING THE PERSISTENCE OF INDICATORS AND PATHOGENS IN NATURAL SURFACE WATERS

This chapter has been published in *Environmental Science & Technology* and is reprinted with permission from *Environ. Sci. Technol.* 2022, 56, 17, 12106–12115. Copyright 2022 American Chemical Society.

### 3.1 Introduction

Persistence modeling in the context of microbial surface water contamination has predominantly relied on first order decay kinetic assumptions (Crane & Moore, 1986; de Brauwere et al., 2014; Pachepsky et al., 2006). Relying on general indicator organism first-order decay rates instead of persistence models capturing site-specific or pathogen-specific decay behaviors may cause misleading predictions about pathogen loads and associated risks, which is of particular concern for water bodies that maintain water quality with assumed natural attenuation. Despite its consistent use and application, the simplicity of applying first-order decay kinetics has long been challenged in the literature (Benham et al., 2006; Blaustein et al., 2013; Easton et al., 1999; Gonzalez, 1995; Pachepsky et al., 2006; Park et al., 2016). Biphasic decay, or two-stage decay kinetics, has been frequently observed in natural and bench-scale experiments (Easton et al., 1999, 2005; Medema et al., 1997; Mitchell & Akram, 2017; Park et al., 2016), for both culture-based and molecular-based fecal indicator bacteria (FIB) targets (Korajkic et al., 2014). Approaches for addressing the observed nonlinear decay patterns have included fitting a log-linear model to two portions of the data, yielding two first-order decay rates, or fitting a delayed Chick Watson model (Easton et al., 2005; Green et al., 2011; Sivaganesan et al., 2003). Some researchers have also explored the application of alternative persistence model forms. For example, Gonzalez et al. (1995) fit the log-linear, Gompertz, and logistic models to survival data of enteric bacteria in fresh and marine waters and found that the

nonlinear models provided a significantly better fit to the data and supplied added information about the lag times, decay rates, and asymptotes of the survival curves (Gonzalez, 1995).

More recently, a suite of two and three parameter models identified primarily in the food microbiology literature have been used to fit survival datasets of bacteria, viruses, indicators, and MST markers in a variety of water matrices (Y. Brooks et al., 2015; Dean et al., 2020; Mitchell & Akram, 2017). Some of the models found to best fit the data most frequently in these previous studies include the exponential damped, the Juneja and Marks 1, the Juneja and Marks 2, and the double exponential models (Dean et al., 2020; Mitchell & Akram, 2017). These four models have two or three-parameters and the model forms can capture two-stage decay kinetics, initial periods of minimal decay (shoulders), and decay rates that taper off over time (tails). Of these four alternative model forms, the Juneja and Marks 1 (jm1) and Juneja and Marks 2 (jm2) models are considered to be more mechanistically motivated and have most frequently been identified as the best fitting models to multiple datasets addressing the persistence of indicators, viruses, and bacteria in water matrices and fomites in previous works (Dean et al., 2020; Enger et al., 2018; Mitchell & Akram, 2017; Tamrakar et al., 2017).

As surface waters encompass a range of water types with varying water quality conditions and environmental stressors, it was posited that the persistence behaviors of pathogens and indicator organisms in surface waters would be more dynamic than the traditionally assumed first-order decay. Based on previous analyses of indicator and pathogen persistence in various water matrices, it was more specifically hypothesized that a two-parameter model, particularly **jm1** or **jm2**, would provide a better fit than the conventional exponential model (Dean et al., 2020; Mitchell & Akram, 2017). It was further hypothesized that the identification and use of an improved model form to describe pathogen and indicator persistence

behaviors will yield more accurate predictions of persistence values of interest, and exploratory factor analyses using these values will lead to novel insights about the factors affecting persistence in natural surface waters.

To test these hypotheses, a previously published systematic literature review was used to generate a database of natural surface water persistence experiments (Dean & Mitchell, 2022b) to facilitate the fitting of alternative persistence models and the exploration of factor-persistence relationships. Although meta-analyses assessing the persistence of multiple targets in surface waters have been completed before (Boehm et al., 2018, 2019; Mitchell & Akram, 2017), this meta-analysis is unique because it is (i) the most comprehensive to date providing a larger weight of evidence for the findings; (ii) supported by novel quantitative analysis including the fitting of alternative persistence models to each mined dataset to address both data and model uncertainty; and (iii) able to refine relationships and elucidate interactions between indicator and pathogen persistence and the natural environmental conditions present in surface waters through rigorous data analytic techniques. The insights garnered by this analysis can inform quantitative microbial risk assessments (QMRA), which are used to establish acceptable levels of waterborne pathogens and set water monitoring protocols (Haas et al., 2014), and suggest enhancements to future experimental designs that aim to explore pathogen decay dynamics.

## 3.2 Methods

### 3.2.1 Data Mining

Data for this study were extracted from 61 studies previously identified in a systematic literature review that evaluated the state of the science with regards to the persistence of FIB, bacteria, viruses, bacteriophages, or protozoa over time in natural surface waters (Dean & Mitchell, 2022b). The five target groups were selected to help highlight potential differences in

persistence behaviors for targets used as indicator organisms (FIB and bacteriophages) and targets capable of causing waterborne disease (bacteria, viruses, and protozoa). Water type, sunlight presence, predation presence, method of detection, and water temperature were documented for each experiment, in addition to any other relevant water quality factors (Dean & Mitchell, 2022b). Each experiment needed to include four or more time points and the associated concentrations to be included in the meta-analysis. Available data were then extracted from the studies or data shown graphically were digitized using the open access Web Plot Digitizer tool (<https://automeris.io/WebPlotDigitizer/>). Concentration data were transformed into log reduction values ( $\log_{10}(N_t/N_0)$ ) to facilitate the fitting of the persistence models shown in Table 3.1 and for ease of interpretation.

### 3.2.2 Persistence Model Fitting

Each dataset was first assessed for a general negative trend by fitting a basic linear model to each dataset with the log-reduction values as the dependent variable, time as the independent variable, and a set intercept of zero. If the slope of the linear model was not a statistically significant ( $p < 0.05$ ) negative value, the dataset was not included in the analysis. A selection of models (Table 3.1) previously determined to be well-suited to fit persistence data were fit to the datasets with Maximum Likelihood Estimation methods in R using the basic **optim** function (R Core Team, 2020). For each of the models shown in Table 3.1,  $N_t$  is the concentration at time  $t$  and  $N_0$  is the concentration at time zero. Both sides of each equation have been  $\log_{10}$ -transformed to facilitate the plotting of log reduction values over time. Briefly, the exponential model (**ep**), also referred to as the log-linear model or Chick-Watson model, assumes a constant rate of decay ( $k_t$ ) over time, as observed in first-order chemical reactions (Chick, 1908; Watson, 1908). The exponential damped model (**epd**), also referred to as the exponentially damped polynomial

model, reduces to the exponential when  $k_2$  is zero and returns to the initial level (log reduction=0) when both  $k_1$  and  $k_2$  are positive (McKellar & Lue, 2004; Whiting, & Buchanan, Robert, L., 2001). Although this model lends itself to capturing a tailing effect and has been one of the best fitting models in our previous work (Dean et al., 2020; Mitchell & Akram, 2017; Tamrakar et al., 2017), the U-shape tendency has the potential to limit its applicability to microbial persistence data (Enger et al., 2018). The **jm1** model, more commonly known as the multi-hit target theory model, is a two-parameter model that reduces to the exponential when  $k_2=1$ . It was originally derived in the field of radiation biology, and its premise is that the death of an organism occurs from the inactivation of multiple critical sites on the cell (Atwood & Norman, 1949; Juneja & Marks, 2001; Little, 1968; Nomiya, 2013). This behavior is reflected in the commonly seen “shoulders”, where it is hypothesized that some damage is absorbed by the cell before death occurs (Little, 1968). The **jm2** function has been used to model *Salmonella* and *Listeria monocytogenes* survival in response to various treatments (Carlier et al., 1996; Juneja et al., 2003; Juneja & Marks, 2003). The sigmoidal function of **jm2** is based on the logistic probability distribution and  $k_1$  and  $k_2$  can be considered as dispersal and location ( $k_2>0$ ) parameters, respectively (Juneja & Marks, 2003). Finally, the double exponential model (**dep**) was proposed by Shull et al. (1963) and analyzed by Abraham et al. (1990) by applying the model to *Bacillus stearothermophilus* spore inactivation data. The model is based on the premise that when a population of spores is first exposed to heat, there is a population of activated spores and inactivated spores. The  $k_3$  parameter, as shown in Table 3.1, is a constant that depends on the two inactivation rates ( $k_1$  and  $k_2$ ) and on the ratio of dormant spores in the test solution. The **dep** model is the only three-parameter model included in this analysis, as it is the three-parameter

model that has most frequently served as a best fitting model in some prior literature (Enger et al., 2018; Mitchell & Akram, 2017).

Table 3.1: Persistence Models Selected for the Meta-Analysis

Model	Equation	Reference
Exponential (ep)	$\log_{10} \left( \frac{N_t}{N_o} \right) = \log_{10}(e^{-k_1 t})$	(Chick, 1908; Watson, 1908)
Exponential Damped (epd)	$\log_{10} \left( \frac{N_t}{N_o} \right) = \log_{10}(e^{-k_1 t} e^{-k_2 t})$	(McKellar & Lue, 2004; Whiting, & Buchanan, Robert, L., 2001)
Juneja and Marks 1 (jm1)	$\log_{10} \left( \frac{N_t}{N_o} \right) = \log_{10}(1 - (1 - e^{-k_1 t})^{k_2})$	(Atwood & Norman, 1949; Nomiya, 2013)
Juneja and Marks 2 (jm2)	$\log_{10} \left( \frac{N_t}{N_o} \right) = \log_{10} \left( \frac{1}{1 + e^{k_1 + k_2 \ln t}} \right)$	(Carlier et al., 1996; Juneja & Marks, 2003)
Double Exponential (dep)	$\log_{10} \left( \frac{N_t}{N_o} \right) = \log_{10}(k_3 e^{-k_1 t} + (1 - k_3) e^{-k_2 t})$	(Abraham et al., 1990; Shull et al., 1963)

Models were removed from the analysis if the fitting routine found unstable parameter estimates, if the model failed to converge, or if the resulting model was U-shaped (i.e., predicted a sustained increase in concentrations after substantial decay). For each model a Bayesian Information Criterion (BIC) value, adjusted R<sup>2</sup> value and normalized root mean square error (nRMSE) value were calculated. The BIC value was calculated with Equation 1, where  $nll$  is the negative log-likelihood value,  $k$  is the number of parameters in the model, and  $n$  is the number of datapoints. BIC values for each model were used for model comparison; a lower BIC value indicates a better fitting model, and BIC values within 2 of one another indicate two models of equal fit (Dziak et al., 2020; Haas et al., 2014).

$$\text{BIC} = 2nll + k \log(n) \quad \text{Eq. 1}$$



Adjusted  $R^2$  and nRMSE values were calculated as indicators of goodness of fit. The  $R^2$  value (Equation 2) was calculated with the log-reduction values predicted by the model ( $LR_p$ ), the observed log-reductions ( $LR_o$ ), and the mean observed log-reduction ( $\overline{LR}_o$ ). As four of the five persistence models were nonlinear, an adjusted  $R^2$  value was calculated as shown in Equation 3, with the consideration of the number of datapoints in the experiment,  $n$ , and the number of parameters in the model,  $p$ . For a normalized RMSE value, the range of observed log-reductions for each dataset was used to normalize the RMSE value as shown in Equation 4. There is not a standard threshold where a model is considered a good fit versus a bad fit with these metrics, and suitable thresholds tend to vary by discipline and data type. A sensitivity analysis was conducted to assess the impact of nRMSE and Adjusted  $R^2$  threshold selection on the identification of the number of models providing good and best fits to the datasets (Appendix A). As the nRMSE values can be expressed in meaningful units, and the sensitivity analysis suggested it was sufficiently stringent, an nRMSE value of less than or equal to 10% was selected as the threshold for goodness of fit in this study.

$$R^2 = 1 - \frac{\sum(LR_p - LR_o)^2}{\sum(LR_o - \overline{LR}_o)^2} \quad \text{Eq. 2}$$

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(n - 1)}{(n - p - 1)} \quad \text{Eq. 3}$$

$$\text{nRMSE} = \frac{\sqrt{\frac{\sum(LR_p - LR_o)^2}{N}}}{\max(LR_o) - \min(LR_o)} \quad \text{Eq. 4}$$

### 3.2.3 Model-Relevant Persistence Values

By comparing BIC values, each dataset was determined to have one or more best fitting models. If the model(s) provided adequate fits to the data, they were used to calculate T90 and T99 values. T90 and T99 values are defined as the amount of time it takes to observe a 90% and

99% reduction, respectively, from the initial concentration. If the dataset had more than one model that provided a good fit to the data, model averaged T90 and T99 values were calculated with the method highlighted in previous work (Dean et al., 2020; Haas et al., 2014). Briefly, a T90 and T99 were predicted with each best fitting model, and then the model's BIC value was used to calculate a weighted average T90 and T99. The calculated T90s and T99s for each dataset were treated as the dependent variable in the exploratory factor analysis techniques (Section 3.2.4) used to explore the relationships between factors and decay. To ensure the dependent variables were truly reflective of the model's fit to the data, model averaged T90 and T99 values were only calculated when a 1 or 2-log reduction was observed within the experimental data, and at least one model provided a good fit to the dataset. If a 1 or 2-log reduction was observed within the experimental data but none of the tested models provided a good fit, the dataset was used to test the analytic techniques when applicable. Basic interpolation was used to determine the observed T90 and T99 for these datasets, henceforth referred as the "testing" data.

### 3.2.4 Factor Analysis Methods

There were six core factors documented for each dataset that was extracted or digitized from the literature review (Dean & Mitchell, 2022b): target type (FIB, bacteria, virus, bacteriophage, or protozoa), water type (fresh, marine, or brackish), sunlight (present/absent), predation (present/absent), method of detection (culture-based/molecular-based), and temperature. Other factors evaluated by the original study were also documented and three additional factors were frequently highlighted: pH, turbidity, and dissolved oxygen. The largest dataset for the exploratory factor analysis included the main core factors but additional datasets that included pH, turbidity, or dissolved oxygen were also assessed on a smaller scale.

Kruskal-Wallis tests, Pairwise-Wilcoxon tests and Spearman correlation coefficients were used to assess the basic trends between the factors and persistence values. Spearman correlation coefficients were calculated for the continuous and binary factors, and Kruskal-Wallis and Pairwise-Wilcoxon tests were calculated for the binary and categorical factors. For the factor analyses, the binary and multi-level categorical factors (sunlight, predation, water type, target type) were dummy-coded. The base condition was represented by FIB, freshwater, culture-based methods of detection, and the absence of sunlight or predation. The dependent variable in each method was the model averaged T90 or T99 values. As highlighted in the systematic review of the literature (Dean & Mitchell, 2022b), a variety of methods have been used to assess the effect of experimental factors on target persistence, including correlation coefficients, ANOVAs, multiple linear regression, and generalized linear mixed models (Ahmed et al., 2019; Avery et al., 2008; Espinosa et al., 2008; Korajkic et al., 2013, 2014, 2019; Levin-Edens et al., 2011; Liang et al., 2017; Tiwari et al., 2019; Wanjugi et al., 2016). In this analysis, these analytic methods were explored and expanded upon, as the magnitude of the dataset allowed for the comparison of technique efficacy. Regression trees, random forests, multiple linear regression, and quantile regressions were fit to the data herein. Each method requires different underlying assumptions about the data (Appendix B, and the validity of those assumptions was evaluated by comparing the performance and predictive power of each method.

To accomplish this, the data was divided into training and testing datasets as described in Section 3.2.3. The regression trees, linear models, quantile regressions, and random forests were fit to the training data. RMSE values calculated from the predicted and observed persistence values in the training data were used to evaluate each method's *performance*. Then the optimized linear models, quantile regressions, regression trees, and random forests were used to predict the

testing data persistence values as a form of method validation, and the RMSE values calculated using these predicted and observed values were used to consider the method's *predictive power*. Details for the methods used to analyze the data with regression trees, linear models, and quantile regressions are included in Appendix B. Consistently, random forests performed the best on the training data and had the greatest predictive power. As such the results of the random forests are discussed in detail herein.

Random forests are an expansion past regression trees, that fit hundreds of trees to bootstrapped training samples (Brieman, 2001; Hastie et al., 2017). During the tree-growing process, a number of input variables ( $m$ ) are randomly selected for splitting. Using the **randomForest** package and the **caret** package for tuning  $m$ , the datasets were explored with the random forest method (Kuhn, n.d.; Liaw & Wiener, 2002). Variable importance in a random forest is calculated as the mean increase in the mean square error when a variable of concern is permuted from the process (Hastie et al., 2017). H-statistics, which calculate the fraction of variance not explained by the sum of partial dependencies of the independent variables, were used to evaluate the contribution of interactions to prediction variance using the **iml** package (Friedman & Popescu, 2008; Molnar et al., 2018).

### 3.3 Results

#### 3.3.1 Persistence Modeling

There were 678 datasets extracted or digitized from the 61 studies identified in the previously published systematic literature review (Dean & Mitchell, 2022b), but only 629 datasets had at least four time points and a significant negative trend. The five persistence models in Table 3.1 were fit to each dataset and models were removed if the parameters during the fitting process were unstable, if the final optimized model was U-shaped, or if there were any

convergence errors (Table C3.1). The dep model was the most prone to unstable parameter estimates and convergence errors, and the epd model was the most likely to be removed because of a final U-shaped model form. At least one model provided a good fit ( $nRMSE \leq 0.10$ ) to 498 of the 629 datasets, with jm2 providing a good fit the most often (71%) to the 629 datasets, followed by jm1 (66%), epd (60%), ep (38%), and dep (23%). The best fitting model for each dataset was determined with BIC values. Of the 498 datasets, there was a single best fitting model for 293 datasets with the remaining 205 datasets having multiple best fitting models as shown in Table 3.2. In total, jm2 was one of (or the only) best fitting model for 52% of the 498 datasets, followed by epd (44%), jm1 (39%), ep (19%), and dep (12%). It should be noted that though the ep model provided the best fit to 93 datasets, it was the single best fitting model for only seven datasets.

Table 3.2: Best Fitting Model Frequency as Determined by BIC Values

<b>Model</b>	<b>Number of Datasets with a Best Fit (n=498)</b>	<b>Number of Datasets with One Best Fitting Model (n=295)</b>	<b>Number of Datasets with Multiple Best Fitting Models (n=203)</b>
ep	93	7	86
epd	218	91	127
jm1	193	48	145
jm2	261	133	128
dep	58	14	44

For the factor analysis, model averaged T90 and T99s were calculated for the datasets in which at least one model provided a good fit to the data, and either a 1 log-reduction and 2 log-reduction was observed in the experiment, respectively. Out of the 629 datasets, a 1-log reduction was observed in only 568 datasets, and at least one model provided a good fit to 458 of the datasets. A 2-log reduction was observed in 427 of the 629 datasets, and at least one model provided a good fit to 353 of the datasets. Thus, 458 and 353 datasets were used to train the

methods for the T90 and T99 factor analyses, respectively, and 110 and 74 datasets were used to test the methods for the T90 and T99 analyses, respectively.

### *3.3.1.1 First-Order Kinetic Comparison*

It was hypothesized in this analysis that the persistence values calculated with the best fitting persistence models identified herein, would yield a greater understanding of the relationships between factors and persistence than analyses that have previously relied upon first-order decay rates. To allow for a direct comparison, the exponential model (Table 3.1) was fit to each of the core factors datasets ( $n_{T90}=458$ ,  $n_{T99}=353$ ) and used to calculate T90 and T99 values. The distribution of exponential-predicted T99s was not significantly different ( $p=0.07$ ) from the distribution of the T99s predicted with the best fitting models for each dataset (Figure 3.1), as evaluated with a Kruskal-Wallis test. Although the distributions of T99s did not significantly differ, the exponential model did predict a higher maximum T99 value for the training datasets, and the exponential-predicted T99 distribution overall had higher central tendency values and more variance. When analyzing the data by target type (Tables D3.1-D3.2), the exponential model predicted higher central tendency and maximum values for the bacteria, bacteriophage, virus, and protozoa T99s than the model-averaging approach. The exponential model also predicted higher T90 values for the bacteria, virus, and bacteriophage targets than the model averaging approach, although the T90 distributions did not significantly differ ( $p=0.34$ ) (Table D3.2).

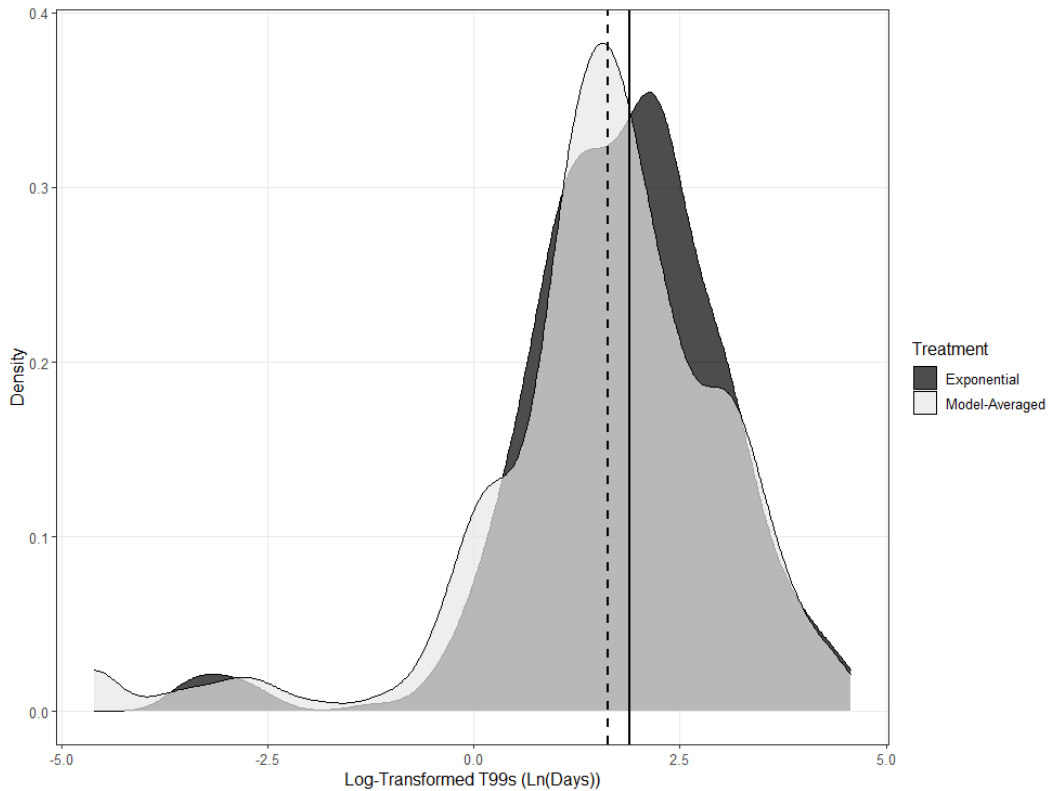


Figure 3.1: Distribution of the log-transformed exponential and model-averaged predicted T99 values for all targets and factors (n=353) combined; dashed line is the median for the model-averaged T99s, and the solid line is the median for the exponential T99s

### 3.3.2 Factor Analysis

For the core factors and T99 values, the 353 datasets used to train the factor analysis methods represented each factor as follows: target type of FIB (n=136)/Bacteria (n=120)/Virus (n=64)/Bacteriophage (n=25)/Protozoa (n=8), sunlight presence (n=106)/absence (n=247), predation presence (n=274)/absence (n=79), water type of fresh (n=246)/marine (n=76)/brackish (n=31) and culture-based (n=314)/molecular-based (n=39) methods of detection. The breakdown of the T90 dataset was similar (Table D3.1). Figure 3.2 shows the distributions of the log-transformed T99 values for each factor grouping. The number of datasets for the T99 factor analyses with pH, turbidity, and dissolved oxygen were 216, 148, and 71, respectively.

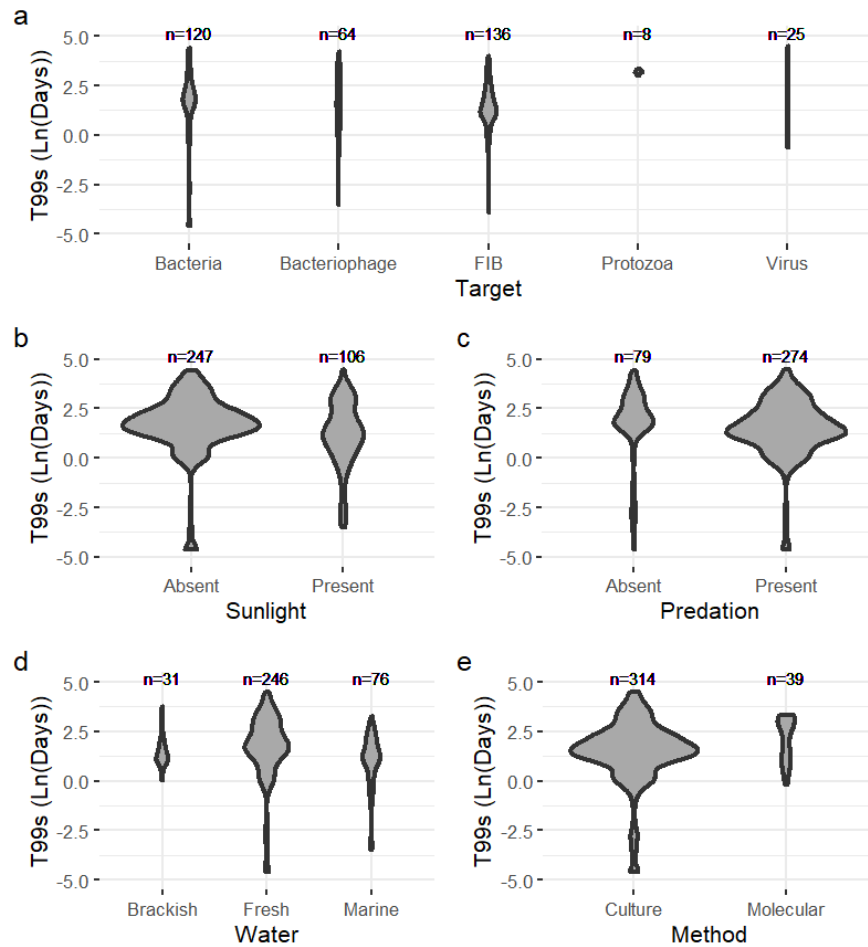


Figure 3.2: Violin plots demonstrating the distribution of log-transformed T99 values for a) targets, b) sunlight, c) predation, d) water type, and e) method of detection (scaled with count and annotated with number of datasets for each factor level)

### 3.3.2.1 Correlations

The datasets were first explored with various correlation measures. All the continuous and binary core factors were significantly correlated with the persistence values, as shown in Table 3.3. Increasing temperatures, the presence of sunlight, and the presence of predation were associated with lower persistence values, and molecular-based methods were associated with higher persistence values. Kruskal Wallis tests of the binary and categorical factors indicated that there were significant ( $p < 0.05$ ) differences between the levels of sunlight, predation, target type, water type and method of detection for the dependent variables. Per the Pairwise Wilcoxon tests,



the differences between water types were driven by the differences between fresh and marine waters and fresh and brackish waters. For the T99s, there were significant differences ( $p < 0.05$ ) between protozoa and all target types, a trend also observed for the T90s, and FIB and bacteriophages. The pairwise Wilcoxon tests for the T90 data also suggested significant differences between FIB and both bacteria and bacteriophages.

When pH was included as a factor, temperature, predation, method of detection, and pH were significantly correlated with the persistence values. Sunlight did not have a significant correlation. When turbidity was assessed, temperature, predation, method of detection, and turbidity were significantly correlated with the persistence values. For the datasets that included dissolved oxygen, temperature, predation, and method of detection were significantly associated with both the T90 and T99 values.

Table 3.3: Spearman Rank Correlation Coefficients

Factors Assessed	n	Factor	T90	n	Factor	T99
<b>Core factors</b>	458	Temperature	<b>-0.32</b>	355	Temperature	<b>-0.22</b>
		Sunlight	<b>-0.11</b>		Sunlight	<b>-0.17</b>
		Predation	<b>-0.22</b>		Predation	<b>-0.19</b>
		Method	<b>0.10</b>		Method	<b>0.14</b>
<b>pH</b>	255	Temperature	<b>-0.28</b>	215	Temperature	<b>-0.22</b>
		Sunlight	0.05		Sunlight	0.02
		Predation	<b>-0.25</b>		Predation	<b>-0.22</b>
		Method	<b>0.21</b>		Method	<b>0.28</b>
		pH	<b>-0.23</b>		pH	<b>-0.32</b>
<b>Turbidity</b>	180	Temperature	<b>-0.32</b>	148	Temperature	<b>-0.21</b>
		Sunlight	-0.04		Sunlight	-0.11
		Predation	<b>-0.27</b>		Predation	<b>-0.27</b>
		Method	<b>0.34</b>		Method	<b>0.39</b>
		Turbidity	<b>0.15</b>		Turbidity	<b>0.18</b>
<b>DO</b>	92	Temperature	<b>-0.31</b>	71	Temperature	<b>-0.24</b>
		Sunlight	-0.07		Sunlight	0.09
		Predation	<b>-0.45</b>		Predation	<b>-0.41</b>
		Method	<b>0.37</b>		Method	<b>0.56</b>
		DO	-0.01		DO	0.02

\*Bolded coefficients indicate a significance of  $p < 0.05$

### 3.3.2.2 Method Evaluation

As described in Section 3.2.4 (*Factor Analysis Methods*), there were a variety of methods implemented to explore the relationship between the documented factors and the modeled persistence metrics. The results were evaluated with RMSE values to assess method performance and predictive power. The results of each method for the T99 core factor datasets are shown in Table 3.4 and the results of each method for the T90 core factor datasets are shown in Table D3.3. The random forest method has the lowest RMSE values for each persistence value (T90 and T99) and for each data type (training and testing). Thus, the random forest method was the best suited to identify and describe the relationships between the various factor and persistence values. Notably, random forests allow for nonlinear relationships between independent and dependent variables and can also account for interactions between independent variables. As random forests are best suited for this data, this suggests there are nonlinear relationships and interactions occurring between the variables studied in this analysis.

Table 3.4: Evaluation of Factor Analysis Methods with Core Factor Data for T99 Values

Method	RMSE Values	
	Performance (n <sub>T99</sub> =353)	Predictive Power (n <sub>T99</sub> =74)
Regression Trees	11.8	16.7
Random Forest	<b>11.4</b>	<b>15.4</b>
Linear Models	13.8	17.0
Quantile Regressions*	13.4	17.5

\*Performance and predictive power RMSE values for the median quantile only

### 3.3.2.3 Random Forests

Figure 3.3 shows the variable importance from the random forests fit to the T99 core factor training data. Temperature was the most important factor affecting the persistence values, and the differentiation of protozoa from the other target types was the second most important variable. As the baseline conditions for these analyses were FIB, the elevated importance of

protozoa indicates that the T99s for protozoa differed the most from FIB compared to the other targets. This was visually evident in Figure 3.2 with there being less datasets for protozoa, and those datasets being associated with higher, less variable T99 values compared to the other targets. Marine water, method of detection and predation were the next highest ranked variables of importance for the T99 random forest. Predation was the second most important variable for the T90 values (Figure D3.1), followed by bacteriophages, brackish water, and marine water.

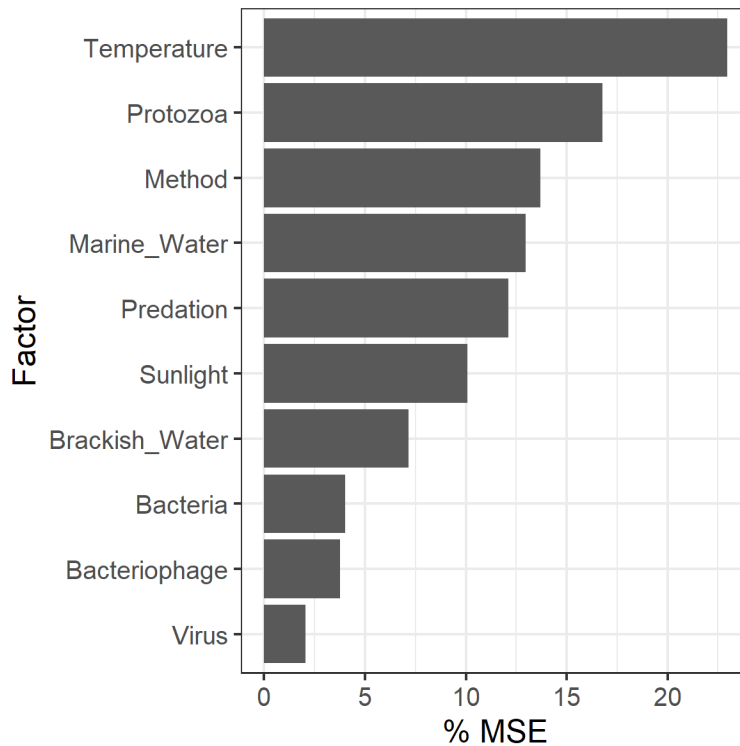


Figure 3.3: Variable importance for the T99 random forest analysis with the core factors (target type, sunlight, predation, water type, and method of detection) ranked by %MSE (percent change in mean square error when the factor is not included in the analysis)

For the T99 values, the partial dependence plots (Figure D3.2) indicated higher values typically associated with non-FIB target types and molecular-based methods. The increase in T99 was the most pronounced for the protozoa target type. Lower T99s were associated with increasing temperatures, predation presence, and marine waters instead of fresh. Sunlight presence was associated with decreasing T99s but not T90s (Figure D3.3). The literature review

suggested possible interactions between sunlight and water type, sunlight and method, water type and predation, temperature and predation, and water type and method(Dean & Mitchell, 2022b). H-statistics for general interactions (Table D3.4) indicated that temperature and predation are the factors that interact the most with other factors in relation to both T90s and T99s. For the T99 random forest, the two-way H-statistics indicated that temperature is predominantly interacting with predation, and that predation is interacting with sunlight (27%), the bacteria targets (23%), and temperature (25%). The partial dependence plots for predation and temperature suggest that for both T90s and T99s, the effects of predation may be more pronounced at the lowest temperatures evaluated, however the effect is more obvious for the T90s than the T99s.

The other interactions noted in the literature review of sunlight and water type, predation and water type, and water type and method of detection were not noted in the partial dependence plots (Dean & Mitchell, 2022b). There was however a visible interaction between sunlight and method of detection in the T99 random forest; the presence of sunlight clearly reduced the average T99 for culture-based methods, but no significant change was observed for the average T99 for molecular-based detection methods. In addition to the interactions suggested in the literature review (Dean & Mitchell, 2022b), the partial dependence plots suggested a possible sunlight and predation interaction; sunlight resulted in a greater reduction in the persistence values when predation was absent. This interaction was more obvious for the T90s than the T99s. Notably, there were fewer general interactions between the documented factors in the T99 random forest than in the T90 random forest (Table D3.4). This suggests that there are additional water quality and environmental factors affecting persistence at later time points that may not be captured in this analysis.

Although the random forests had the best performance and predictive power, the random forests fit to the core factors only described about 18% of the variance in the T99 data. When the random forests included pH, turbidity, and dissolved oxygen, a higher percentage of the variance was explained. For the T99 values, the random forest with pH, turbidity and dissolved oxygen explained 30%, 32%, and 40% of the variance, respectively (Tables D3.5-D3.6). Figure D3.4 shows the variable importance for each random forest. pH was the most important variable, turbidity was the third most important variable, and dissolved oxygen was one of the least important variables affecting T99 values in their respective random forests.

#### *3.3.2.4 Impact of Model Selection on Factor Analyses*

Random forests were fit to the exponential-predicted T90 and T99s to identify any variation in the factor analyses. When the exponential-predicted T90s and T99s were used as the dependent variables, the documented factors explained similar amounts of the observed variance (24-41%), as when the model-averaged values were used (18-36%). As shown in Figure D3.1 and D.3.5, temperature and predation were some of the most important variables affecting early persistence behaviors regardless of dependent variable treatment, and the protozoa T99s differed the most from FIB (Figure 3.4). H-statistics identified temperature and predation as the factors most involved in interactions with other factors in the random forest fit to the T90 values (Table D3.7) similarly to the model-averaged random forest (Table D3.4).

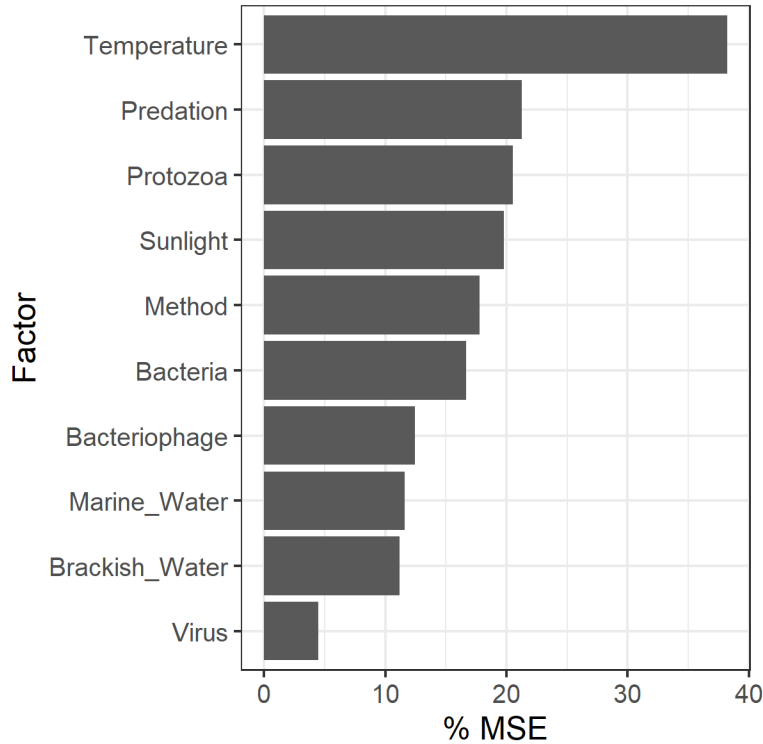


Figure 3.4: Variable importance for the exponential-predicted T99 random forest analysis with the core factors (target type, sunlight, predation, water type, and method of detection) ranked by %MSE (percent change in mean square error when the factor is not included in the analysis)

### 3.4 Discussion

The analyses presented herein consider model uncertainty in addition to data uncertainty within waterborne pathogen persistence studies utilizing a total of five candidate persistence models fit to each dataset mined from the literature review process. In the context of this study, model uncertainty was holistically evaluated with respect to goodness of fit, model selection, and applicability beyond the observable data in each study. Previous reviews and meta-analyses have represented each persistence dataset with a first-order decay rate (Boehm et al., 2018, 2019; L. E. Brooks & Field, 2016), whereas this analysis evaluated the fit of one-parameter (first-order), two-parameter, and three-parameter persistence models and then used the best fitting models to calculate persistence values (T90 or T99) as dependent variables of interest. In this analysis, 498 of the tested 629 datasets were well-described by at least one of the tested models. Consequently,

first order decay kinetics provided the best fit to only 93 of those 498 datasets. The exponential model was the sole best fitting model to only 1% of these datasets, indicating that for most of the persistence data studied herein, the other evaluated model forms could provide an equally good or improved fit to the data observed in each study. The jm2 model, based on the logistic probability distribution, was the best fitting model most often. The two-parameters in the jm2 model facilitate a sigmoidal shape, which can capture both shouldering and tailing behaviors so it may likely provide better estimates outside the range of observed data. This work further supports previous studies, in which the jm1 and jm2 models frequently have been found to describe natural decay dynamics more accurately than the ep model (Dean et al., 2020; Mitchell & Akram, 2017).

Despite it not providing the best fit to the majority of the datasets, there was no statistically significant difference between the distributions of predicted persistence values when the ep model was used compared to the best fitting models for the persistence values selected (T90 and T99). In general, the ep model predicted higher maximum and central tendency values for most of the target types (Figure 3.1). This suggests that although the jm2 model may provide a better fit to the data more frequently than the ep, the assumption of first-order decay is still useful for more immediate log-reduction values of interest (i.e., within the 1 or 2 log-reductions observed in most studies). However, it was evident that the reduction in model uncertainty becomes more pronounced as log-reduction values increase past the 2-log reduction point (Dean & Mitchell, 2022a).

It was hypothesized that a persistence value based on a more accurate persistence profile would facilitate more in-depth factor analyses. Previous meta-analyses and factor analyses have analyzed the factor-persistence relationships with correlation coefficients, ANOVAs, multiple

linear regression, generalized linear mixed models, and Bayesian hierarchical linear models (Ahmed et al., 2019; Avery et al., 2008; Boehm et al., 2019; L. E. Brooks & Field, 2016; Dean & Mitchell, 2022b; Espinosa et al., 2008; Korajkic et al., 2013, 2014, 2019; Levin-Edens et al., 2011; Liang et al., 2017; Tiwari et al., 2019; Wanjugi et al., 2016). After testing several methods (Appendix B) this study identified random forests as the method able to provide the most insight into factor-persistence relationships for the analyzed data. Random forests can account for nonlinear relationships between independent and dependent variable, and the results of the analyses presented herein suggest that there are nonlinear factor-persistence relationships for temperature and possibly turbidity, dissolved oxygen, and pH.

Previous global models identified sunlight, temperature, and various target dummy variables as significant factors (Boehm et al., 2018). The random forests fit in this analysis identified temperature, predation, method of detection, and marine water as some of the most important variables influencing T99 values. The literature review preceding this analysis suggested relevant interactions between temperature and predation (Dean & Mitchell, 2022b); the interaction was confirmed for both persistence values, with the effects of predation being more pronounced in the lowest evaluated temperature ranges. The way predation was classified in the preceding literature review (Dean & Mitchell, 2022b) (as the presence or absence of indigenous microbiota) encompasses the effects of both predation and competition. As microorganisms typically persist for longer at lower temperatures, it is possible that the significant interaction between temperature and predation is due to a greater number of populations available to prey on other microorganisms or compete for nutrients in the lower end of the evaluated temperature range. There were also possible interactions identified between sunlight and predation, and sunlight and method. The interaction between sunlight and method is



further supported by a previous meta-analysis in which the effect of light was only found to significantly affect culturable indicators (Brooks & Field, 2016). These interactions could be responsible for sunlight having a lower level of importance in these analyses compared to previous meta-analyses (Boehm et al., 2018). Furthermore, the effects of sunlight were reduced to a binary presence/absence status in this analysis, which greatly simplifies the complexities of photoinactivation compared to metrics that consider factors such as sunlight intensity and water absorbance (Boehm et al., 2018; Maraccini et al., 2016; Mattioli et al., 2017). The protozoa target group was found to differ the most from the FIB group, a result that is not unexpected given that the review of the literature suggested that temperature, sunlight, and other stressors had minimal effect of protozoa decay (Dean & Mitchell, 2022b) and that previous studies have highlighted the lack of concurrence between indicator and pathogen data (Craun et al., 1997; Harwood et al., 2005).

The inclusion of turbidity in the analyses reduced the importance of water type, suggesting turbidity may affect persistence more than the general presence or absence of salinity. When turbidity and pH were included in the analyses, method of detection was the second most important variable affecting the T99s. The literature review identified potential interactions between water type and method of detection (Dean & Mitchell, 2022b). Although this analysis did not clearly confirm the presence of interactions between method of detection and water type (fresh or marine or brackish), the results of the random forest analyses that included pH and turbidity may suggest that the interactions observed in the literature for method of detection and water type may have been the result of more specific water quality characteristics such as pH and turbidity rather than salinity. Notably, there were experiments evaluated in the systematic literature review that suggested turbidity and sunlight interactions, with elevated turbidity or

shading minimizing the inactivation effects of sunlight(Dean & Mitchell, 2022b). The random forest analyses, however, did not suggest a significant interaction between the two factors, and as such turbidity does not seem to be a driving force for sunlight being of lower importance than the other documented factors in the analyses presented herein.

Although it was hypothesized that the alternative persistence models fit in this analysis would provide a more accurate dependent variable for factor analysis methods, and that this would elucidate the finer relationships between persistence and water quality and environmental factors, the effects were subtle though the finding still significant given that this hypothesis had previously not been rigorously tested. The predominant relationships between the factors and persistence metrics and the identified interactions were similar between the model-averaged and exponential-predicted T90 and T99 values. The random forests fit to exponential-predicted persistence values assigned a higher importance to sunlight compared to other factors than the forests fit to the model-averaged datasets. Previous meta-analyses identified sunlight as one of the most important factors affecting decay (Boehm et al., 2018, 2019), and the results of this analysis suggest that some of the elevated importance may be due to dependent variable selection in addition to the aforementioned interactions and simplifications. As the effect of reduced model uncertainty is more evident at log-reduction values greater than two, it is possible that the effect of model selection would influence factor analyses with dependent variables more reflective of later time points of decay.

There were limitations to the analyses presented herein. Frequently, datasets were digitized from plots of experimental data and although the digitized datasets were visually compared to each original plot, there is still the potential for there to be slight differences between the concentrations used in this analysis and the true concentrations documented from

each study. As the models in this analysis are fit to log-reduction values, the effects of slight differences between the actual and digitized concentrations are not expected to have greatly influenced the analysis. Despite being the method best suited to describe the datasets, the random forests explained only 18-36% of the observed variance in the persistence values dependent on the core factors, suggesting there are persistence dynamics not captured by the factors documented by this or any previous analysis. The simplifications in factor designation are likely a large contributor to this unexplained variance. Indicators and pathogens were separated into broad groupings (FIB, bacteria, bacteriophages, viruses, protozoa) to attempt to maximize sample sizes and aid in the detection of interaction terms. Even within the FIB category, the literature review preceding this analysis frequently documented differences between *E. coli* or enterococci persistence behavior (Dean & Mitchell, 2022b; Jeanneau et al., 2012; Korajkic et al., 2014; Walters & Field, 2009). Additionally, the protozoa target group was composed of less than ten observations, and as such, relationships between the factors and protozoa could not be adequately explored.

It has also been noted by previous studies that the selection of the factors impacting more general water quality metrics, such as standard exceedance or FIB concentration, can be extremely site-specific (de Brauwere et al., 2014; Franczy et al., 2013). The design of this study resulted in any site-specific characteristics being summarized with a single factor with only three levels (fresh, brackish, and marine water types). It is possible that site-specific water quality and environmental factors are responsible for the unexplained variance. This is supported by the importance of pH and turbidity in their respective random forests. Generalized decay rates based on indicator behavior are typically relied upon, as site-specific and pathogen-specific decay parameters are not typically known. This assumption, however, could lead to misleading

predictions about decay, which may be especially important in water bodies that depend on natural attenuation for maintaining water quality. Based on gaps noted in this study and in previous analyses (Boehm et al., 2018), future studies aimed at evaluating mechanisms of decay should document more specific sunlight-related metrics, such as sunlight intensity or water absorbance, and site-specific water quality factors to attempt to generate more accurate descriptive and predictive models.

The meta-analysis presented herein confirmed that two-parameter persistence models more frequently describe the natural persistence profiles observed for indicators and pathogens in surface waters than the traditionally applied first-order decay kinetics. The two-parameter model based on the logistic probability distribution, **jm2**, was identified as the model best equipped to fit a variety of persistence datasets and conditions. Although the assumption of first-order decay may be adequate for the prediction of the time required for one or two-log reductions, the effect of model selection on reduced uncertainty becomes more pronounced with higher log-reduction values of interest. This study also highlighted the potential nonlinear relationships between common water quality factors (temperature, pH, turbidity and dissolved oxygen) and persistence values as well as significant interactions between the documented factors, both of which can be incorporated into future predictive models. Although temperature has frequently been highlighted as a key factor affecting persistence, the inclusion of predation in this analysis and the identification of it as important is novel compared to other meta-analyses (Boehm et al., 2018, 2019; Brooks & Field, 2016). These analyses further suggest that it could be advantageous for future researchers to document pH and turbidities for water bodies of concern, as they appear to interact with other relevant factors to influence persistence. This study improves our understanding of indicator and pathogen persistence, and the data presented herein provides

valuable information for future QMRAs which can be used to inform water management and monitoring decisions.

## REFERENCES

- Abraham, G., Debray, E., Candau, Y., & Piar, G. (1990). Mathematical Model of Thermal Destruction of *Bacillus stearothermophilus* Spores. *Applied and Environmental Microbiology*, 56(10), 3073–3080. <https://doi.org/10.1128/aem.56.10.3073-3080.1990>
- Ahmed, W., Zhang, Q., Kozak, S., Beale, D., Gyawali, P., Sadowsky, M. J., & Simpson, S. (2019). Comparative decay of sewage-associated marker genes in beach water and sediment in a subtropical region. *Water Research*, 149, 511–521. <https://doi.org/10.1016/j.watres.2018.10.088>
- Atwood, K. C., & Norman, A. (1949). On the Interpretation of Multi-Hit Survival Curves. *Proceedings of the National Academy of Sciences*, 35(12), 696–709. <https://doi.org/10.1073/pnas.35.12.696>
- Avery, L. M., Williams, A. P., Killham, K., & Jones, D. L. (2008). Survival of *Escherichia coli* O157:H7 in waters from lakes, rivers, puddles and animal-drinking troughs. *Science of The Total Environment*, 389(2–3), 378–385. <https://doi.org/10.1016/j.scitotenv.2007.08.049>
- Benham, B. L., Baffaut, C., Mankin, K., & Pachepsky, Y. (2006). Modeling bacteria fate and transport in watershed models to support TMDLs. *Transactions of the ASABE*, 49(4), 987–1002.
- Blaustein, R. A., Pachepsky, Y., Hill, R. L., Shelton, D. R., & Whelan, G. (2013). *Escherichia coli* survival in waters: Temperature dependence. *Water Research*, 47(2), 569–578. <https://doi.org/10.1016/j.watres.2012.10.027>
- Boehm, A. B., Graham, K. E., & Jennings, W. C. (2018). Can We Swim Yet? Systematic Review, Meta-Analysis, and Risk Assessment of Aging Sewage in Surface Waters. *Environmental Science & Technology*, 52(17), 9634–9645. <https://doi.org/10.1021/acs.est.8b01948>
- Boehm, A. B., Silverman, A. I., Schriewer, A., & Goodwin, K. (2019). Systematic review and meta-analysis of decay rates of waterborne mammalian viruses and coliphages in surface waters. *Water Research*, 164, 114898. <https://doi.org/10.1016/j.watres.2019.114898>
- Brieman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Brooks, L. E., & Field, K. G. (2016). Bayesian meta-analysis to synthesize decay rate constant estimates for common fecal indicator bacteria. *Water Research*, 104, 262–271. <https://doi.org/10.1016/j.watres.2016.08.005>
- Brooks, Y., Aslan, A., Tamrakar, S., Murali, B., Mitchell, J., & Rose, J. B. (2015). Analysis of the persistence of enteric markers in sewage polluted water on a solid matrix and in liquid suspension. *Water Research*, 76, 201–212. <https://doi.org/10.1016/j.watres.2015.02.039>

- Carlier, V., Augustin, J., & Rozier, J. (1996). Heat Resistance of *Listeria monocytogenes* (Phagovar 2389/2425/3274/2671/47/108/340): 0- and z-Values in Ham. *Journal of Food Protection*, 59(6), 588–591.
- Chick, H. (1908). An Investigation of the Laws of Disinfection. *Journal of Hygiene*, 8(1), 92–158. <https://doi.org/10.1017/S0022172400006987>
- Crane, S. R., & Moore, J. A. (1986). Modeling enteric bacterial die-off: A review. *Water, Air, & Soil Pollution*, 27(3–4), 411–439. <https://doi.org/10.1007/BF00649422>
- Craun, G. F., Berger, P. S., & Calderon, R. L. (1997). Coliform bacteria and waterborne disease outbreaks. *Journal - American Water Works Association*, 89(3), 96–104. <https://doi.org/10.1002/j.1551-8833.1997.tb08197.x>
- de Brauwere, A., Ouattara, N. K., & Servais, P. (2014). Modeling Fecal Indicator Bacteria Concentrations in Natural Surface Waters: A Review. *Critical Reviews in Environmental Science and Technology*, 44(21), 2380–2453. <https://doi.org/10.1080/10643389.2013.829978>
- Dean, K., & Mitchell, J. (2022a). Exploring the effects of natural stressors on indicators and pathogens in surface waters with persistence modeling. *Proceedings of the Water Environment Federation. Public Health and Water Conference & Wastewater Disease Surveillance Summit*, Cincinnati, OH. <https://www.accesswater.org/?id=10080808&fromsearch=true#iosfirsthighlight>
- Dean, K., & Mitchell, J. (2022b). Identifying water quality and environmental factors that influence indicator and pathogen decay in natural surface waters. *Water Research*, 211, 118051. <https://doi.org/10.1016/j.watres.2022.118051>
- Dean, K., Wissler, A., Hernandez-Suarez, J. S., Nejadhashemi, A. P., & Mitchell, J. (2020). Modeling the persistence of viruses in untreated groundwater. *Science of The Total Environment*, 717, 134599. <https://doi.org/10.1016/j.scitotenv.2019.134599>
- Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, 21(2), 553–565. <https://doi.org/10.1093/bib/bbz016>
- Easton, J. H., Gauthier, J. J., Lalor, M. M., & Pitt, R. E. (2005). DIE-OFF OF PATHOGENIC E. COLI O157:H7 IN SEWAGE CONTAMINATED WATERS. *Journal of the American Water Resources Association*, 41(5), 1187–1193. <https://doi.org/10.1111/j.1752-1688.2005.tb03793.x>
- Easton, J. H., Lalor, M., Gauthier, J. J., Pitt, R., Newman, D., & Meyland, S. (1999). Determination of Survival Rates for Selected Bacterial and Protozoan Pathogens from Wet Weather Discharges. *Proceedings of the Water Environment Federation. Water Environment Federation 72nd Annual Conference & Exposition*, New Orleans, LA.

- Enger, K. S., Mitchell, J., Murali, B., Birdsell, D. N., Keim, P., Gurian, P. L., & Wagner, D. M. (2018). Evaluating the long-term persistence of *Bacillus* spores on common surfaces. *Microbial Biotechnology*, *11*(6), 1048–1059. <https://doi.org/10.1111/1751-7915.13267>
- Espinosa, A. C., Mazari-Hiriart, M., Espinosa, R., Maruri-Avidal, L., Méndez, E., & Arias, C. F. (2008). Infectivity and genome persistence of rotavirus and astrovirus in groundwater and surface water. *Water Research*, *42*(10–11), 2618–2628. <https://doi.org/10.1016/j.watres.2008.01.018>
- Francy, D. S., Stelzer, E. A., Duris, J. W., Brady, A. M. G., Harrison, J. H., Johnson, H. E., & Ware, M. W. (2013). Predictive Models for Escherichia coli Concentrations at Inland Lake Beaches and Relationship of Model Variables to Pathogen Detection. *Applied and Environmental Microbiology*, *79*(5), 1676–1688. <https://doi.org/10.1128/AEM.02995-12>
- Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics*, *2*(3). <https://doi.org/10.1214/07-AOAS148>
- Gonzalez, J. M. (1995). Modelling enteric bacteria survival in aquatic systems. *Hydrobiologia*, *316*, 109–116.
- Green, H. C., Shanks, O. C., Sivaganesan, M., Haugland, R. A., & Field, K. G. (2011). Differential decay of human faecal Bacteroides in marine and freshwater: Differential decay of Bacteroides. *Environmental Microbiology*, *13*(12), 3235–3249. <https://doi.org/10.1111/j.1462-2920.2011.02549.x>
- Haas, C. N., Rose, J. B., & Gerba, C. P. (2014). *Quantitative Microbial Risk Assessment* (2nd ed.). Wiley.
- Harwood, V. J., Levine, A. D., Scott, T. M., Chivukula, V., Lukasik, J., Farrah, S. R., & Rose, J. B. (2005). Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. *Applied and Environmental Microbiology*, *71*(6), 3163–3170. <https://doi.org/10.1128/AEM.71.6.3163-3170.2005>
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning* (2nd ed.). Springer.
- Jeanneau, L., Solecki, O., Wéry, N., Jardé, E., Gourmelon, M., Communal, P.-Y., Jadas-Hécart, A., Caprais, M.-P., Gruau, G., & Pourcher, A.-M. (2012). Relative Decay of Fecal Indicator Bacteria and Human-Associated Markers: A Microcosm Study Simulating Wastewater Input into Seawater and Freshwater. *Environmental Science & Technology*, *46*(4), 2375–2382. <https://doi.org/10.1021/es203019y>
- Juneja, V. K., & Marks, H. M. (2001). *Discussion of Nonlinear Survival Curves. Presented at the Annual Meeting of the Society for Risk Analysis. Seattle, WA.*
- Juneja, V. K., & Marks, H. M. (2003). Mathematical description of non-linear survival curves of *Listeria monocytogenes* as determined in a beef gravy model system at 57.5 to 65 °C.



*Innovative Food Science & Emerging Technologies*, 4(3), 307–317.  
[https://doi.org/10.1016/S1466-8564\(03\)00025-0](https://doi.org/10.1016/S1466-8564(03)00025-0)

Juneja, V. K., Marks, H. M., & Mohr, T. (2003). Predictive Thermal Inactivation Model for Effects of Temperature, Sodium Lactate, NaCl, and Sodium Pyrophosphate on *Salmonella* Serotypes in Ground Beef. *Applied and Environmental Microbiology*, 69(9), 5138–5156. <https://doi.org/10.1128/AEM.69.9.5138-5156.2003>

Korajkic, A., McMinn, B. R., Ashbolt, N. J., Sivaganesan, M., Harwood, V. J., & Shanks, O. C. (2019). Extended persistence of general and cattle-associated fecal indicators in marine and freshwater environment. *Science of The Total Environment*, 650, 1292–1302. <https://doi.org/10.1016/j.scitotenv.2018.09.108>

Korajkic, A., McMinn, B. R., Shanks, O. C., Sivaganesan, M., Fout, G. S., & Ashbolt, N. J. (2014). Biotic Interactions and Sunlight Affect Persistence of Fecal Indicator Bacteria and Microbial Source Tracking Genetic Markers in the Upper Mississippi River. *Applied and Environmental Microbiology*, 80(13), 3952–3961. <https://doi.org/10.1128/AEM.00388-14>

Korajkic, A., Wanjugi, P., & Harwood, V. J. (2013). Indigenous Microbiota and Habitat Influence *Escherichia coli* Survival More than Sunlight in Simulated Aquatic Environments. *Applied and Environmental Microbiology*, 79(17), 5329–5337. <https://doi.org/10.1128/AEM.01362-13>

Kuhn, M. (n.d.). caret: Classification and Regression Training. *R Package Version 6.0-88*. <https://CRAN.R-project.org/package=caret>

Levin-Edens, E., Bonilla, N., Meschke, J. S., & Roberts, M. C. (2011). Survival of environmental and clinical strains of methicillin-resistant *Staphylococcus aureus* [MRSA] in marine and fresh waters. *Water Research*, 45(17), 5681–5686. <https://doi.org/10.1016/j.watres.2011.08.037>

Liang, L., Goh, S. G., & Gin, K. Y. H. (2017). Decay kinetics of microbial source tracking (MST) markers and human adenovirus under the effects of sunlight and salinity. *Science of The Total Environment*, 574, 165–175. <https://doi.org/10.1016/j.scitotenv.2016.09.031>

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22.

Little, J. B. (1968). Cellular effects of ionizing radiation. *The New England Journal of Medicine*, 272(7), 369–376.

Maraccini, P. A., Mattioli, M. C. M., Sassoubre, L. M., Cao, Y., Griffith, J. F., Ervin, J. S., Van De Werfhorst, L. C., & Boehm, A. B. (2016). Solar Inactivation of Enterococci and *Escherichia coli* in Natural Waters: Effects of Water Absorbance and Depth. *Environmental Science & Technology*, 50(10), 5068–5076. <https://doi.org/10.1021/acs.est.6b00505>

- Mattioli, M. C., Sassoubre, L. M., Russell, T. L., & Boehm, A. B. (2017). Decay of sewage-sourced microbial source tracking markers and fecal indicator bacteria in marine waters. *Water Research*, *108*, 106–114. <https://doi.org/10.1016/j.watres.2016.10.066>
- McKellar, R., C., & Lue, X. (2004). Primary Models. In *Modeling Microbial Responses in Food*. CRC Press.
- Medema, G., Bahara, M., & Schets, F. M. (1997). Survival of *Cryptosporidium parvum*, *Escherichia coli*, faecal enterococci, and *Clostridium perfringens* in river water: Influence of temperature and autochthonous microorganisms. *Water Science and Technology*, *35*(11/12), 249–252. [https://doi.org/10.1016/S0273-1223\(97\)00267-9](https://doi.org/10.1016/S0273-1223(97)00267-9)
- Mitchell, J., & Akram, S. (2017). Pathogen Specific Persistence Modeling Data. In *Global Water Pathogens Project* (J.B. Rose and B. Jiménez-Cisneros). UNESCO.
- Molnar, C., Casalicchio, G., & Bischl, B. (2018). iml: An R package for Interpretable Machine Learning. *The Journal of Open Source Software*, *3*(26), 786. <https://doi.org/10.21105/joss.00786>
- Nomiya, T. (2013). Discussions on target theory: Past and present. *Journal of Radiation Research*, *54*(6), 1161–1163. <https://doi.org/10.1093/jrr/rrt075>
- Pachepsky, Y. A., Sadeghi, A. M., Bradford, S. A., Shelton, D. R., Guber, A. K., & Dao, T. (2006). Transport and fate of manure-borne pathogens: Modeling perspective. *Agricultural Water Management*, *86*(1–2), 81–92. <https://doi.org/10.1016/j.agwat.2006.06.010>
- Park, Y., Pachepsky, Y., Shelton, D., Jeong, J., & Whelan, G. (2016). Survival of Manure-borne *Escherichia coli* and Fecal Coliforms in Soil: Temperature Dependence as Affected by Site-Specific Factors. *Journal of Environmental Quality*, *45*(3), 949–957. <https://doi.org/10.2134/jeq2015.08.0427>
- R Core Team. (2020). *R: A language and environmnet for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Shull, J. J., Cargo, G. T., & Ernst, R. R. (1963). Kinetics of Heat Activation and of Thermal Death of Bacterial Spores. *Applied Microbiology*, *11*, 485–487.
- Sivaganesan, M., Rice, E. W., & Mariñas, B. J. (2003). A Bayesian method of estimating kinetic parameters for the inactivation of *Cryptosporidium parvum* oocysts with chlorine dioxide and ozone. *Water Research*, *37*(18), 4533–4543. [https://doi.org/10.1016/S0043-1354\(03\)00412-3](https://doi.org/10.1016/S0043-1354(03)00412-3)
- Tamrakar, S. B., Henley, J., Gurian, P. L., Gerba, C. P., Mitchell, J., Enger, K., & Rose, J. B. (2017). Persistence analysis of poliovirus on three different types of fomites. *Journal of Applied Microbiology*, *122*(2), 522–530. <https://doi.org/10.1111/jam.13299>

- Tiwari, A., Kauppinen, A., & Pitkänen, T. (2019). Decay of *Enterococcus faecalis*, *Vibrio cholerae* and MS2 Coliphage in a Laboratory Mesocosm Under Brackish Beach Conditions. *Frontiers in Public Health*, 7, 269. <https://doi.org/10.3389/fpubh.2019.00269>
- Walters, S. P., & Field, K. G. (2009). Survival and persistence of human and ruminant-specific faecal *Bacteroidales* in freshwater microcosms. *Environmental Microbiology*, 11(6), 1410–1421. <https://doi.org/10.1111/j.1462-2920.2009.01868.x>
- Wanjugi, P., Fox, G. A., & Harwood, V. J. (2016). The Interplay Between Predation, Competition, and Nutrient Levels Influences the Survival of *Escherichia coli* in Aquatic Environments. *Microbial Ecology*, 72(3), 526–537. <https://doi.org/10.1007/s00248-016-0825-6>
- Watson, H. E. (1908). A Note on the Variation of the Rate of Disinfection with Change in the Concentration of the Disinfectant. *Epidemiology and Infection*, 8(4), 536–542. <https://doi.org/10.1017/S0022172400015928>
- Whiting, R. C., & Buchanan, Robert, L. (2001). Predictive Modeling and Risk Assessment. In *Food Microbiology Fundamentals and Frontiers* (2nd ed.). ASM Press.

## APPENDIX A: GOODNESS OF FIT METRIC SENSITIVITY ASSESSMENT

A sensitivity assessment was conducted to evaluate the influence of different metric (nRMSE vs Adjusted R<sup>2</sup>) and threshold (increment changes of 5%) selections on the determination of the number of datasets that had a model provide a good fit to the data, and the subsequent number of times each model type was the best fit to the data (as determined by BIC values). nRMSE values from 0.30 to 0.05 (less stringent to more stringent), and Adjusted R<sup>2</sup> values of 0.70 to 0.95 (less stringent to more stringent) were tested, and the results are shown in Tables A3.1 and A3.2. As expected, the more stringent thresholds resulted in less datasets being determined to have a good fitting model, as shown in Table A3.1. The differences in datasets selected was the greatest for the shift from an nRMSE 0.10 to 0.05 and an Adjusted R<sup>2</sup> of 0.90 to 0.95, with more than 100 datasets being excluded. The other increments of 0.05 for both metrics were associated with smaller changes in the number of datasets, and thus the thresholds of an nRMSE  $\leq 0.10$  and Adjusted R<sup>2</sup> of 0.90 were further considered for application with the data analyzed in this study. The nRMSE metric was the final selected metric for use because of it has a more data-relevant and interpretable meaning (difference in predicted to observed log-reduction values divided by the total range of observed log-reduction values). Notably, regardless of metric or threshold selection, the percentage of each good fitting model providing the best fit to the data remained relatively constant (Table A3.2) with the following rank: jm2 (45-58%), epd (42-44%), jm1 (35-41%), ep (12-24%) and dep (9-13%).

Table A3.1: Sensitivity Assessment of the Selected Goodness of Fit Thresholds; Impact on Models Identified as a Good Fit (GF)

Metric	Threshold	Number of Datasets with a GF	Number of Datasets with a Good Fitting Model					Percentage out of Total (n=629)				
			ep	epd	jm1	jm2	dep	%ep	%epd	%jm1	%jm2	%dep
nRMSE	$\leq 0.05$	304	55	200	186	190	56	0.18	0.66	0.61	0.62	0.18
	$\leq 0.10$	498	242	379	418	448	143	0.49	0.76	0.84	0.9	0.29
	$\leq 0.15$	575	394	452	523	548	193	0.69	0.79	0.91	0.95	0.34
	$\leq 0.20$	603	504	478	584	588	222	0.84	0.79	0.97	0.98	0.37
	$\leq 0.25$	617	563	488	600	600	243	0.91	0.79	0.97	0.97	0.39
	$\leq 0.30$	621	590	489	607	605	244	0.95	0.79	0.98	0.97	0.39
Adjusted R <sup>2</sup>	$\geq 0.95$	354	114	236	227	253	62	0.32	0.67	0.64	0.71	0.18
	$\geq 0.90$	460	222	340	365	384	110	0.48	0.74	0.79	0.83	0.24
	$\geq 0.85$	512	288	386	431	459	133	0.56	0.75	0.84	0.9	0.26
	$\geq 0.80$	541	353	412	474	500	152	0.65	0.76	0.88	0.92	0.28
	$\geq 0.75$	559	398	432	504	526	166	0.71	0.77	0.9	0.94	0.3
	$\geq 0.70$	579	445	450	537	548	180	0.77	0.78	0.93	0.95	0.31

Table A3.2: Sensitivity Assessment of the Selected Goodness of Fit Thresholds on the Models Identified as Providing the Best Fit to the Data

Metric	Threshold	Number of Datasets with a GF	Number of Datasets with a Best Fitting Model					Percentage out of Number with a GF				
			ep	epd	jm1	jm2	dep	%ep	%epd	%jm1	%jm2	%dep
nRMSE	≤0.05	304	35	129	105	136	26	0.12	0.42	0.35	0.45	0.09
	≤0.10	498	93	218	193	261	58	0.19	0.44	0.39	0.52	0.12
	≤0.15	575	124	248	225	323	67	0.22	0.43	0.39	0.56	0.12
	≤0.20	603	140	263	246	348	74	0.23	0.44	0.41	0.58	0.12
	≤0.25	617	149	265	254	357	79	0.24	0.43	0.41	0.58	0.13
	≤0.30	621	151	265	256	359	79	0.24	0.43	0.41	0.58	0.13
Adjusted R <sup>2</sup>	≥0.95	354	47	153	123	168	30	0.13	0.43	0.35	0.47	0.08
	≥0.90	460	80	201	170	242	48	0.17	0.44	0.37	0.53	0.1
	≥0.85	512	94	220	189	277	55	0.18	0.43	0.37	0.54	0.11
	≥0.80	541	103	228	200	299	61	0.19	0.42	0.37	0.55	0.11
	≥0.75	559	116	239	213	311	62	0.21	0.43	0.38	0.56	0.11
	≥0.70	579	128	250	227	327	65	0.22	0.43	0.39	0.56	0.11

## APPENDIX B: FACTOR ANALYSIS METHODS

1. **Regression Trees:** Basic regression trees were used to further assess the factor-persistence relationships following the Classification and Regression Tree (CART) algorithm (Breiman, Freidman, Olshen and Stone 1984; Therneau & Atkinson, 2019). The regression tree methodology gives an indication of some of the more important variables influencing persistence and predicts an average response based on those determined criteria using mean values and assuming linear relationships. Regression trees can be useful for identifying important variables in an analysis, however they are prone to over fitting the training data and thus have poor predictive power (Hastie, Tibshirani, & Friedman, 2017).
2. **Linear Models:** The use of multiple linear regressions for the evaluation of experimental factors is noted in the literature (Liang et al., 2017; Levin-Edens et al., 2011) and recent meta-analyses used multiple linear regression to explore the effect of water quality and environmental factors on first-order decay rates of targets in natural surface waters (Boehm et al., 2018; Boehm et al., 2019). To apply the multiple linear regression methodology in this analysis, it was necessary to log-transform the dependent variable. A monotonic modification to the log-transformation was used for datasets that included a predicted T90 value of zero. As shown in Equation S1, the minimum model-averaged T90 value that was greater than zero was halved and added to each model-averaged T90 value before being log-transformed.

$$T90_{LT} = \ln\left(T90_{MA} + \frac{\min(T90_{MA})}{2}\right) \quad \text{Eq. S1}$$

Basic linear models were fit to the data using the **lm()** function in R, following the basic format shown in Equation S2, where  $f(X)$  is the distribution of T90 or T99 values,  $X_i$  is the value for factor  $i$ ,  $B_i$  is the coefficient for factor  $i$ , and  $B_0$  is the intercept value (R Core Team, 2020; Hastie, Tibshirani, & Friedman, 2017). Linear models were also fit that incorporated interactions noted in the systematic literature review (Dean & Mitchell, 2021): sunlight and water type, sunlight and method, water type and predation, temperature and predation, and water type and method.

$$f(X) = \beta_0 + \sum_{i=1}^n X_i \beta_i \quad \text{Eq. S2}$$

3. **Quantile Regressions:** Regression trees, random forests, and linear models rely on mean values within the data. The distribution of model averaged T90 and T99 values had a very high variance, indicating that median values may be more appropriate for identifying factor-persistence relationships within this data. To facilitate this, quantile regressions were also fit to the data with the **quantreg** package. The quantile regressions utilized the same formula as the linear regressions (Equation 5). Unlike linear regressions, quantile regressions evaluate the effect of independent variables on the quantiles of the dependent variable and do not make any assumptions about the distributions of the data (Das, Krzywinski, & Altman, 2019). When the data was available, three quantiles, the 0.10, median, and 0.90, were fit to each dataset.

## APPENDIX C: MODEL FITTING DETAILS

Table C3.1: Models Removed During the Fitting Process

Reason for Removal	Number of Models (Percent of n=629)				
	ep	epd	jm1	jm2	dep
Unstable Parameters	0 (0%)	4 (1%)	12 (2%)	22 (3%)	202 (32%)
U-Shaped Model Form	0 (0%)	135 (21%)	0 (0%)	0 (0%)	37 (6%)
Convergence Error	0 (0%)	2 (1%)	4 (1%)	0 (0%)	151 (24%)

**APPENDIX D: ADDITIONAL FACTOR ANALYSIS DETAILS AND RESULTS**

Table D3.1: Descriptive Characteristics for the Model-Averaged T90s/T99s (Days)

Factor & Levels	T90 (T99) in Days					
	N	Mean	Median	Variance	Min	Max
Total	458 (353)	7.9 (10.4)	3.0 (5.2)	189.3 (193.7)	0 (0.01)	116.0 (88.9)
Target Type						
FIB	201 (136)	5.2 (7.5)	2.6 (3.9)	84.5 (83.9)	0.01 (0.02)	75.7 (54.8)
Bacteria	139 (120)	9.5 (10.8)	3.9 (5.8)	295.6 (249.3)	0 (0.01)	116.0 (83.1)
Bacteriophage	77 (64)	9.6 (12.7)	3.8 (6.0)	158.7 (223.7)	0.01 (0.03)	80.0 (69.3)
Virus	32 (25)	9.6 (5.4)	2.5 (5.4)	276.7 (406.4)	0.2 (0.5)	88.2 (88.9)
Protozoa	9 (8)	23.2 (24.2)	16.9 (24.6)	496.6 (7.5)	11.2 (19.7)	82 (27.7)
Sunlight						
Present	127 (106)	6.1 (8.4)	2.7 (3.4)	67.3 (140.3)	0.01 (0.03)	45.5 (88.9)
Absent	331 (247)	8.6 (11.2)	3.2 (5.9)	234.8 (215.0)	0 (0.01)	116.0 (83.1)
Predation						
Present	351 (274)	6.1 (9.3)	2.7 (4.6)	89.4 (163.1)	0 (0.01)	82.0 (88.9)
Absent	107 (79)	13.7 (14.0)	5.2 (7.0)	476.9 (286.0)	0 (0.01)	116.0 (83.1)
Water Type						
Fresh	297 (246)	10.1 (12.4)	4.0 (6.1)	266.9 (247.2)	0 (0.01)	116.0 (88.9)
Marine	101 (76)	4.1 (5.2)	2.5 (3.6)	17.9 (25.9)	0.01 (0.03)	20.8 (26.8)
Brackish	60 (31)	3.4 (6.7)	2.2 (3.6)	26.3 (74.9)	0.5 (1.0)	36.6 (43.6)
Method of Detection						
Culture-Based	410 (314)	7.9 (10.0)	2.9 (4.9)	206.6 (205.3)	0 (0.1)	116.0 (88.9)
Molecular-Based	48 (39)	8.2 (13.4)	6.3 (17.0)	43.1 (93.2)	0.2 (0.8)	21.3 (27.7)



Table D3.2: Descriptive Characteristics for the Exponential Predicted T90s/T99s (Days)

Factor & Levels	T90 (T99) in Days					
	N	Mean	Median	Variance	Min	Max
Total	458 (353)	7.9 (11.2)	3.5 (6.6)	200.2 (202.5)	0.02 (0.03)	140.4 (96.8)
Target Type						
FIB	201 (136)	4.8 (7.0)	2.4 (3.7)	77.3 (61.8)	0.02 (0.04)	85.3 (51.9)
Bacteria	139 (120)	9.6 (12.7)	4.5 (8.4)	331.1 (245.3)	0.02 (0.03)	140.4 (89.9)
Bacteriophage	77 (64)	10.0 (14.0)	4.5 (8.1)	151.9 (269.1)	0.5 (1.0)	80.2 (76.8)
Virus	32 (25)	11.2 (14.1)	4.7 (6.5)	324.3 (470.1)	0.2 (0.3)	91.7 (96.8)
Protozoa	9 (8)	20.9 (26.6)	13.2 (26.2)	527.0 (35.7)	9.4 (18.8)	81.7 (35.7)
Sunlight						
Present	127 (106)	4.8 (8.0)	1.9 (3.4)	39.5 (147.3)	0.02 (0.03)	48.4 (96.8)
Absent	331 (247)	9.1 (12.5)	4.1 (8.1)	257.1 (220.8)	0.3 (0.7)	140.4 (89.9)
Predation						
Present	351 (274)	6.1 (10.2)	3.0 (5.9)	85.4 (169.0)	0.02 (0.04)	81.7 (96.8)
Absent	107 (79)	13.8 (14.6)	5.7 (8.7)	535.1 (306.3)	0.02 (0.03)	140.4 (89.9)
Water Type						
Fresh	297 (246)	10.2 (13.4)	4.3 (8.1)	286.0 (262.1)	0.03 (0.1)	140.4 (96.8)
Marine	101 (76)	4.3 (6.4)	2.6 (3.7)	20.2 (34.6)	0.02 (0.03)	21.4 (24.2)
Brackish	60 (31)	2.4 (5.2)	2.0 (3.7)	2.0 (13.8)	0.9 (1.9)	8.8 (17.5)
Method of Detection						
Culture-Based	410 (314)	7.9 (10.7)	3.3 (6.2)	220.3 (214.0)	0.02 (0.03)	140.4 (96.8)
Molecular-Based	48 (39)	8.0 (14.6)	8.0 (17.0)	29.4 (98.9)	0.5 (1.1)	20.5 (35.7)

Table D3.3: Evaluation of Factor Analysis Methods with Core Factor Data for Model-Averaged T90 Values

Method	RMSE Values	
	Performance (n <sub>T90</sub> =458)	Predictive Power (n <sub>T90</sub> =110)
Regression Trees	10.2	15.7
Random Forest	<b>8.3</b>	<b>14.6</b>
Linear Models	13.0	18.8
Quantile Regressions*	12.7	18.7

\*Performance and predictive power RMSE values for the median quantile only

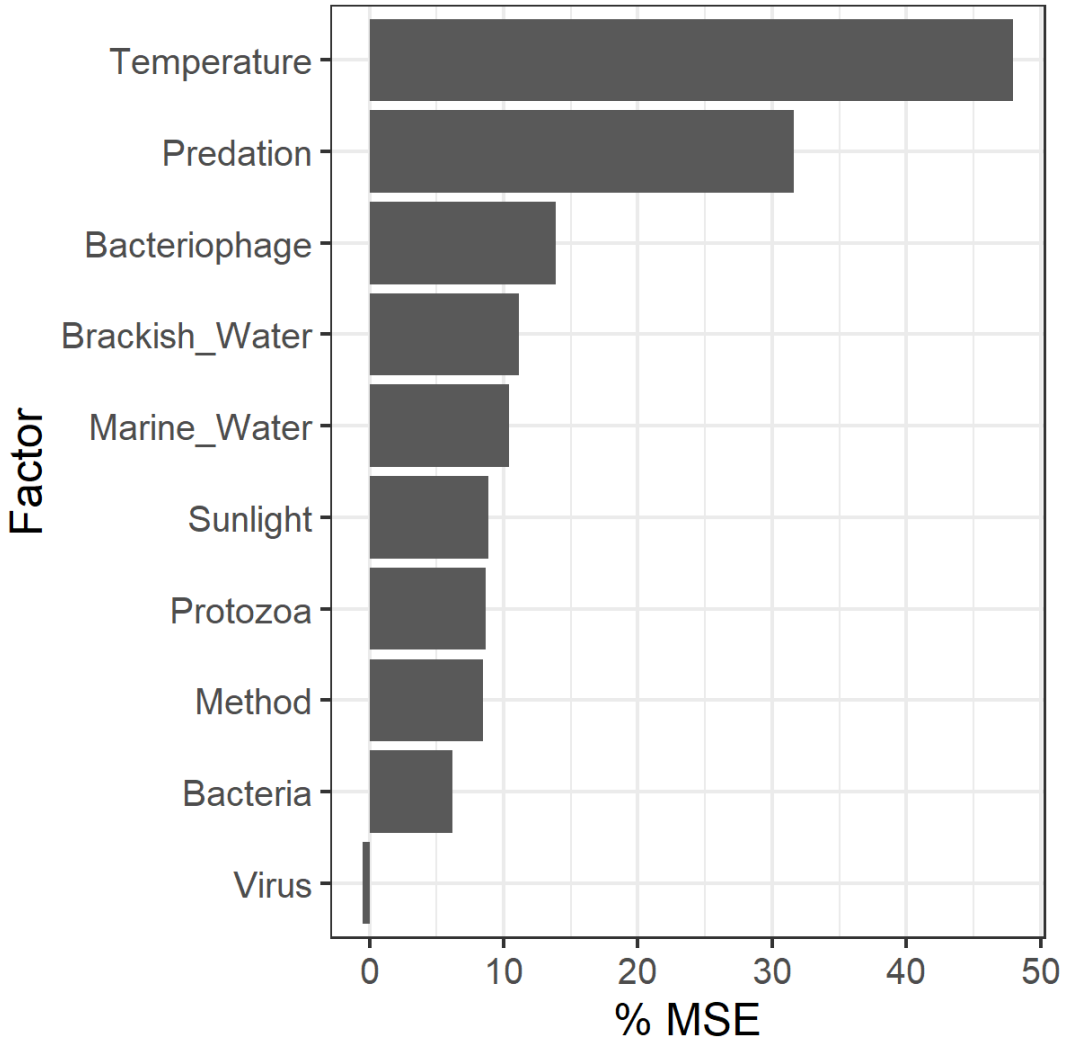


Figure D3.1: Variable importance for the random forest fit to the T90 core factor training dataset ranked by %MSE (percent change in mean square error when a factor is not included in the analysis)

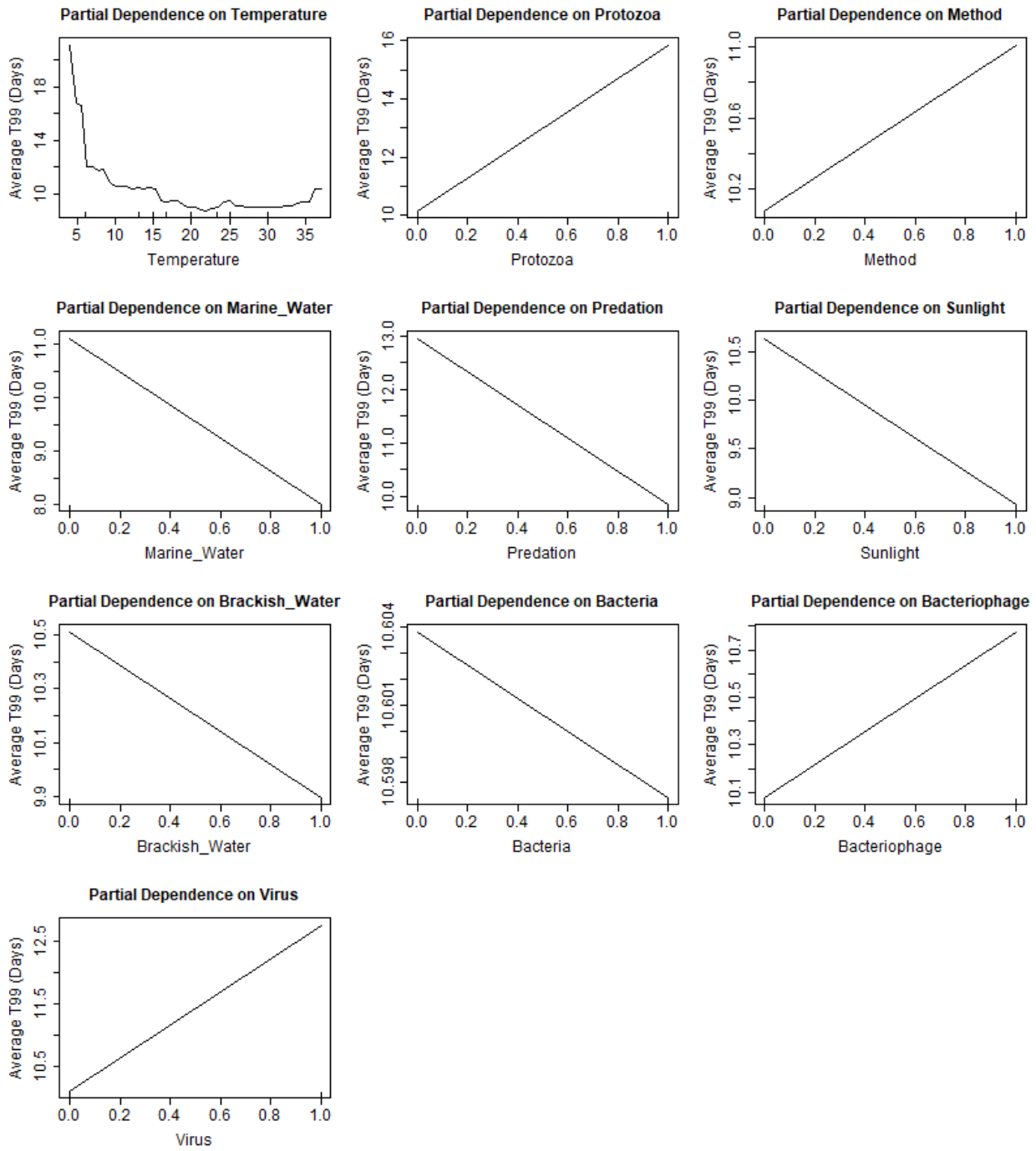


Figure D3.2: Partial dependence plots for the T99 random forest illustrating the effect of changing factor status or values on the average T99 value

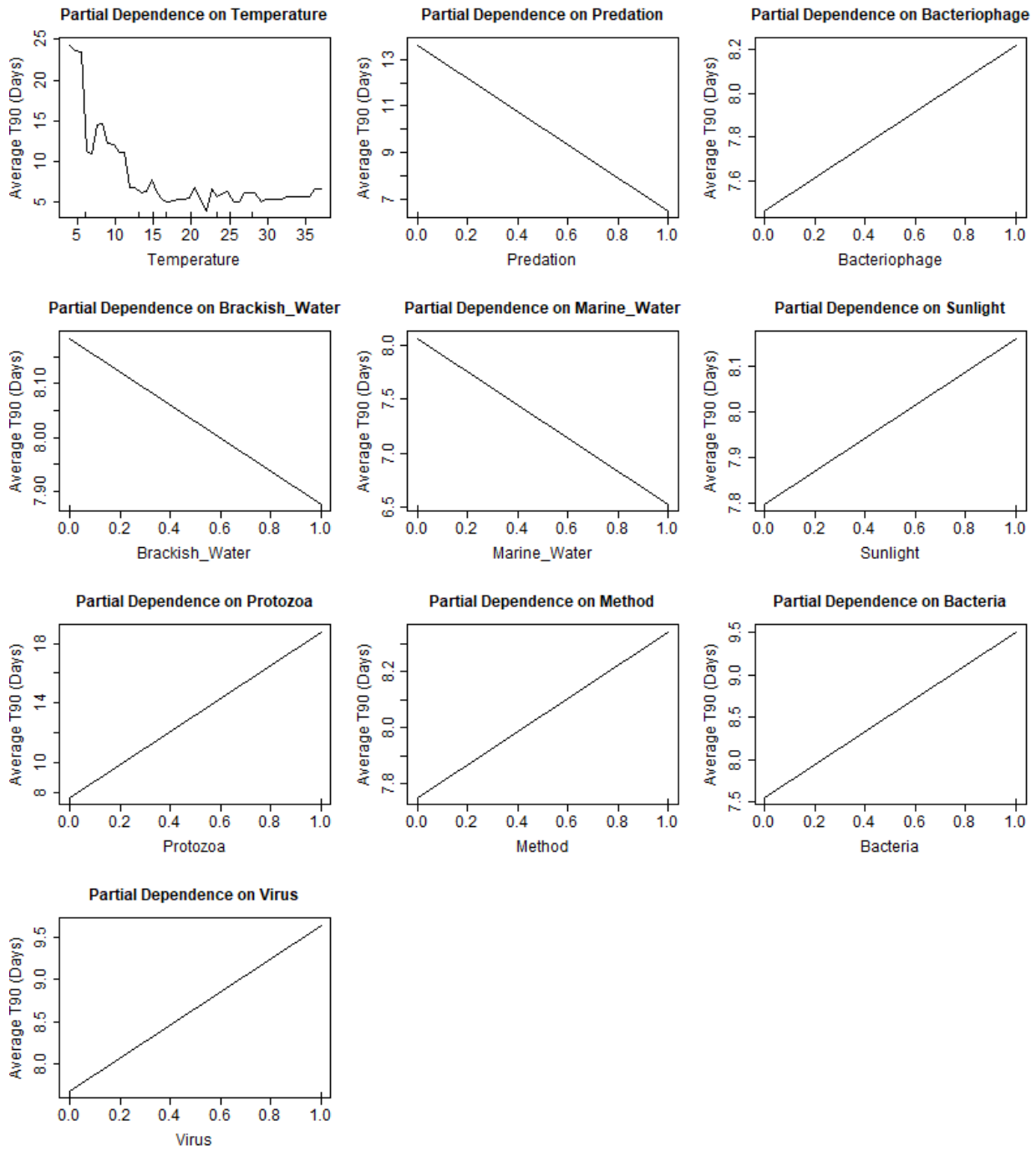


Figure D3.3: Partial dependence plots for the T90 random forest illustrating the effect of changing factor status or values on the average T90 value

Table D3.4: H-Statistics for Core Factor Random Forests fit the Model-Averaged Predicted T90s and T99s

General Interactions*						
Factor	T90			T99		
	Average	Lower Bound	Upper Bound	Average	Lower Bound	Upper Bound
Temperature	0.67	0.64	0.69	0.29	0.28	0.30
Sunlight	0.16	0.15	0.18	0.22	0.21	0.23
Predation	0.64	0.61	0.66	0.23	0.22	0.24
Marine Water	0.18	0.16	0.20	0.16	0.14	0.18
Brackish Water	0.13	0.12	0.15	0.11	0.10	0.13
Bacteria	0.22	0.20	0.23	0.14	0.13	0.15
Virus	0.12	0.10	0.14	0.15	0.13	0.18
Bacteriophage	0.25	0.22	0.27	0.12	0.12	0.13
Protozoa	0.08	0.06	0.10	0.06	0.05	0.08
Method	0.11	0.10	0.13	0.19	0.17	0.21
Specific Two-Way Interactions with Temperature*						
Factor	T90			T99		
	Average	Lower Bound	Upper Bound	Average	Lower Bound	Upper Bound
Sunlight	0.12	0.11	0.14	0.14	0.13	0.15
Predation	0.89	0.83	0.96	0.24	0.24	0.25
Marine Water	0.10	0.09	0.11	0.07	0.07	0.08
Brackish Water	0.16	0.14	0.17	0.15	0.13	0.16
Bacteria	0.17	0.16	0.18	0.11	0.10	0.11
Virus	0.07	0.07	0.08	0.07	0.06	0.08
Bacteriophage	0.12	0.11	0.13	0.09	0.09	0.1
Protozoa	0.05	0.04	0.07	0.01	0.01	0.01
Method	0.05	0.04	0.05	0.07	0.06	0.07
Specific Two-Way Interactions with Predation*						
Factor	T90			T99		
	Average	Lower Bound	Upper Bound	Average	Lower Bound	Upper Bound
Temperature	0.84	0.78	0.90	0.25	0.24	0.26
Sunlight	0.16	0.15	0.17	0.27	0.25	0.28
Marine Water	0.05	0.05	0.05	0.05	0.04	0.05
Brackish Water	0.04	0.04	0.05	0.12	0.10	0.13
Bacteria	0.07	0.07	0.07	0.23	0.23	0.24
Virus	0.02	0.01	0.02	0.02	0.02	0.03
Bacteriophage	0.07	0.06	0.07	0.15	0.15	0.15
Protozoa	0.07	0.06	0.09	0.04	0.03	0.05
Method	0.07	0.07	0.07	0.07	0.07	0.07

\*H-statistics averaged over 100 iterations

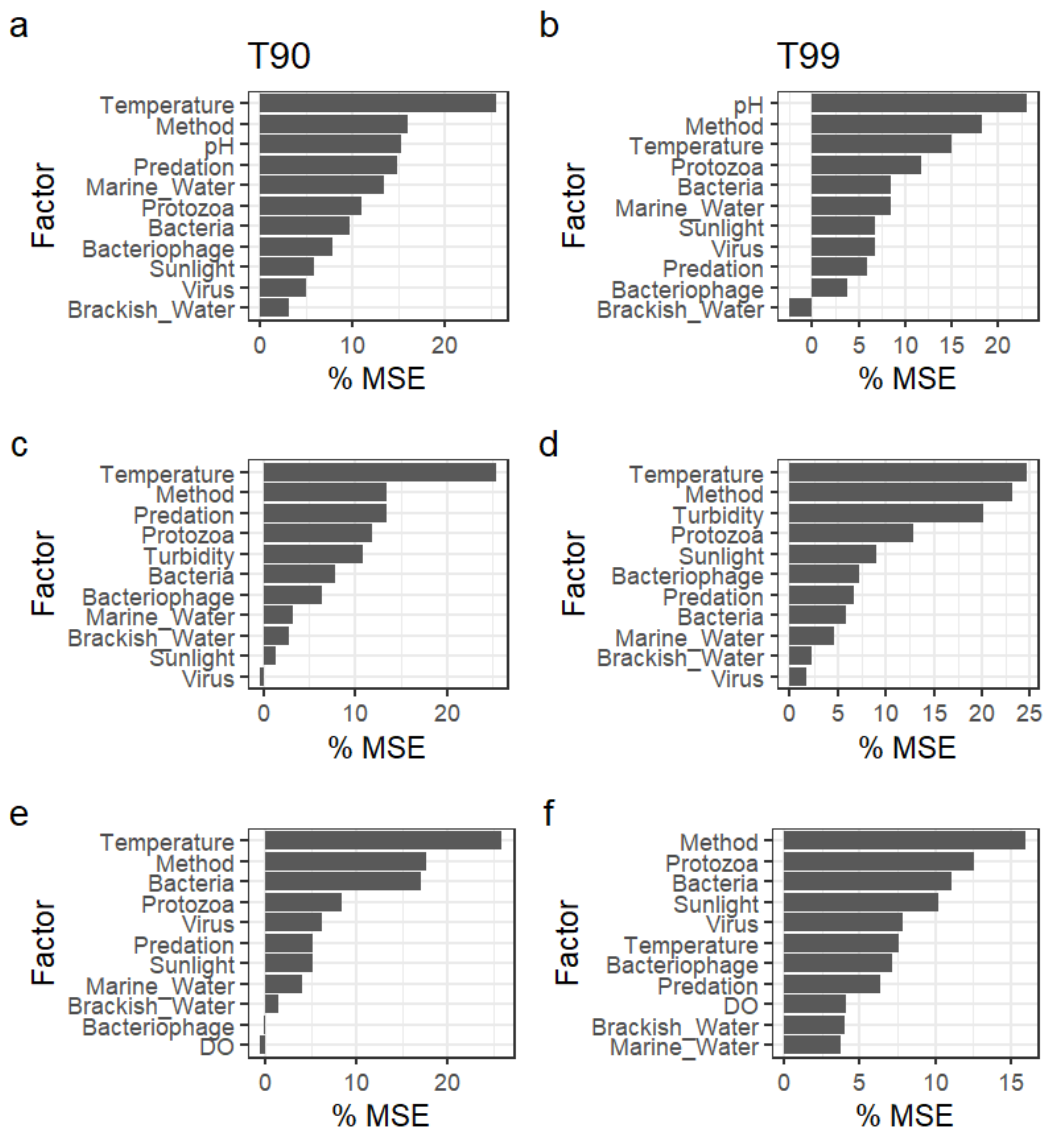


Figure D3.4: Random forest variable importance for datasets with the core factors and a/b) pH, c/d) Turbidity, or e/f) Dissolved Oxygen for a/c/e) T90 values and b/d/f) T99 values.

Table D3.5: T90 Random Forest Variable Importance for Datasets Including pH, Turbidity, and Dissolved Oxygen

T90 w/ pH		T90 w/ Turbidity		T90 w/ Dissolved Oxygen	
Variable	%IncMSE	Variable	%IncMSE	Variable	%IncMSE
Temperature	25.5	Temperature	25.4	Temperature	25.9
Method	16.0	Method	13.5	Method	17.6
pH	15.2	Predation	13.4	Bacteria	17.0
Predation	14.8	Protozoa	11.8	Protozoa	8.3
Marine Water	13.4	Turbidity	10.8	Virus	6.2
Protozoa	10.9	Bacteria	7.8	Predation	5.2
Bacteria	9.7	Bacteriophage	6.3	Sunlight	5.1
Bacteriophage	7.9	Marine Water	3.2	Marine Water	4.0
Sunlight	5.9	Brackish Water	2.9	Brackish Water	1.4
Virus	5.0	Sunlight	1.3	Bacteriophage	-0.1
Brackish Water	3.2	Virus	-0.4	DO	-0.6
Performance (n=255)	6.6	Performance (n=180)	7.6	Performance (n=92)	5.3
Predictive Power (n=51)	22.7	Predictive Power (n=43)	17.4	Predictive Power (n=30)	13.9
Variance Explained	57%	Variance Explained	59%	Variance Explained	79%

Table D3.6: T99 Random Forest Variable Importance for Datasets Including pH, Turbidity, and Dissolved Oxygen

T99 w/ pH		T99 w/ Turbidity		T99 w/ Dissolved Oxygen	
Variable	%IncMSE	Variable	%IncMSE	Variable	%IncMSE
pH	23.0	Temperature	24.6	Method	16.0
Method	18.3	Method	23.1	Protozoa	12.6
Temperature	15.0	Turbidity	20.2	Bacteria	11.1
Protozoa	11.7	Protozoa	12.8	Sunlight	10.2
Bacteria	8.5	Sunlight	9.1	Virus	7.8
Marine Water	8.4	Bacteriophage	7.3	Temperature	7.5
Sunlight	6.8	Predation	6.7	Bacteriophage	7.2
Virus	6.7	Bacteria	5.9	Predation	6.4
Predation	5.9	Marine Water	4.6	DO	4.1
Bacteriophage	3.8	Brackish Water	2.3	Brackish Water	4.0
Brackish Water	-2.4	Virus	1.8	Marine Water	3.7
Performance (n=216)	9.9	Performance (n=148)	7.9	Performance (n=71)	7.2
Predictive Power (n=39)	18.5	Predictive Power (n=35)	15.8	Predictive Power (n=25)	19.7
Variance Explained	30%	Variance Explained	32%	Variance Explained	40%

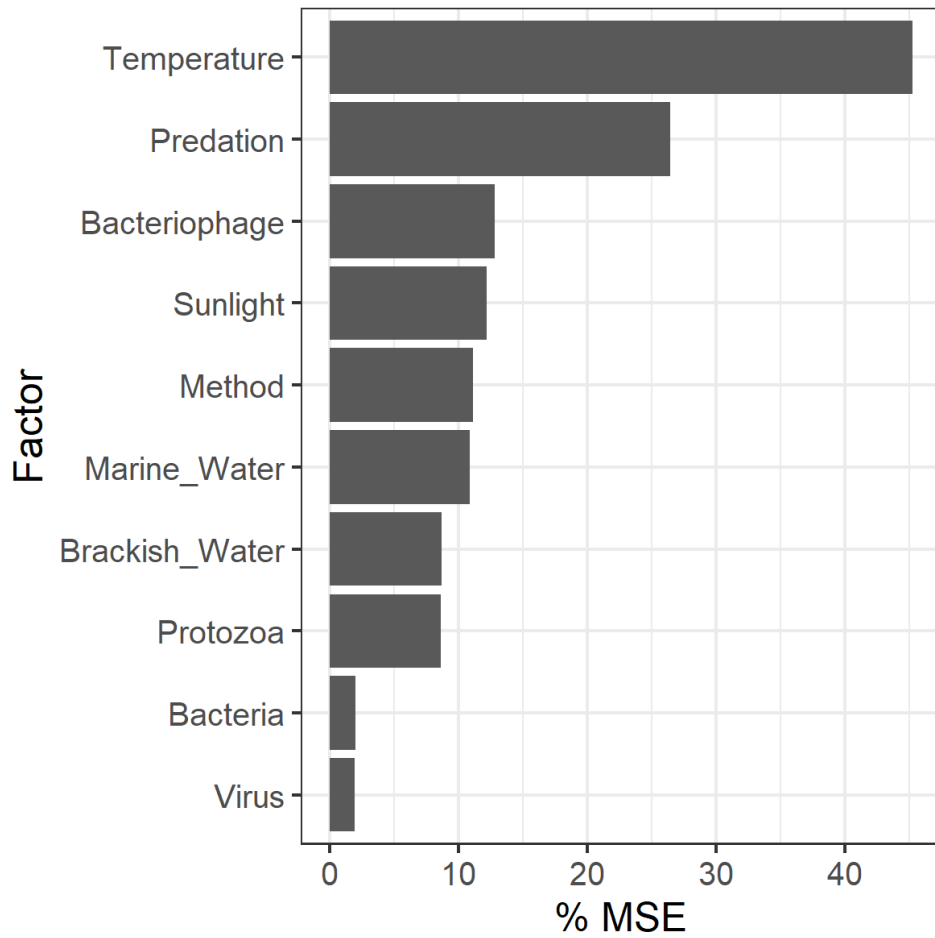


Figure D3.5: Variable importance for the random forest fit to the exponential-predicted T90s and core factor training dataset ranked by %MSE (percent change in mean square error when a factor is not included in the analysis)



Table D3.7: H-Statistics for Core Factor Random Forests fit the First-Order Decay Kinetics Predicted T90s and T99s

General Interactions*						
Factor	T90			T99		
	Average	Lower Bound	Upper Bound	Average	Lower Bound	Upper Bound
Temperature	0.61	0.59	0.64	0.56	0.54	0.57
Sunlight	0.16	0.14	0.18	0.32	0.29	0.34
Predation	0.60	0.57	0.62	0.4	0.38	0.42
Marine Water	0.16	0.14	0.18	0.18	0.17	0.2
Brackish Water	0.09	0.08	0.11	0.11	0.1	0.12
Bacteria	0.19	0.17	0.20	0.31	0.29	0.32
Virus	0.12	0.10	0.14	0.18	0.16	0.21
Bacteriophage	0.24	0.22	0.27	0.26	0.24	0.29
Protozoa	0.08	0.05	0.10	0.05	0.04	0.05
Method	0.10	0.08	0.11	0.21	0.19	0.23
Specific Two-Way Interactions with Temperature*						
Factor	T90			T99		
	Average	Lower Bound	Upper Bound	Average	Lower Bound	Upper Bound
Sunlight	0.07	0.06	0.07	0.29	0.26	0.31
Predation	0.77	0.72	0.82	0.45	0.44	0.46
Marine Water	0.05	0.05	0.06	0.17	0.15	0.18
Brackish Water	0.04	0.04	0.05	0.08	0.07	0.08
Bacteria	0.19	0.18	0.2	0.37	0.35	0.10
Virus	0.07	0.07	0.08	0.15	0.13	0.09
Bacteriophage	0.16	0.15	0.17	0.27	0.26	0.09
Protozoa	0.06	0.04	0.08	0.03	0.03	0.02
Method	0.05	0.04	0.05	0.13	0.12	0.08
Specific Two-Way Interactions with Predation*						
Factor	T90			T99		
	Average	Lower Bound	Upper Bound	Average	Lower Bound	Upper Bound
Temperature	0.81	0.76	0.86	0.46	0.45	0.47
Sunlight	0.18	0.17	0.19	0.27	0.26	0.29
Marine Water	0.07	0.07	0.07	0.08	0.07	0.08
Brackish Water	0.04	0.03	0.04	0.04	0.04	0.05
Bacteria	0.05	0.05	0.05	0.32	0.32	0.33
Virus	0.03	0.02	0.03	0.03	0.03	0.03
Bacteriophage	0.06	0.06	0.07	0.14	0.14	0.14
Protozoa	0.06	0.05	0.08	0.02	0.02	0.03
Method	0.04	0.04	0.04	0.06	0.06	0.06

\*H-statistics averaged over 100 iterations

## CHAPTER 4: TESTING A GENERAL MODEL FOR PATHOGEN PERSISTENCE IN SURFACE WATERS

### 4.1 Introduction

The conventionally used first-order decay model has a single parameter ( $k$ ) that describes the constant rate of decay over time. Recent work has indicated that the two-parameter model, Juneja and Marks (JM2), is better able to describe the persistence of pathogens and indicators in surface waters (Dean & Mitchell, 2022a) than first order decay kinetics. With JM2 providing a better fit to the majority of persistence data for microbial targets in surface waters in the peer-reviewed literature, it can be inferred that: i) microbial targets do not typically decay at a constant rate in the environment; and ii) there are common nonlinear decay dynamics across pathogen and indicator species. The  $k_1$  and  $k_2$  parameters of the JM2 model (Eq. 1) dictate nonlinear shape of the curve for log-reductions,  $\log_{10}\left(\frac{N_t}{N_0}\right)$ , over time,  $t$ . If  $k_2$  is less than 1, the curve is convex in shape and the fastest rate of decay occurs immediately with decay tapering off over time. If  $k_2$  is greater than 1, there is a point of inflection where the curve shifts from concave to convex; the concave portion of the curve represents a period of minimal or no measurable decay (a shoulder), and the fastest rate of decay occurs at the point of inflection. It is evident from prior analyses (Dean & Mitchell, 2022a) that the ability for the JM2 model to account for more dynamic behaviors is an improvement upon assumptions of constant decay, however the frequency and magnitude of tapering rates and shoulders across target groups has yet to be explored.

$$\log_{10}\left(\frac{N_t}{N_0}\right) = \log_{10}\left(\frac{1}{1+e^{k_1+k_2 \ln(t)}}\right) \quad \text{Eq. 1}$$

It was hypothesized in this study that the persistence behaviors of indicators and pathogens are inter-related and that differences between target groups can be quantified using the

JM2 model form, Bayesian hierarchical modeling methods, and a comprehensive dataset representing the state of the science mined from the peer-reviewed literature. Quantifying the differences between target group persistence is an important goal, as it is well established that the reliance on the indicator-pathogen paradigm for surface water management is a prominent source of uncertainty (Harwood et al., 2005; Korajkic et al., 2018). The use of the JM2 model has narrowed the uncertainty associated with both pathogen-specific and setting-specific persistence modeling (Dean & Mitchell, 2022a). Therefore, a logical next step was to determine how much knowledge the model may add to predictions of pathogen persistence in surface waters more broadly. A similar approach was used in Oishi et al. (2020) to develop a universal inactivation model for pathogens in stored excreta, however the dependent variable was log-reduction values, whereas this study uses the parameters of the JM2 model to facilitate the evaluation of the occurrence of persistence behaviors such as shoulders and tapering rates.

These efforts led to the development of a novel general model that estimates the typical persistence behaviors of indicators and pathogens in highly varied surface water conditions. Although a general model is not expected to replace the need for pathogen and site-specific persistence data and models, the general model construction leveraged the knowledge of over 400 persistence experiments to provide general information for the consideration of time within surface water decision making processes relying on only indicator or minimal monitoring data. This work assesses the feasibility of a general model for pathogen and indicator persistence in surface waters and uses the general model to make inferences about typical persistence behaviors and the differences between fecal indicator bacteria (FIB), bacteriophages, bacteria, viruses, and protozoa more broadly.

## 4.2 Methods

### 4.2.1 Model Components

A previously completed systematic literature review (Dean and Mitchell, 2022b) identified persistence datasets in the literature for common indicator organisms (FIB and bacteriophages) and pathogens of interest (bacteria, viruses, and protozoa). A meta-analysis of the mined datasets found that the JM2 model provided a good fit to 458 datasets in which at least one-log reduction was observed (Dean & Mitchell, 2022a). Of note, the 458 datasets are composed of 201 fecal indicator bacteria (FIB), 139 pathogenic bacteria, 77 bacteriophage, 32 virus, and 9 protozoa datasets, and the persistence experiments were conducted in freshwater, brackish water, and marine water, at temperatures ranging from 4°C to 37°C, in the presence and absence of sunlight, in the presence and absence of predation, and targets were enumerated with molecular-based and culture-based methods. A factorial analysis to better understand the influence of these factors on the observed decay parameters was reported in Dean & Mitchell, 2022b.

The water quality factors driving persistence behaviors were identified as temperature, predation, and water type (Dean & Mitchell, 2022a). Thus, the independent variables included within the general model are temperature (continuous), predation (binary), and water type (binary). Water type in this analysis is presented as a binary variable with a “0” status representing freshwater conditions and a “1” status representing not freshwater conditions, as the prior meta-analysis identified significant differences between fresh and brackish waters and fresh and marine waters, but not between brackish and marine waters. The estimates of the JM2 parameters for each of the 458 datasets by Likelihood Estimation (MLE) were used as the dependent variable for the general model developed herein. Kruskal-Wallis tests and Spearman

correlation coefficients were completed to evaluate the effect of the independent variables on each individual JM2 parameter directly. Factors that significantly influenced ( $p < 0.05$ ) the parameter values were selected as independent variables in the generalizable model.

#### 4.2.2 Model Form

The general model was developed in Rstan (Stan Development Team, 2021). Under the Bayesian framework, the best fitting (MLE) parameters from each dataset are assumed to be derived from a common probability distribution (a hyperdistribution) describing the plausible range of values for each parameter. The hyperdistributions for  $k_1$  and  $k_2$  each contain *hyperparameters*. As  $k_1$  is an unbounded continuous variable, a normal distribution is the simplest assumption for its hyperdistribution as shown in Equation 2, with hyperparameters  $\mu_{k1}$ , the mean  $k_1$  parameter, and  $\sigma_{k1}$ , the standard deviation of the  $k_1$  parameter. The  $k_2$  parameter is a positive real number, and thus can be log-transformed and also described with a normal distribution as shown in Equation 3.

$$k_{1h} \sim N(\mu_{k1}, \sigma_{k1}) \quad \text{Eq. 2}$$

$$\log(k_{2h}) \sim N(\mu_{\log(k2)}, \sigma_{\log(k2)}) \quad \text{Eq. 3}$$

The general model takes the form of a varying-intercept linear model (Figure 4.1), with the intercepts being grouped by target type as described in Equation 4-5 where the parameters for each dataset,  $i$ , are determined by an intercept for each target type  $j$  ( $\alpha_j$ ) and regression coefficients ( $\beta_k$ ) for each independent variable  $k$  observed in the data  $x_i$ .

$$\mu_{k1} = \alpha_{1j} + \beta_{1k} x_i \quad \text{Eq. 4}$$

$$\mu_{\log(k2)} = \alpha_{2j} + \beta_{2k} x_i \quad \text{Eq. 5}$$

The target type specific intercepts are drawn from normal distributions of the population intercepts described by mean values of  $\alpha_1$  and  $\alpha_2$  and standard deviations of  $\sigma_{\alpha_1}$  and  $\sigma_{\alpha_2}$  as shown in Equations 6-7.

$$\alpha_{1j} \sim N(\alpha_1, \sigma_{\alpha_1}) \quad \text{Eq. 6}$$

$$\alpha_{2j} \sim N(\alpha_2, \sigma_{\alpha_2}) \quad \text{Eq. 7}$$

The two model parameters of the JM2 model have been shown to be correlated but still possible to estimate independently (Dean & Mitchell, 2022c). Thus, the model described by Equations 2-7 assume that  $k_1$  and  $k_2$  are not correlated. Several steps were taken to test the impact of the aforementioned model assumptions prior to the selection of a final model form: (1) Other plausible probability distribution types in addition to the normal distribution, for the hyperdistributions of  $k_1$  and  $k_2$  were tested in the event of poor characterization by the normal distributions; (2) A varying-intercept and varying-slope model grouped by target was evaluated along with a non-linear transformation of the temperature variable in-line with known relationships (Dean & Mitchell, 2022a; Dean & Mitchell, 2022b).

The structure of the multilevel hierarchical model described is shown in Figure 4.1, where  $y_1$  to  $y_n$  describe the individual datasets ( $n=458$ ), and each dataset's best fitting parameters are  $k_{1-y_1}$ ,  $k_{2-y_1}$ , to  $k_{1-y_n}$ ,  $k_{2-y_n}$ . In a basic linear regression, the MLE parameters for each dataset (Dean & Mitchell, 2022b) were predicted with intercept values drawn from target type-level distributions (mean values of  $\alpha_{k1-FIB}$ ,  $\alpha_{k2-FIB}$ , etc.) that are inter-related via hyperdistributions for each parameter, governed by the hyperparameters  $\mu_{k1}$ ,  $\sigma_{k1}$ ,  $\mu_{k2}$ ,  $\sigma_{k2}$ . The basic linear regressions for the MLE parameters for each dataset also include population-level coefficients for temperature, predation, and water type variables ( $\beta_{Temperature}$ ,  $\beta_{Predation}$ ,  $\beta_{Water Type}$ ).

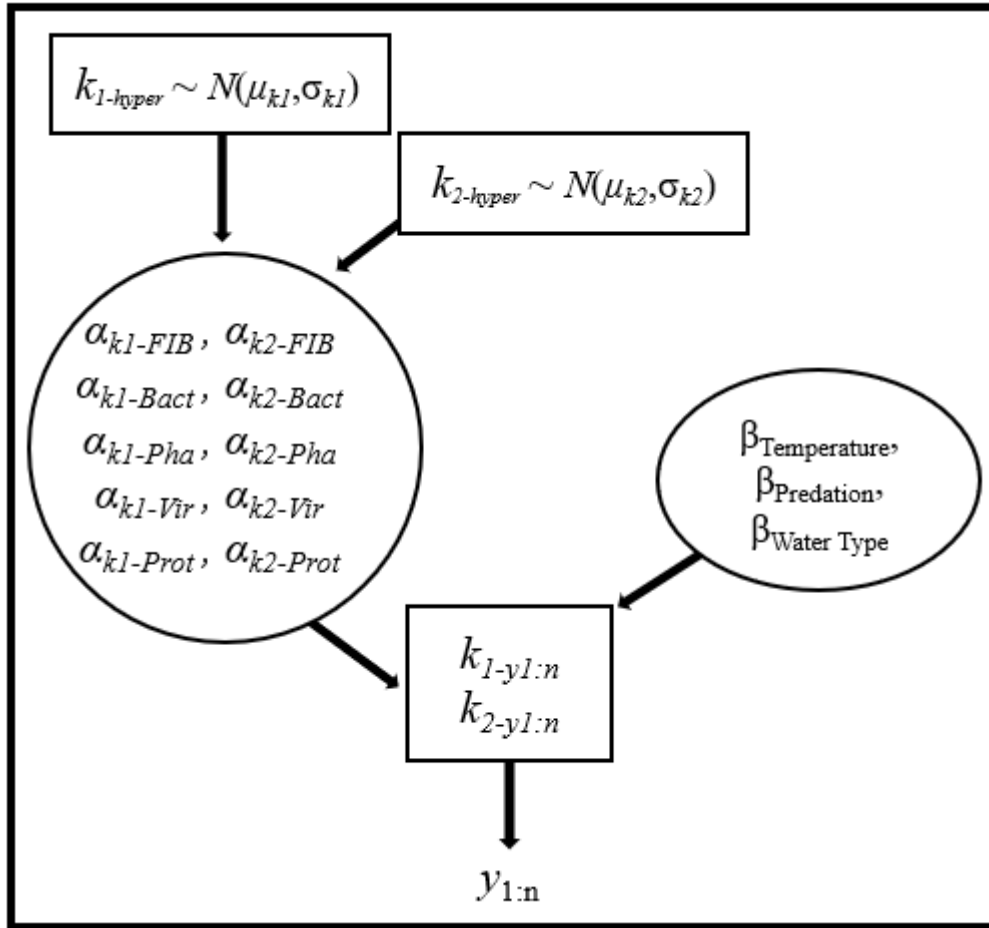


Figure 4.1: Structure of the hierarchical model framework

#### 4.2.3 Priors

In prior, published literature, the JM2 model was fit to pathogen and indicator persistence data in groundwater (Dean et al., 2020) and other various water matrices (Kline et al., 2022; Mitchell & Akram, 2017). Therefore, the range of best fitting  $k_1$  and  $k_2$  parameter values for each of these datasets (Appendix A) were used to construct weakly informative priors for the population-level intercepts of each parameter - shown in Equations 8-9. The weakly informative priors for the hyperdistribution scale and sigma parameters are shown in Equations 10-11. Uninformative priors were selected for the regression coefficient values, and the population intercept standard deviation values as shown in Equations 12-14

$$\alpha_1 \sim N(0,8) \quad \text{Eq. 8}$$

$$\alpha_2 \sim N(0.5,2) \quad \text{Eq. 9}$$

$$\sigma_{k_1} \sim \gamma(1,0.5) \quad \text{Eq. 10}$$

$$\sigma_{\log(k_2)} \sim \gamma(5,2) \quad \text{Eq. 11}$$

$$\sigma_{\alpha_1} \sim \gamma(5,1) \quad \text{Eq. 12}$$

$$\sigma_{\alpha_2} \sim \gamma(5,5) \quad \text{Eq. 13}$$

$$\beta_k \sim N(0,2) \quad \text{Eq. 14}$$

The influence of the weakly informative prior distributions on the posterior distributions were explored with a sensitivity analysis. The initial priors (Equations 9-12) were considered the base condition. Three alternative priors were selected including an informative prior ( $\alpha_1 \sim N(0,4)$ ,  $\alpha_2 \sim N(0.5,1)$ ,  $\sigma_{k_1} \sim \gamma(1,1)$ ,  $\sigma_{\log(k_2)} \sim \gamma(5,4)$ ), another weakly informative prior ( $\alpha_1 \sim N(0,16)$ ,  $\alpha_2 \sim N(0.5,4)$ ,  $\sigma_{k_1} \sim \gamma(1,0.25)$ ,  $\sigma_{\log(k_2)} \sim \gamma(5,1)$ ), and an uninformative prior ( $\alpha_1 \sim N(0,25)$ ,  $\alpha_2 \sim N(0.5,10)$ ,  $\sigma_{\log(k_1)} \sim \gamma(1,0.1)$ ,  $\sigma_{\log(k_2)} \sim \gamma(5,0.5)$ ). The hierarchical model was run in Rstan with the four priors and changes in the percent bias of selected metrics (median, 2.5%, 97.5%) from the base prior condition were calculated and compared.

#### 4.2.4 Model Fitting and Evaluation

Rstan was run with two chains, 3000 iterations, and a 1000 iteration burn-in. Convergence was evaluated with Rhat values ( $< 1.1$ ) and the effective sample size for each parameter ( $N_{\text{eff}}/N > 0.001$ ). Graphical posterior predictive checking was used to compare observed data to simulated data and to evaluate the appropriateness of the selected distributions. Kolmogorov-Smirnoff tests were completed to compare the distributions of observed and



simulated data. Root mean square error (RMSE) values were calculated for the observed  $k_1$  and  $k_2$  parameters and the simulated median  $k_1$  and  $k_2$  parameters. RMSE values were also calculated for the T90 values predicted with the observed and simulated median parameter estimates.

Alternative model forms were tested to evaluate the impact of various assumptions on the performance of the general model (Section 4.2.2). Leave-one-out cross validation with the **loo()** package was completed to compare the various models fit to the same data. The performance of the final selected model forms was further evaluated by comparing each model's predictive power on a testing dataset. The testing data was composed of 106 datasets extracted from the literature in which a 1-log reduction was observed but the JM2 model did not provide a good fit to the data (Dean and Mitchell 2022a; Dean and Mitchell, 2022b). The testing data did not have  $k_1$  and  $k_2$  parameters for direct evaluation of the model, and thus T90s general fit of the model to the unseen testing data was evaluated visually to identify any clear misrepresentations of the data.

#### 4.2.5 Uncertainty Factor Quantification

The median parameter estimates for each target group were used to calculate uncertainty factors for value of interest for surface water managers and decision makers. The two-parameters of the JM2 model ( $k_1$  and  $k_2$ ) are modified rate and shape parameters from the log-logistic probability distribution, which in certain forms can be used to describe more mechanistic features of the persistence profile. The JM2 function and  $k_1$  and  $k_2$  can be used to describe two of the commonly observed decay dynamics: the rate of tapering of the decay rate (tailing in the curve), and the length of time where initially minimal to no decay is observed (a shoulder period). As JM2 is based on the log-logistic probability distribution, the underlying assumption is that the rate of decay is tapering off at a constant rate when considering a log-timescale. The

rate of tapering ( $r_t$ ) can thus be described with Equation 1 where change in  $\Delta t$  represents the change in time in days between datapoints.

$$r_t = \frac{\log_{10}([\Delta t]^{k_2})}{\ln(\Delta t)} \quad \text{Eq. 14}$$

In the original derivation of the equation (Carlier et al., 1996), specified that if  $k_2$  was greater than 1, there was a point of inflection in the function where the concave curve shifted to a convex shape and the tapering of the decay rate began. If  $k_2$  was equal to or less than 1, there was no point of inflection, and the curve was convex in shape only. With these known constructs, the point of inflection can be considered a surrogate for the length of a shoulder period,  $S$ , in the persistence profile as shown in Equation 15.

$$S = \begin{cases} \left[ \frac{k_2 - 1}{e^{k_1}} \right]^{1/k_2}, & k_2 > 1 \\ 0, & k_2 \leq 1 \end{cases} \quad \text{Eq. 15}$$

Uncertainty factors were calculated for the rate of tapering, the length of the shoulder, and the time for 1-log (T90) and 2-log (T99) reductions to occur using the median parameter values for each target group.

## 4.3 Results

### 4.3.1 Model Comparison and Selection

Kruskal-Wallis tests for the categorical independent variables (predation, water type); grouping factor (target type); and the dependent variables ( $k_1$  and  $k_2$ ) indicated that the  $k_1$  parameter was significantly influenced by predation, water type and target type ( $p < 0.05$ ). The  $k_2$  parameter was only significantly influenced by water type and target type ( $p < 0.05$ ). Temperature significantly influences  $k_1$  ( $p < 0.05$ ) but not  $k_2$ , as determined by Spearman correlation coefficients. In the linear regressions considered the base model form for estimating  $k_1$  and  $k_2$ ,  $k_1$

was dependent upon three variables (temperature, predation, and water type), and  $k_2$  was dependent upon one variable (water type).

The initial selection of the normal hyperdistribution for the  $k_1$  and  $\log(k_2)$  parameter resulted in a poor fit to the parameter  $k_1$  data. This was visually evident with posterior predictive checks (Figure 4.2a) and Kolmogorov-Smirnov tests which indicated that it was highly unlikely that the observed  $k_1$  data and the simulated  $k_1$  values with the normal hyperdistribution were drawn from the same distribution ( $p \ll 0.05$ ). As evident in Figure 4.2b, the normal hyperdistribution was found to well-describe the log-transformed  $k_2$  parameters, with an average Kolmogorov-Smirnov test p-value of 0.33 for the 4000 replicates of the  $k_2$  data.

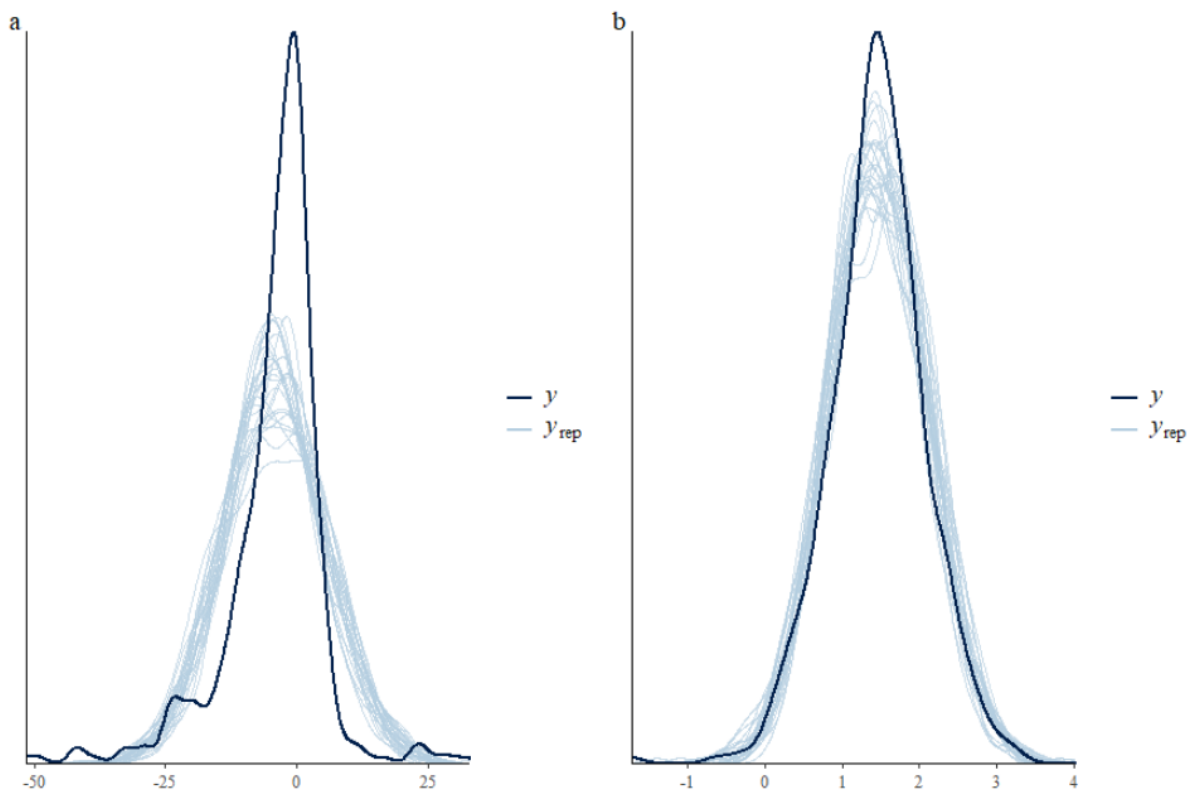


Figure 4.2: Comparison of the observed ( $y$ ) and predicted ( $y_{\text{rep}}$ ) density of the a) normally distributed  $k_1$  parameters, and b) normally distributed  $\log(k_2)$  parameters

As shown in Figure 4.2, the distribution of simulated  $k_I$  values under the normal hyperdistribution was more symmetric and wider than that of the observed data. Thus, the hyperdistribution of the  $k_I$  parameter was next assumed to be a double exponential (Laplace) distribution. The double exponential distribution is a continuous distribution with heavier tails than the normal distribution, and it is described with location ( $\mu$ ) and scale ( $\sigma$ ) parameters. The expected value of the distribution is  $\mu$  and the variance is  $2\sigma^2$  (Gelman et al., 2014). The weakly informative priors and the range of priors tested in the sensitivity analysis described in Section 4.2.3 for the  $k_I$  hyperdistribution are for the double exponential location parameter  $\mu_{k_I}$  (Eq. 8) and the scale parameter  $\sigma_{k_I}$  (Eq. 10).

Figure 4.3a compares the simulated  $k_I$  values under the double exponential hyperdistribution to the observed values. Visually, the double exponential distribution was more representative of the observed  $k_I$  values, with the frequency of the values between -20 and 10 being more accurately represented than when normally distributed (Figure 4.2a). The posterior distribution for  $k_I$  however, still did not capture the increase in frequency of values from -25 to -20 that is evident in the observed data, and a higher occurrence of values from 10 to 20 is being predicted than what was observed. The Kolmogorov-Smirnov test indicated that the double exponential model was an improvement from the normal, but that it was still unlikely that the observed and simulated  $k_I$  values were drawn from the same distribution ( $p=0.03$ ). An evaluation of selected test statistics (median, 2.5% estimate, 97.5% estimate) indicated that the distribution differences were being driven by the upper and lower range of simulated  $k_I$  values, as the simulated median  $k_I$  values were well-estimated ( $p>0.05$ ).

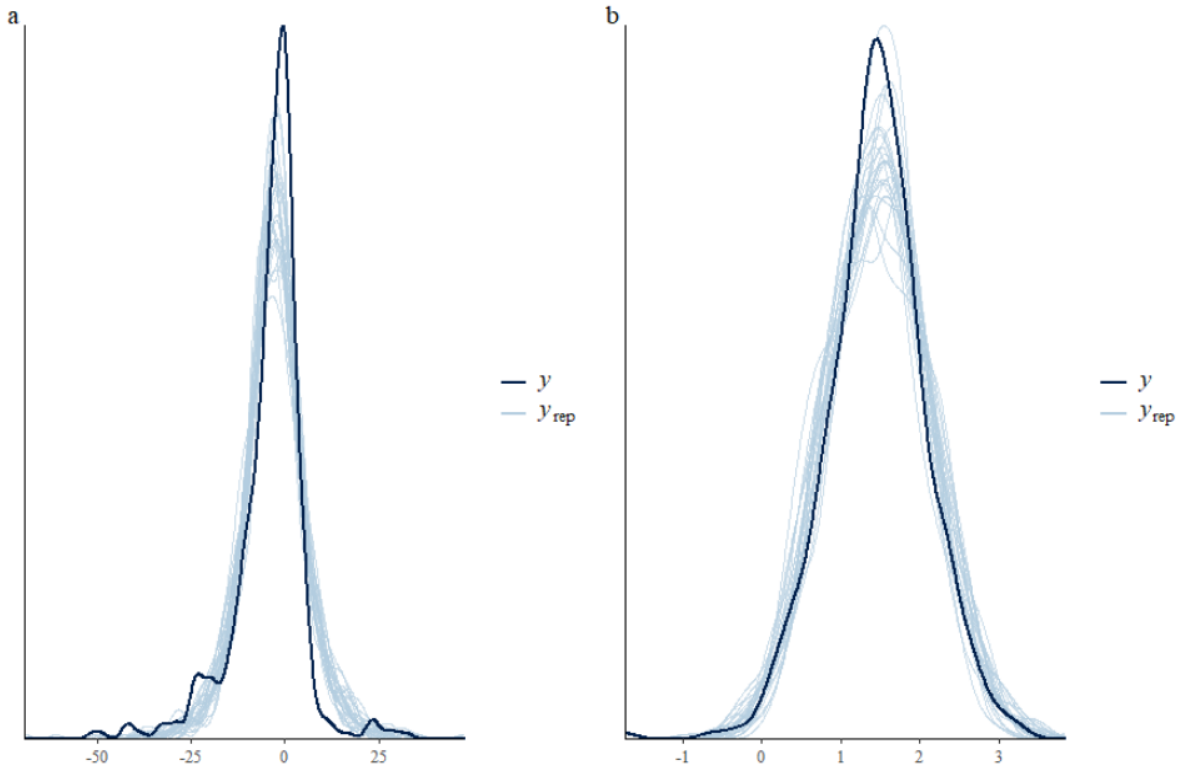


Figure 4.3: Comparison of the observed ( $y$ ) and predicted ( $y_{\text{rep}}$ ) density of the a) double exponentially distributed  $k_1$  parameters, and b) normally distributed  $\log(k_2)$  parameters

As the median  $k_1$  parameters were well-represented with double exponential hyperdistribution, the base model form assumed the hyperdistribution of  $k_1$  was double exponentially distributed and the hyperdistribution for the log-transformed  $k_2$  was normally distributed. The optimized population and target-level intercepts and coefficients for the base model are shown in Table 4.1. The  $k_1$  population-level intercept values were found to be normally distributed with a  $\mu$  of approximately -9.8 and a  $\sigma$  of 3.6. Figure 4.4 illustrates the distributions of each target types  $k_1$  intercept value, and the targets with the least amount of data (viruses, protozoa) were associated with the greatest range in intercept values. As shown in Table 4.1, temperature and predation were found to significantly impact the estimate  $k_1$  parameters, whereas the coefficient for water type was not found to significantly differ from zero. The log-transformed  $k_2$  population-level intercept values were normally distributed with a  $\mu$  of 1.5 and a

$\sigma$  of 0.4. The central tendencies of the normally distributed log-transformed  $k_2$  intercepts were found to be similar across target types, ranging from 1.2 (viruses) to 1.6 (bacteria, bacteriophages). The coefficient for water type in the  $\log(k_2)$  varying-intercept model did not significantly differ from zero (Table 4.1).

Table 4.1: Summary Values of Posterior Distribution ( $N(\mu, \sigma)$ ) for Population and Target Group Parameter Values

Parameter	$\mu$	$\sigma$	95% Confidence Interval (0.025, 0.975)
$\alpha_1$ Population	-9.8	1.99	(-13.61, -5.57)
$\alpha_1$ FIB	-8.7	1.19	(-10.98, -6.44)
$\alpha_1$ Bacteria	-9.4	1.18	(-11.73, -7.14)
$\alpha_1$ Bacteriophage	-10.9	1.27	(-13.37, -8.37)
$\alpha_1$ Virus	-8.2	1.55	(-11.31, -5.22)
$\alpha_1$ Protozoa	-13.9	1.90	(-17.64, -10.06)
$\alpha_1\sigma$	3.6	1.47	(1.49, 7.11)
$\alpha_2$ Population	1.5	0.22	(0.98, 1.88)
$\alpha_2$ FIB	1.5	0.05	(1.37, 1.58)
$\alpha_2$ Bacteria	1.6	0.06	(1.45, 1.67)
$\alpha_2$ Bacteriophage	1.6	0.07	(1.43, 1.7)
$\alpha_2$ Virus	1.2	0.11	(0.99, 1.41)
$\alpha_2$ Protozoa	1.5	0.18	(1.13, 1.84)
$\alpha_2\sigma$	0.4	0.22	(0.14, 0.99)
$\beta_1$ temp	0.2	0.04	(0.14, 0.31)
$\beta_1$ pred	2.4	0.76	(0.88, 3.88)
$\beta_1$ water	1.1	0.68	(-0.26, 2.46)
$\beta_2$ water	-0.1	0.06	(-0.24, 0.01)
$\sigma_{k1}$	5.7	0.27	(5.17, 6.24)
$\sigma_{k2}$	0.6	0.02	(0.58, 0.66)

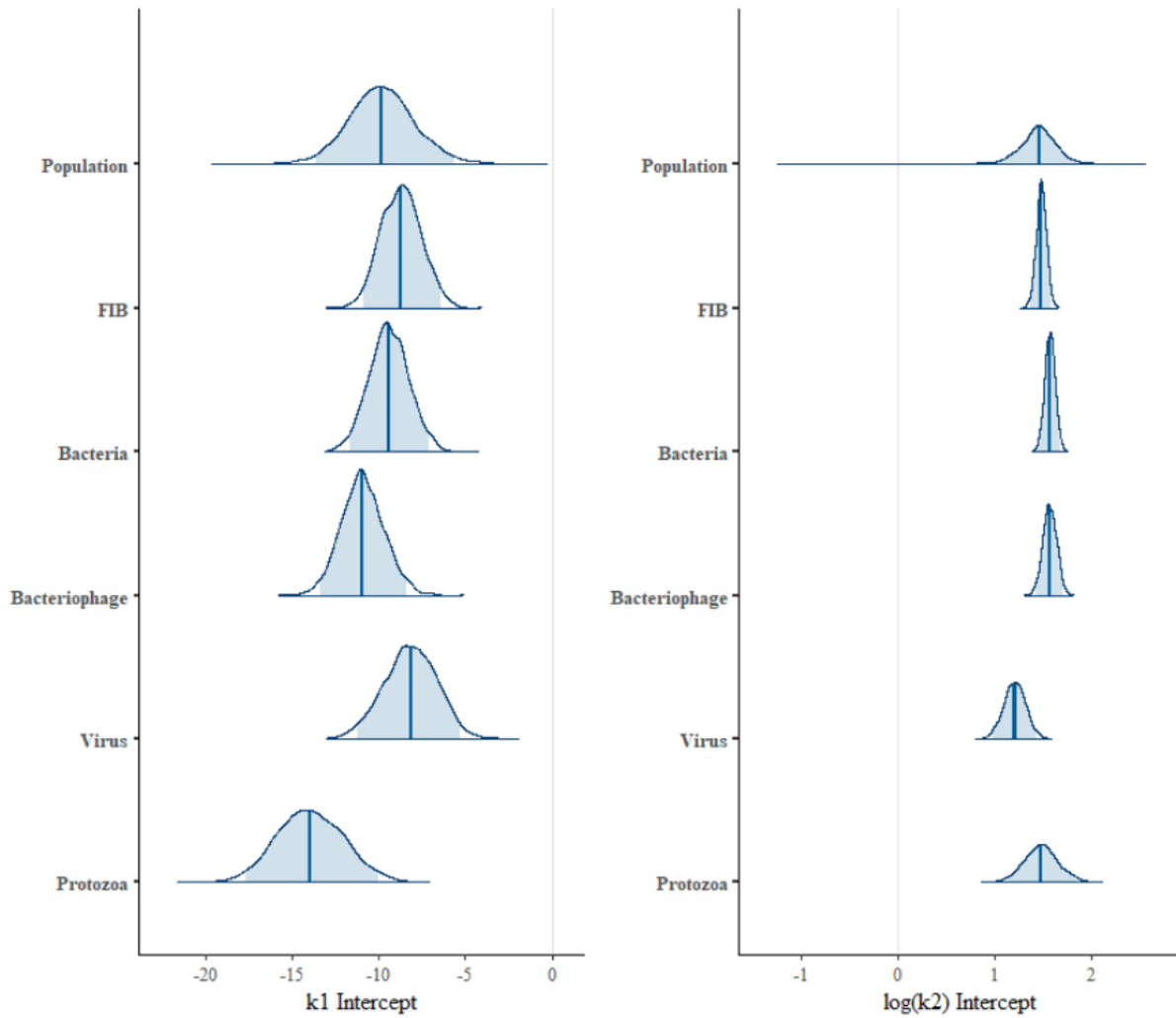


Figure 4.4: Population and target-type distributions of intercept values for a)  $k_1$  and b)  $\log(k_2)$

The application of the median parameter estimates for the population and target-level  $k_1$  and  $k_2$  estimates resulted in the median JM2 models shown in Figure 4.5 for a range of temperatures and the most relevant environmental conditions: freshwater and predation present. Note that the coefficients for water type were not significantly different from zero, and thus the median models for freshwater can be considered representative of all water types. As expected, more rapid decay is predicted for the highest temperature conditions (Figure 4.5c) compared to the lowest temperature (Figure 4.5a). The median model estimates predict the greatest

persistence of viruses and protozoa over time and shows FIB persistence conservatively estimating the persistence of pathogenic bacteria. The bacteriophage JM2 model predicts greater persistence than the FIB, but still does not capture the persistence of viruses or protozoa over time.

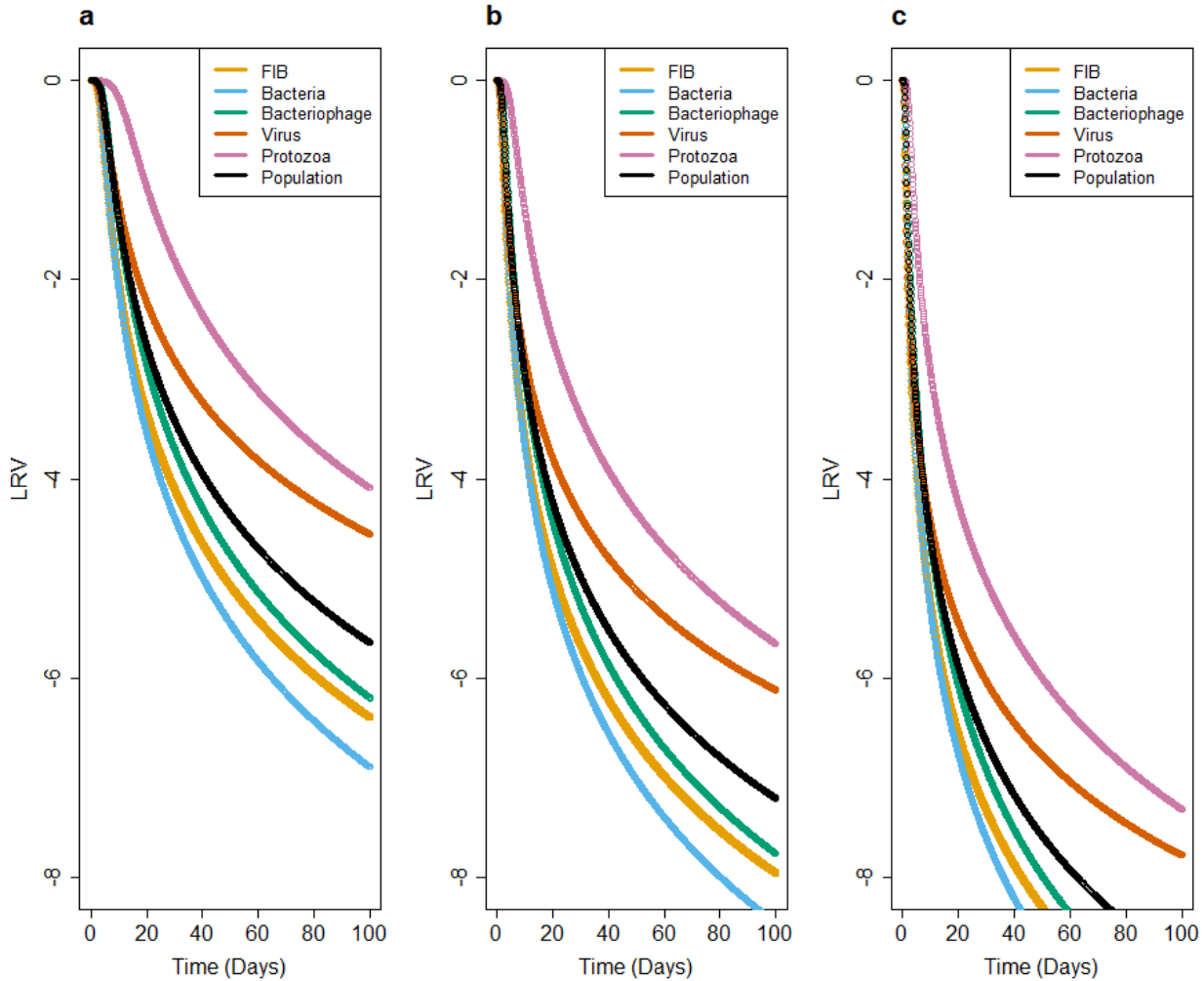


Figure 4.5: Median JM2 models for the most common conditions (freshwater, predation present) for temperatures of a) 4°C, b) 20°C, and 37°C

Alternative model forms (a. varying slope and varying intercept, b. varying intercept with non-linear transformation of temperature) were also evaluated. Leave-one-out cross validation determined that the base model (Eq. 4-5) provided the best fit to the data, and thus the base



model form presented thus far (Table 4.1, Figures 4.3-4.4) is considered the final model form. A sensitivity analysis of the priors used in the final model form did not indicate strong bias (average tendency for the simulated values to over or underestimate the base value) within the posterior distributions as the result of weakly informative priors that were selected. Changing the priors from informative to uninformative distributions resulted in at most a 3.2% change in the percent bias of the median estimate of  $k_1$  and a 1.2% change in the percent bias of the 95% confidence interval for  $k_1$ . The change in percent bias for the median and 95% confidence interval estimates of the log-transformed  $k_2$  were all less than 0.5%. Table A4.1 in Appendix A summarizes the changes in percent bias, and Figures A4.3-A4.4 visually demonstrates the minimal change in the resulting posterior distributions.

#### 4.3.2 Performance

The final model form was used to predict median, 2.5% and 97.5% estimates of T90 values for each of the training datasets, and the results are shown in Table 4.2. The median observed T90 fell within the 95% confidence interval of the predicted T90 values for 17 of the 18 tested conditions. The median observed bacteriophage T90 in non-freshwater conditions with predation absent had a median T90 greater than the upper bound of the predicted 95% confidence interval. In general, the upper bound of the observed T90 values was poorly approximated by the model for most conditions. This mischaracterization of the upper estimates of the observed T90 values resulted in high RMSE values for the training data predicted T90s, as shown in Figure 4.6.

Table 4.2: Training Dataset Observed and Predicted T90 Values with the General Model Form

Target	Conditions			Observed T90s (Days)		Predicted T90s (Days)	
	Temperature	Predation	Water Type	Median	2.5%-97.5%	Median	2.5%-97.5%
FIB	4-25	Absent	Fresh	5	2-82	3-10	2-18
	4-30	Present	Fresh	3	0-14	2-6	1-10
	21	Absent	Brackish/ Marine	5	5	4	2-7
	10-31.2	Present	Brackish/ Marine	2	0-6	1-4	1-7
Bacteria	4-37	Absent	Fresh	4	0-112	2-9	1-17
	4-37	Present	Fresh	4	0-26	1-6	1-10
	9.5-35	Absent	Brackish/ Marine	3	0-14	2-7	1-13
	9.5-30	Present	Brackish/ Marine	2	1-3	1-4	1-7
Bacteriophage	30	Absent	Fresh	6	5-30	4-4	2-7
	4-30	Present	Fresh	4	1-35	2-8	1-15
	30	Absent	Brackish/ Marine	14	7-20	3-3	2-7
	5-30	Present	Brackish/ Marine	3	1-6	2-7	1-16
Virus	4-37	Absent	Fresh	9	2-75	2-16	1-57
	4-30	Present	Fresh	8	0-31	1-8	1-26
	22	Absent	Brackish/ Marine	2	2	4	1-14
	7-37	Present	Brackish/ Marine	1	0-4	1-6	0-19
Protozoa	15-25	Absent	Freshwater	16	11-20	10-17	4-104
	4-25	Present	Freshwater	15	10-72	6-18	2-108

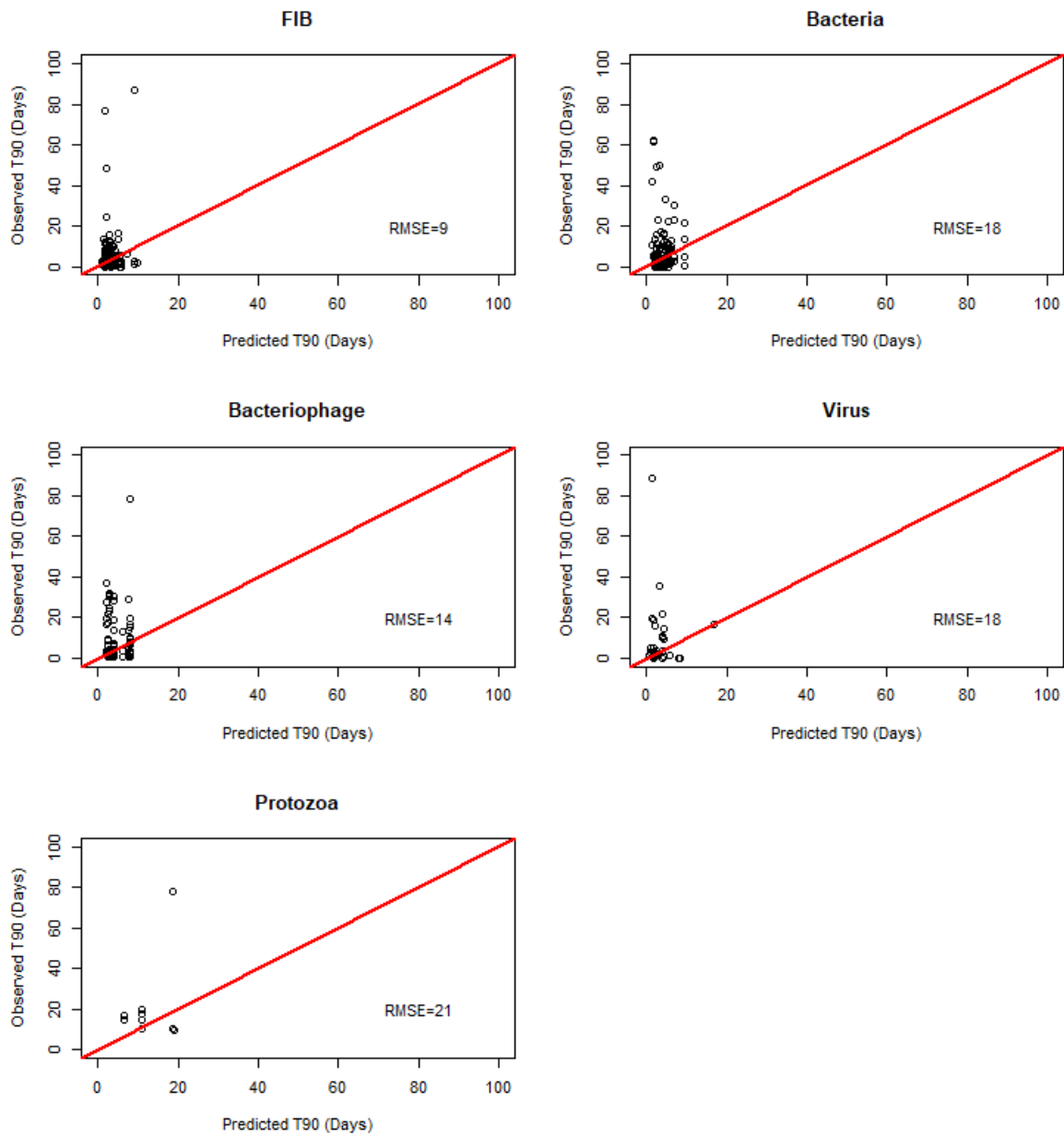


Figure 4.6: Comparison of the observed training T90s to the T90 values predicted with the median parameter estimates of the general model form; red line indicates ideal 1:1 ratio

The predictive power of the final model form was evaluated by fitting the general model to a testing dataset composed of 106 surface water persistence experiments. Table B4.1 compares the observed T90 values to the predicted T90 values, and the comparison is shown visually in Figure 4.7. In-line with the performance on the training data, the general model fails

to capture the upper estimates of the T90 values for each target, however the median T90 values were estimated within the 95% confidence interval of the model for 10/14 conditions (Table B4.1).

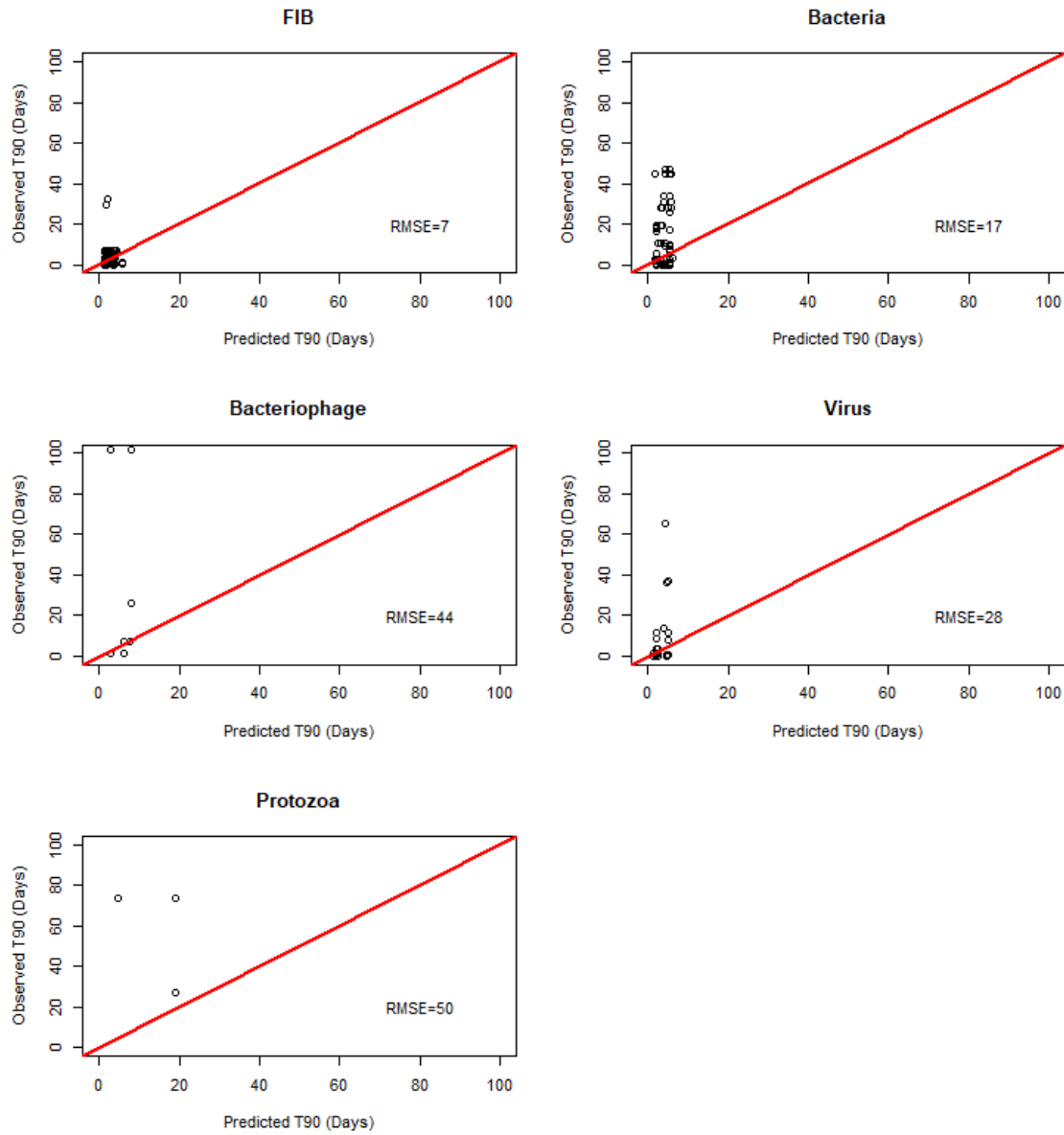


Figure 4.7: Comparison of the observed T90s in the testing data to the T90 values estimated with the median general model parameters; red line is ideal 1:1 ratio

### 4.3.3 Target Persistence Behaviors and Uncertainty Factors

The  $k_1$  and  $\log(k_2)$  intercept estimates for FIB were approximately -8.7 and 1.5, respectively. The viral targets had the largest  $k_1$  intercept (-8.2), followed by FIB (-8.7), bacteria (-9.4), bacteriophages (-10.9), and protozoa (-13.9). The bacteria and bacteriophage targets each had the largest intercept value for the  $\log(k_2)$  parameters (1.6), followed by FIB and protozoa (1.5), and then viruses (1.2). The median, 2.5%, and 97.5% parameter estimates were used to calculate estimates of the rate of tapering (Eq. 14), the length of a shoulder period (Eq. 15), T90s, and T99s for the most relevant environmental conditions (freshwater, predation present) at three temperatures (4°C, 20°C, 37°C).

Table 4.3 summarizes the ratio of bacteria, bacteriophage, virus, and protozoa estimates to FIB estimates for the different metrics of interest. The bacterial shoulders, T90s, and T99s minimally differ from the FIB estimates, however the bacteriophage, virus and protozoa metrics are on average 1.5-3.5x the length of the FIB estimates. Regardless of temperature, the protozoa targets have the longest median shoulders (3-15 days) and the rate of decay tapers off the most quickly for viruses (Figure 4.5). When the bacteriophage data are treated as the baseline, the virus and protozoa metrics are 1-3x the length of the bacteriophage estimates.

Table 4.3: Uncertainty Factors for Metrics of Interest

Temperature	Target	Ratio to FIB Length (95% CI)				
		Shoulder	T90	T99	T99.9	T99.99
4°C	FIB	-	-	-	-	-
	Bacteria	1.0 (1,1)	1.0 (1,1)	1.0 (1,1)	0.9 (1,1)	0.9 (1,1)
	Bacteriophage	1.4 (1,2)	1.4 (1,2)	1.3 (1,2)	1.3 (1,2)	1.2 (1,1)
	Virus	1.2 (1,3)	1.5 (1,4)	1.7 (1,5)	2.0 (1,7)	2.4 (1,9)
	Protozoa	3.3 (1,15)	3.4 (1,18)	3.4 (1,22)	3.4 (1,26)	3.4 (1,30)
20°C	FIB	-	-	-	-	-
	Bacteria	1.1 (1,1)	1.1 (1,1)	1 (1,1)	1 (1,1)	0.9 (1,1)
	Bacteriophage	1.6 (1,2)	1.5 (1,2)	1.4 (1,2)	1.4 (1,2)	1.3 (1,2)
	Virus	0.9 (1,2)	1.1 (1,2)	1.3 (1,3)	1.6 (1,4)	1.9 (1,6)
	Protozoa	3.3 (2,11)	3.3 (2,14)	3.3 (1,16)	3.3 (1,20)	3.4 (1,23)

Table 4.3 (cont'd)

37°C	FIB	-	-	-	-	-
	Bacteria	1.2 (1,2)	1.2 (1,1)	1.1 (1,1)	1.1 (1,1)	1 (1,1)
	Bacteriophage	1.7 (2,3)	1.6 (1,2)	1.5 (1,2)	1.5 (1,2)	1.4 (1,2)
	Virus	0.7 (1,1)	0.8 (1,2)	1.0 (1,2)	1.2 (1,3)	1.4 (1,4)
	Protozoa	3.3 (2,12)	3.3 (2,10)	3.3 (2,14)	3.3 (2,15)	3.4 (1,17)

#### 4.4 Discussion

The analysis presented herein: (1) identified plausible distributional forms for hyperdistributions of the JM2 model's  $k_1$  and  $k_2$  decay parameters; (2) evaluated the feasibility of a generalized model for indicator and pathogen persistence in surface waters; and (3) quantified the uncertainty between the persistence behaviors (shoulders) and commonly used decay metrics (T90s, T99s) for indicators and pathogens. A double exponential distribution and normal distribution were identified as the most appropriate distributions for the hyperdistributions of  $k_1$  and  $\log(k_2)$ , respectively. The double exponential distribution has heavier tails than the normal distribution and is able to accommodate a narrower, symmetric distribution than the normal distribution (Gelman et al., 2014). Although there is a large range of observed  $k_1$  values in previous studies (Figure A4.1), and in the data analyzed herein, there is a greater probability density associated with a smaller range of central values than is predicted with the normal distribution. The double exponential model was selected herein, however Kolmogorov-Smirnov tests and the evaluation of selected metrics suggested that the upper and lower bound estimates of the  $k_1$  parameter were poorly approximated. Visually, the  $k_1$  parameter data are almost bimodal in nature (Figure 4.3), something that should be explored in future work as there may be a driver of this behavior that can be used to improve the accuracy of the form of the hyperdistribution.

LOO methodology selected the varying-intercept linear regression as the optimal model form for the general persistence model. This suggests that the base distributions of parameter estimates are dictated by the target of interest (target-type intercepts), however the water quality

factors have the same magnitude of effect on each target type (population-level coefficients). The general model was used to predict T90 values of each dataset, and the 95% confidence interval of the model captured the median observed T90 for most conditions, the exception being the median T90 for bacteriophages in predation absent, non-freshwater conditions. As most in-situ conditions will consider predation present, the 95% confidence intervals of the general model herein are considered to accurately represent *median* persistence behavior for the indicator and pathogen groups of interest. The general model's 95% confidence intervals, however, do not capture the upper and lower bounds of the observed T90s, leading to high RMSE values for the training and testing data. The high RMSE values are not unexpected, as a general model is expected to provide general persistence behaviors and patterns but is not expected to attain the accuracy of pathogen and location-specific models previously fit for each dataset (Dean & Mitchell, 2022a). As the general model consistently under-predicts the upper bounds of the observed T90s, only the median data was used for uncertainty factor quantification.

The median parameter estimates for the general model were used to quantify uncertainty between the indicator and pathogen groups. As the model only characterizes median persistence behaviors, it is important to acknowledge that the uncertainty factors are representative of average population behaviors only. On average, the T90s and T99s of bacteria and FIB can be considered similar, as evident in Figure 4.5. Both FIB and bacteria persistence were associated with the shortest shoulders of the evaluated targets, with the median length of a shoulder in the persistence curve ranging from 1-5 days for waters of 4°C-37°C. The T90s and T99s of bacteriophages, viruses, and protozoa were approximately 1.5-3.5x the length of the FIB T90s and T99s. This difference is most evident at the lowest temperature range, where the average FIB T90 and T99 were 6 and 10 days, respectively, compared to 19 and 33 days for protozoa. There

is a clear shoulder presence in the protozoa persistence profile (Figure 4.5) which is partially driving this difference. Notably, the median JM2 model for bacteriophage persistence is shown to be a more conservative indicator of the virus persistence data in terms of T90s, T99s, and the length of a shoulder in the persistence curve (Table 4.3). However, it is evident in Figure 4.5 that the virus persistence data delineates from the bacteriophage persistence past the 2-log reduction point at 4°C and the 4-log reduction point at 37°C. The use of coliphages as an indicator is under consideration by the Environmental Protection Agency (EPA, 2015), and with regards to persistence behaviors, this study suggests that bacteriophages are a more conservative estimator of virus persistence than FIB, but there are still significant differences in behavior past the 2-to-4-log reduction time-point depending on water temperature.

The primary findings for surface water managers from this analysis are i) indicators and pathogens persist longer at colder temperatures, ii) the difference between indicator and pathogen persistence is more pronounced at colder temperatures, and iii) the decay of viruses and protozoa may be more than 3x slower than that of commonly monitored FIB. The uncertainty factors generated herein provide managers with additional data to incorporate into decision-making processes when considering elevated levels of FIB. This information is more powerful in cases when decision makers also have knowledge of the likely sources of pathogen contamination, as bacteria persistence behaviors delineate from FIB less than other pathogenic targets. The incorporation of these uncertainty factors with site-specific microbial source tracking data in the future will facilitate optimal decision-making for the use of impacted surface waters.

This study demonstrates the feasibility of developing a general model for characterizing pathogen persistence in surface waters in lieu of site-specific persistence data. Fitting a pathogen



and site-specific persistence model is preferred, when possible, to ensure the most accurate prediction of persistence over time for surface water quality modeling and risk assessment populations. The hierarchical model form evaluated herein, can also be used to improve the fitting of the JM2 model to future datasets, as the described distributions can be used to select more informative priors and ensure maximum utility of limited data.

## REFERENCES

- Carrier, V., Augustin, J.C., & Rozier, J. (1996). Heat resistance of *Listeria monocytogenes* (Phagovar 2389/2425/3274/2671/47/108/340): D- and z-values in ham. *Journal of Food Protection*, 59(6): 588-591.
- Dean, K., Wissler, A., Hernandez-Suarez, J.S., Nejadhashemi, A.P., & Mitchell, J. (2020). Modeling the Persistence of Viruses in Untreated Groundwater. *Science of the Total Environment*, 717(15). <https://doi.org/10.1016/j.scitotenv.2019.134599>
- Dean, K. & Mitchell, J. 2022a. Meta-analysis addressing the implications of model uncertainty in understanding the persistence of indicators and pathogens in natural surface waters. *Environmental Science & Technology*, 56 (17): 12106–12115.
- Dean, K. & Mitchell, J. 2022b. Identifying Water Quality and Environmental Factors that Influence Indicator and Pathogen Decay in Natural Surface Waters. *Water Research*. <https://doi.org/10.1016/j.watres.2022.118051>
- EPA. (2015). Review of coliphages as possible indicators of fecal contamination for ambient water quality. Retrieved from: [https://www.epa.gov/sites/default/files/201607/documents/review\\_of\\_coliphages\\_as\\_possible\\_indicators\\_of\\_fecal\\_contamination\\_for\\_ambient\\_water\\_quality.pdf](https://www.epa.gov/sites/default/files/201607/documents/review_of_coliphages_as_possible_indicators_of_fecal_contamination_for_ambient_water_quality.pdf)
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2014). *Bayesian Data Analysis* (Third). Boca Raton, Florida: Taylor & Francis Group.
- Juneja, V. K., & Marks, H. M. (2003). Mathematical description of non-linear survival curves of *Listeria monocytogenes* as determined in a beef gravy model system at 57.5 to 65 °C. *Innovative Food Science & Emerging Technologies*, 4(3), 307–317. [https://doi.org/10.1016/S1466-8564\(03\)00025-0](https://doi.org/10.1016/S1466-8564(03)00025-0)
- Harwood, V.J., Levine, A. D., Scott, T. M., Chivukula, V., Lukasik, J., Farrah, S. R., & Rose, J. B. (2005). Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. *Applied and Environmental Microbiology*, 71(6), 3163–3170. <https://doi.org/10.1128/AEM.71.6.3163-3170.2005>
- Korajkic, A., McMinn, B., & Harwood, V. (2018). Relationships between Microbial Indicators and Pathogens in Recreational Water Settings. *International Journal of Environmental Research and Public Health*, 15(12), 2842. <https://doi.org/10.3390/ijerph15122842>
- Kline, A., Dean, K., Kossik, A.L., Harrison, J.C., Januch, J.D., Beck, N.K., Zhou, N.A., Shirai, J.H., Boyle, D.S., Mitchell, J., & Meschke, J.S. 2022. Persistence of poliovirus types 2 and 3 in waste-impacted water and sediment. *PLOS One*, 17. <https://doi.org/10.1371/journal.pone.0262761>
- Mitchell, J. & Akram, S. 2017. “Pathogen Specific Persistence Modeling Data.” In: J.B. Rose and B. Jiménez-Cisneros, (eds) *Global Water Pathogens Project*, <http://www.waterpathogens.org> (M. Yates (eds) Part 4 Management of Risk from Excreta

and Wastewater) Accessible at: <http://www.waterpathogens.org/book/pathogen-specific-persistence-modeling-data>. Michigan State University, E. Lansing, MI, UNESCO.

Oishi, W., Kadoya, S., Nishimura, O., Rose, J.B., Sano, D. (2021). Hierarchical Bayesian modeling for predictive environmental microbiology toward a safe use of human excreta: Systematic review and meta-analysis. *Journal of Environmental Management*, 284. <https://doi.org/10.1016/j.jenvman.2021.112088>

Stan Development Team (2021). RStan: the R interface to Stan. R package version 2.21.3. <https://mc-stan.org/>

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P, Paananen T, Gelman A (2020). “loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models.” R package version 2.4.1, <https://mc-stan.org/loo/>

## APPENDIX A: PRIOR DISTRIBUTION SELECTION AND SENSITIVITY ANALYSIS

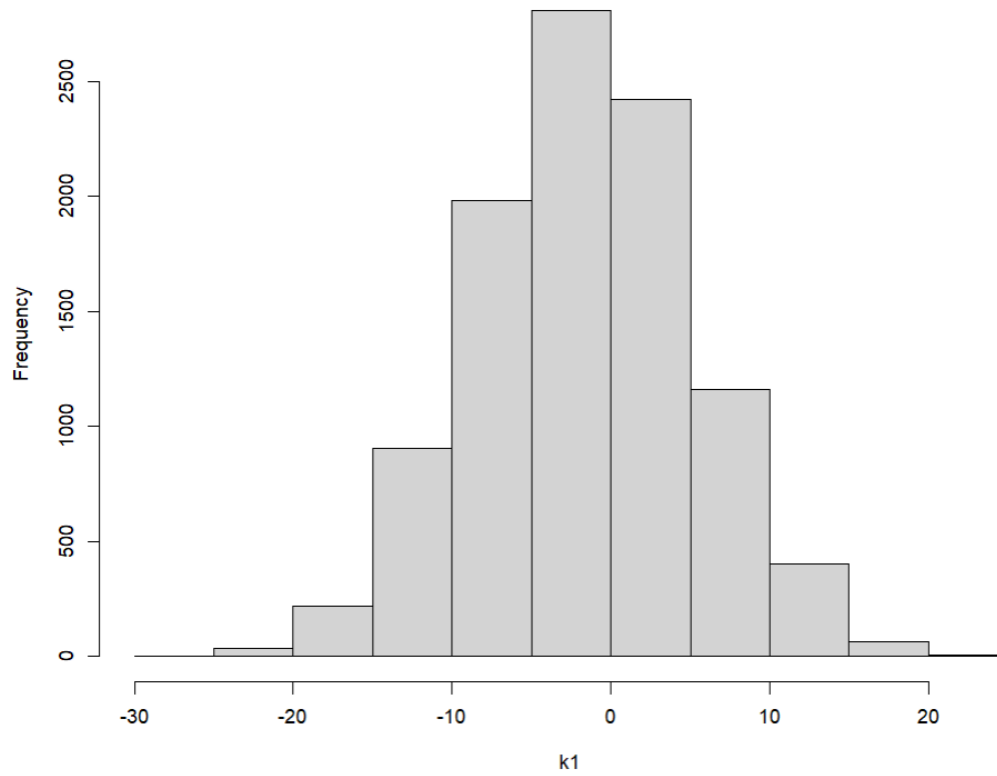


Figure A4.1: Distribution of Optimized  $k_1$  Parameter Values ( $n=55$ ) from Previous Persistence Modeling Efforts (Dean et al., 2020; Kline et al., 2022; Mitchell & Akram, 2017)

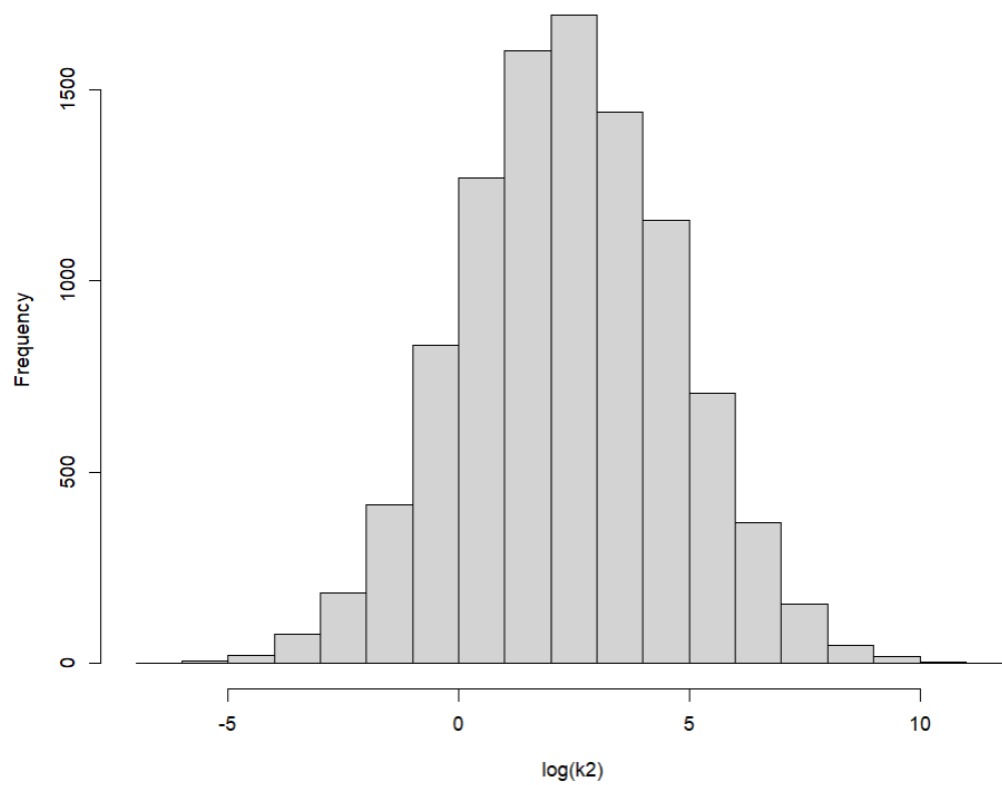


Figure A4.2: Distribution of Optimized  $\log(k_2)$  Parameter Values from Previous Persistence Modeling Efforts (Dean et al., 2020; Kline et al., 2022; Mitchell & Akram, 2017)

Table A4.1: Summary Statistics for MLE Estimates of JM2 Parameters from Previous Analyses (Dean et al., 2020; Kline et al., 2022; Mitchell & Akram, 2017)

	Minimum	Median	Mean	Maximum	Standard Deviation
$k_1$	-33.3	-1.6	-3.2	9.8	6.8
$\log(k_2)$	-0.82	0.86	0.88	2.48	0.67

Table A4.2: Change in Percent Bias from Baseline Prior Conditions with the Use of More and Less Informative Priors

Prior	Test Statistic	Change in Percent Bias (%)	
		k1	log(k2)
Base	Median	NA	NA
Narrower		-3.2	0
Wider		0	0.1
Uninformed		-0.4	0
Base	2.5 Percentile	NA	NA
Narrower		-0.8	0.2
Wider		0.2	-0.1
Uninformed		0.4	-0.1
Base	97.5 Percentile	NA	NA
Narrower		-1	-0.1
Wider		0.4	0.1
Uninformed		1.2	0

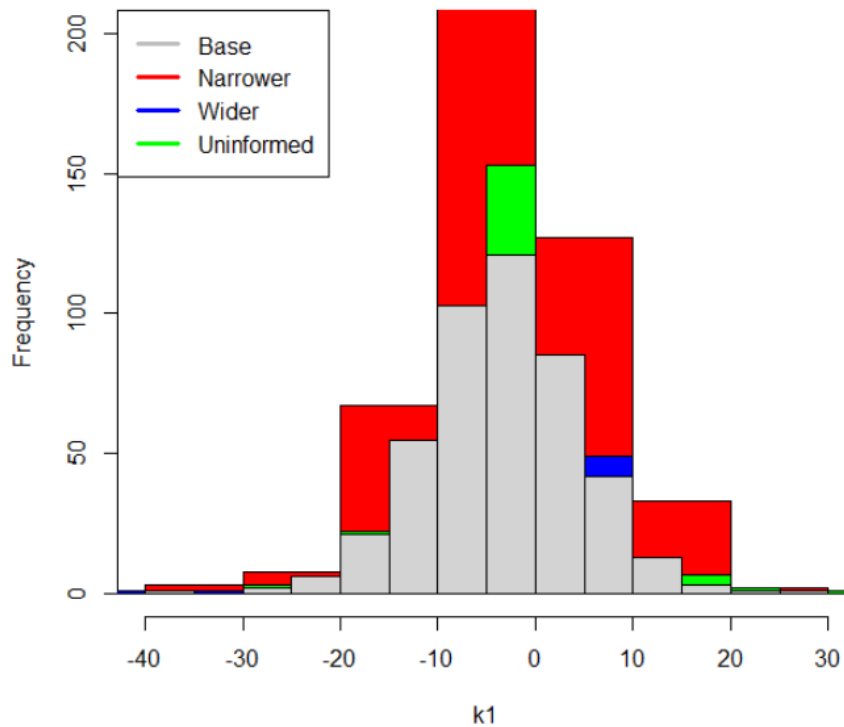


Figure A4.3: Visual comparison of simulated posterior distributions for k1 with baseline weakly informative priors (grey), informative priors (red), broader weakly informative priors (blue), and uninformative priors (green)

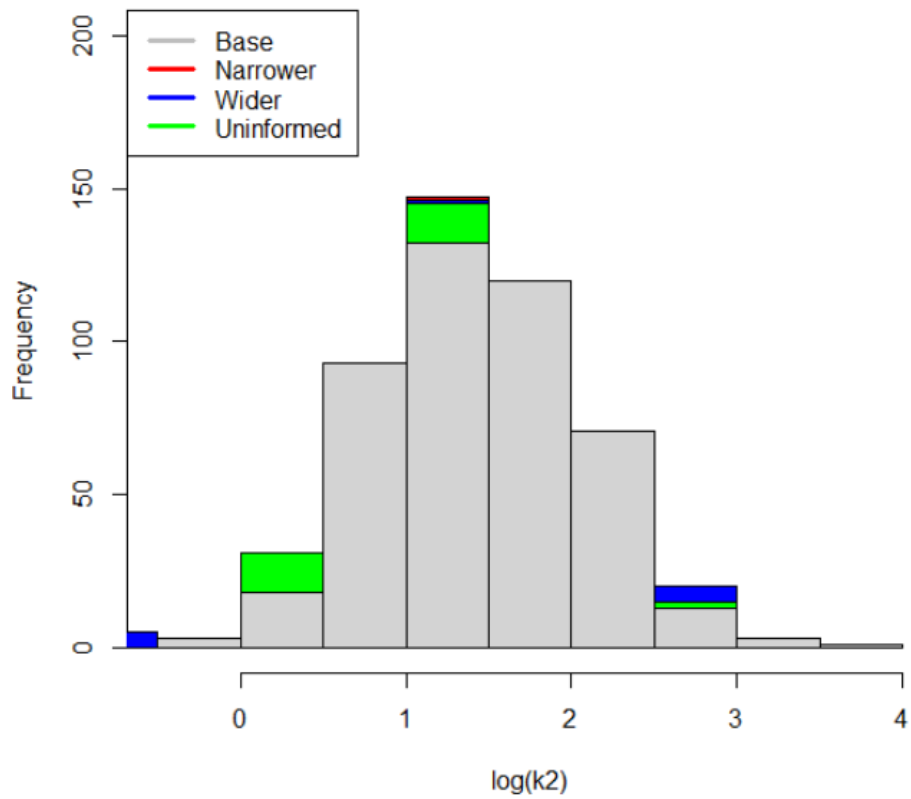


Figure A4.4: Visual comparison of simulated posterior distributions for log-transformed  $k_2$  with baseline weakly informative priors (grey), informative priors (red), broader weakly informative priors (blue), and uninformative priors (green)

## APPENDIX B: EVALUATION OF PREDICTIVE POWER

Table B4.1: Testing Dataset Observed and Predicted T90 Values with the General Model Form

Target	Conditions			Observed T90s (Days)		Predicted T90s (Days)	
	Temperature	Predation	Water Type	Median	2.5%-97.5%	Median	2.5%-97.5%
FIB	14.1-21	Absent	Fresh	0	0-3	4-6	2-11
	13-30	Present	Fresh	4	0-33	2-4	1-6
	21	Absent	Brackish/ Marine	2	2	4	2-7
	15.6-30	Present	Brackish/ Marine	2	0-7	1-3	1-5
Bacteria	13-37	Absent	Fresh	20	0-47	2-6	1-11
	4-30	Present	Fresh	6	0-10	2-6	1-10
	13-35	Absent	Brackish/ Marine	17	2-34	2-6	1-11
	22.7	Present	Brackish/ Marine	2	2	2	1-4
Bacteriophage	4-25	Present	Fresh	7	1-101	3-8	2-15
	5	Present	Brackish/ Marine	26	26	7	4-16
Virus	4-25	Absent	Fresh	4	1-111	4-16	2-57
	22-24	Present	Fresh	1	0-37	2-2	1-7
	10-26	Present	Brackish/ Marine	2	1-12	1-5	0-15
Protozoa	4.2	Present	Fresh	74	74	17	6-107



## **CHAPTER 5: APPLYING PERSISTENCE KNOWLEDGE WITHIN A QMRA CASE STUDY OF A SEWAGE SPILL EVENT**

### **5.1 Introduction**

Pathogen-contaminated surface waters can pose a risk to human health when the waters are used for recreation, as source water for drinking water, or for irrigation purposes. Quantitative microbial risk assessments (QMRA) are used to characterize the health risks associated with microbial contaminants in environmental matrices. QMRAs for fecal contamination in surface waters have commonly assumed minimal decay occurs after the point of introduction to conservatively estimate risk (Ahmed et al., 2018; Soller et al., 2010) or they have applied first order decay mechanics to calculate the change in concentrations over time (Boehm et al., 2018; U.S. EPA, 2010). The former assumptions can limit the information available for decision makers to understand the impact of relying on natural attenuation after a contamination event or the detection of elevated fecal contamination. As highlighted in previous work (Dean & Mitchell, 2022a; Dean & Mitchell, 2022b), the latter assumption is likely introducing model uncertainty into the risk assessment, as the persistence of indicators and pathogens are more dynamic than the constant rate of decay assumed with first-order kinetics.

The ingestion of contaminated surface water during a recreation event is the most direct exposure pathway. Accordingly, this QMRA case study will focus on recreational water ingestion exposure scenarios related to a sewage spill event in the Northeastern United States in 2019. The event resulted in the release of 100,000 gallons of untreated sewage into the environment, predominantly impacting an adjacent retention pond. The spill was contained quickly, and monitoring for levels of fecal indicator bacteria (FIB), MST markers, and pathogens of concern was conducted in the retention pond, stormwater channel, and an impacted tidal stream for 40 days after the containment. Water monitoring officials were informed that the

retention pond is commonly used for recreation within the area, and public health officials were involved to assist in communicating risks to the area post-spill.

A QMRA case study of the impacted retention pond is presented herein to (i) demonstrate the importance of characterizing persistence within surface water QMRAs; (ii) illustrate an application of the uncertainty factors developed to make inferences about pathogen specific decay in surface water; and (iii) benchmark the predictive power of the general persistence model from Chapter 4.

## 5.2 Methods

### 5.2.1 Persistence Data Analysis

Water samples taken from the retention pond were taken on days 0, 1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 15, 22, 33, and 38 after the containment of the spill event following pre-established methods (Worley-Morse et al., 2019). Culturable enterococci (cENT) numbers were quantified in MPN/mL. Molecular assays were used to quantify concentrations of enterococci (mENT), HF183, enterohemorrhagic *E. coli* (EHEC), *Pseudomonas aeruginosa*, crAssphage, pepper mild mottle virus (PMMoV), and human adenovirus (HAV) 41, 42 in gene copies/100mL. As this case study is focused on the utility of the indicator-pathogen paradigm for surface water decision making, only the traditional FIB (cENT, mENT) and the hazards of concern (EHEC, HAV) are analyzed herein. Observations below the limit of detection were assigned the value of the limit of detection as the observed concentration – a conservative simplifying assumption. The concentrations of each target were transformed into log-reduction values, and the datasets were assessed for a negative trend by fitting a simple linear regression with a forced intercept of zero.

The recreational water quality criteria (RWQC) for cENT associated with a risk of gastrointestinal illness of 36/1,000 swimmers are a geometric mean of 35 CFU/100mL and a

standard threshold value of 130 CFU/100mL (U.S. EPA, 2012). Decision makers associated with the spill event evaluated cENT until concentrations fell below 130 CFU/100 mL. The Juneja and Marks 2 (JM2) model was fit to the cENT, EHEC, and HAV datasets using maximum likelihood estimation as in prior works (Dean and Mitchell, 2022b; Dean et al., 2020; Mitchell and Akram, 2017). The JM2 model form is shown in Eq. 1, where the log-reduction from the initial concentration,  $\log_{10}\left(\frac{Nt}{N_0}\right)$ , can be estimated with model parameters ( $k_1$  and  $k_2$ ) and time ( $t$ ).

$$\log_{10}\left(\frac{Nt}{N_0}\right) = \log_{10}\left(\frac{1}{1+e^{k_1+k_2 \ln(t)}}\right) \quad \text{Eq. 1}$$

The best fitting JM2 model for cENT was used to calculate the time required for concentrations to fall below 130 CFU/100 mL and 35 CFU/100 mL, and the best fitting JM2 models for EHEC and HAV were used to estimate the concentration of pathogens present when cENT returns to the aforementioned values.

### 5.2.2 QMRA Considerations and Design

The impacted retention pond is known to be used for recreation by nearby community members, and thus a recreational water QMRA was completed to evaluate: i) how the risk of illness changes over time after the spill event, and ii) the uncertainty associated with relying on traditional indicator data to assess water quality. The impacted retention pond was assumed to be used for full-body recreation, and thus the risk of illness associated with accidentally ingesting water while swimming was evaluated. The concentrations documented in the sampling efforts were the concentrations considered in the QMRA. A probabilistic QMRA was conducted using well established methods (Haas et al 2014). Considerations within each step of the paradigm are outlined in the following sections.

### 5.2.2.1 Hazard Identification

Only *E. coli* 0157:H7 (EHEC) and HAV were considered within the risk assessment, as they are both waterborne pathogens commonly associated with the oral exposure route. EHEC refers to the group of Shiga toxin-producing *E. coli* or Verocytotoxin-producing *E. coli* (CDC, 2014). The most identified EHEC in the United States is *E. coli* 0157:H7, and common symptoms of infection include diarrhea and vomiting. Although EHEC infections can occur in any population, children and elderly populations are more likely to develop serious illness, such as hemolytic uremic syndrome, a life-threatening complication that may lead to kidney failure (CDC, 2014). The primary source of EHEC are the intestinal tracts of animals, and fecal contamination can be introduced to surface waters from a range of sources. Human adenoviruses (HAV) are common in the environment and are associated with a wide range of illnesses including gastroenteritis, the common cold, and pneumonia (CDC, 2019). HAV serotypes 40 and 41 are a major cause of gastroenteritis worldwide and are associated with long persistence in the environment due to their non-enveloped capsid (WHO, 2017).

### 5.2.2.2 Exposure Assessment

The exposure pathway evaluated was the ingestion of pathogen-contaminated recreational water while swimming immediately after containment and after several days of decay, as shown in Figure 5.1. Molecular methods were used to quantify the concentrations of EHEC and HAV in the retention pond. The location that the primer sets target are found once in the genome in the assays used, thus it was assumed that 1 gene copy corresponded to one CFU and virion for EHEC and HAV, respectively. However, the concentrations may be overestimated and considered conservative since all virions or bacteria detected may not be infectious. The initial concentrations of EHEC and HAV ( $C_{in}$ ) in the retention pond were 30,700 CFU/100mL

and 34,700 PFU/100 mL, respectively, based on the monitoring data. The change in concentration over time for EHEC and HAV were modeled with JM2 models fit to the sampling data (Section 5.2.1) and the number of days required for cENT to fall below 130 CFU/100 mL and 35 CFU/100 mL were used to calculate concentrations of EHEC and HAV at those respective time points ( $C_t$ ).

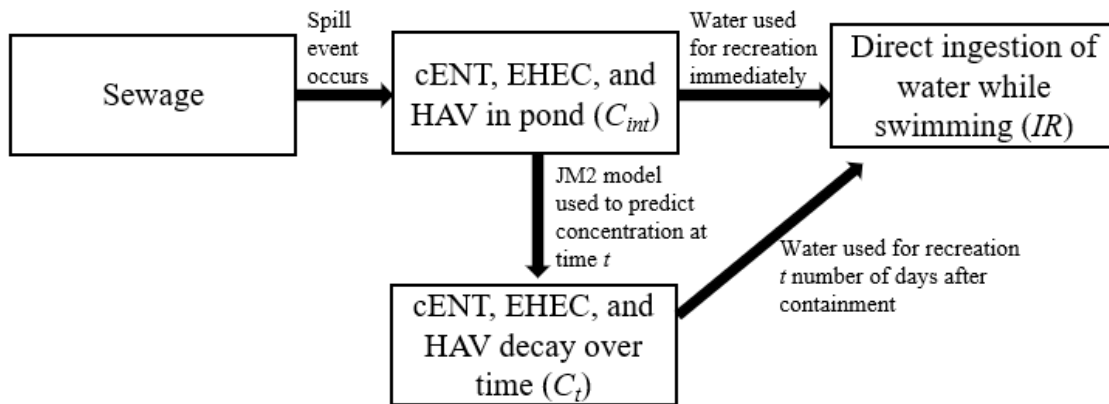


Figure 5.1: Exposure pathway for QMRA case study of sewage spill-impacted recreational waterbody

A study by Schets et al. (2011) determined that women, men, and children ingest an average of 18, 27, and 37 mL of water while swimming, respectively. Schets et al. (2011) also concluded that the variability in ingestion rates could be described by a gamma distribution. The parameters of the gamma distribution for each population are shown in Table 5.1. As shown in Equation 2, the concentrations of the pathogens,  $C$  (CFU/100 mL or PFU/100 mL), and the ingestion rates,  $IR$  (mL), for women, men, and children were used to calculate exposure doses,  $d$  (CFU or PFU), with Equation 2.

$$d = \frac{C_{int}}{100} \times IR \quad \text{Eq. 2}$$

### 5.2.2.3 Dose Response Assessment

An optimal model to describe the relationship between exposure dose and probability of response for EHEC has been contested in the literature (Teunis et al., 2004; Haas et al., 2000) and hence not a straightforward selection. The recommended model on the QMRA Wiki (qmrawiki.org) evaluated the infection endpoint in pigs after oral exposure to doses varying from 170 CFU to 40,000 CFU of *E. coli* 0157:H7 (Cornick & Helgerson, 2004). The exponential model provided the best fit to the data with a median  $k$  parameter estimate of 2.18E-04 (Weir et al., 2003; Cornick & Helgerson, 2004), where  $k$  represents the probability that an organism survives the host defenses to initiate an infection (Equation 3).

$$P(d) = 1 - e^{-kd} \quad \text{Eq. 3}$$

An assessment conducted by Powell et al. (2000) to develop a dose-response model for EHEC evaluated the illness endpoint in humans after oral exposure to *Shigella dysenteriae* strains and enteropathogenic *E. coli* (EPEC). The exact beta-Poisson model, shown in Equation 4, provided the best fit to each study's data, and the authors considered the *S. dysenteriae* beta-Poisson model ( $\alpha=0.16$   $\beta=9.17$ ) as the upper bound of the dose-response relationship for EHEC, and the beta-Poisson model for EPEC ( $\alpha=0.22$ ,  $\beta=3,112,348$ ) as the lower bound. The most likely parameter estimates for EHEC were estimated to be an  $\alpha$  of 0.22 and  $\beta$  of 8,722 (Powell et al., 2004). As shown in Equation 4, the exact beta-Poisson model assumes heterogeneity in the ability of an organism to survive and initiate an infection in the host. The heterogeneity follows a beta distribution with parameters  $\alpha$  and  $\beta$ . The exact beta-Poisson model is calculated with a hypergeometric function using the `gsl` package in R.

$$P(d) = 1 - \text{hyperg}_1F1(\alpha, \alpha + \beta, -d) \quad \text{Eq. 4}$$

An epidemiological investigation of an outbreak in Japan determined that 25% of exposed children and 16% of exposed adults were infected after the ingestion of 31 CFU and 35 CFU, respectively (Teunis et al., 2004). The observed outbreak responses were more closely approximated with the exact beta-Poisson model fit to the Shigella data in Powell et al. (2000), as opposed to the EPEC or Shigella/EPEC exact beta-Poisson models (Powell et al., 2000). In a review of data from eight EHEC outbreaks, Teunis et al. (2008) used hierarchical methods to account for the heterogeneity in exposure. The best fitting parameter estimates for  $\alpha$  and  $\beta$  were 0.37 and 39.7, respectively, and the model form is shown in red in Figure 5.2. Previous EHEC risk assessments have conservatively used the Shigella dose response model (Powell et al., 2000; Westrell et al., 2004; Strachan et al., 2001; Ryan et al., 2014), or the EHEC dose response model fit to the outbreak data (U.S. EPA, 2010). As the beta-Poisson dose response model for the infection endpoint (Teunis et al., 2008) is based on EHEC data specifically and has been deemed a conservative selection in prior risk assessments (U.S. EPA, 2010) it was the model selected herein. As illness is the endpoint of interest within this case study, a morbidity rate ranging from 20%-40% was used to calculate risk of illness from EHEC ingestion based on previous outbreak data (U.S. EPA, 2010; Teunis et al., 2004; Bielaszewska et al., 1997). Point estimates for the selected  $\alpha$  and  $\beta$  parameters are shown in Table 5.1.

### Comparison of D-R Models for EHEC

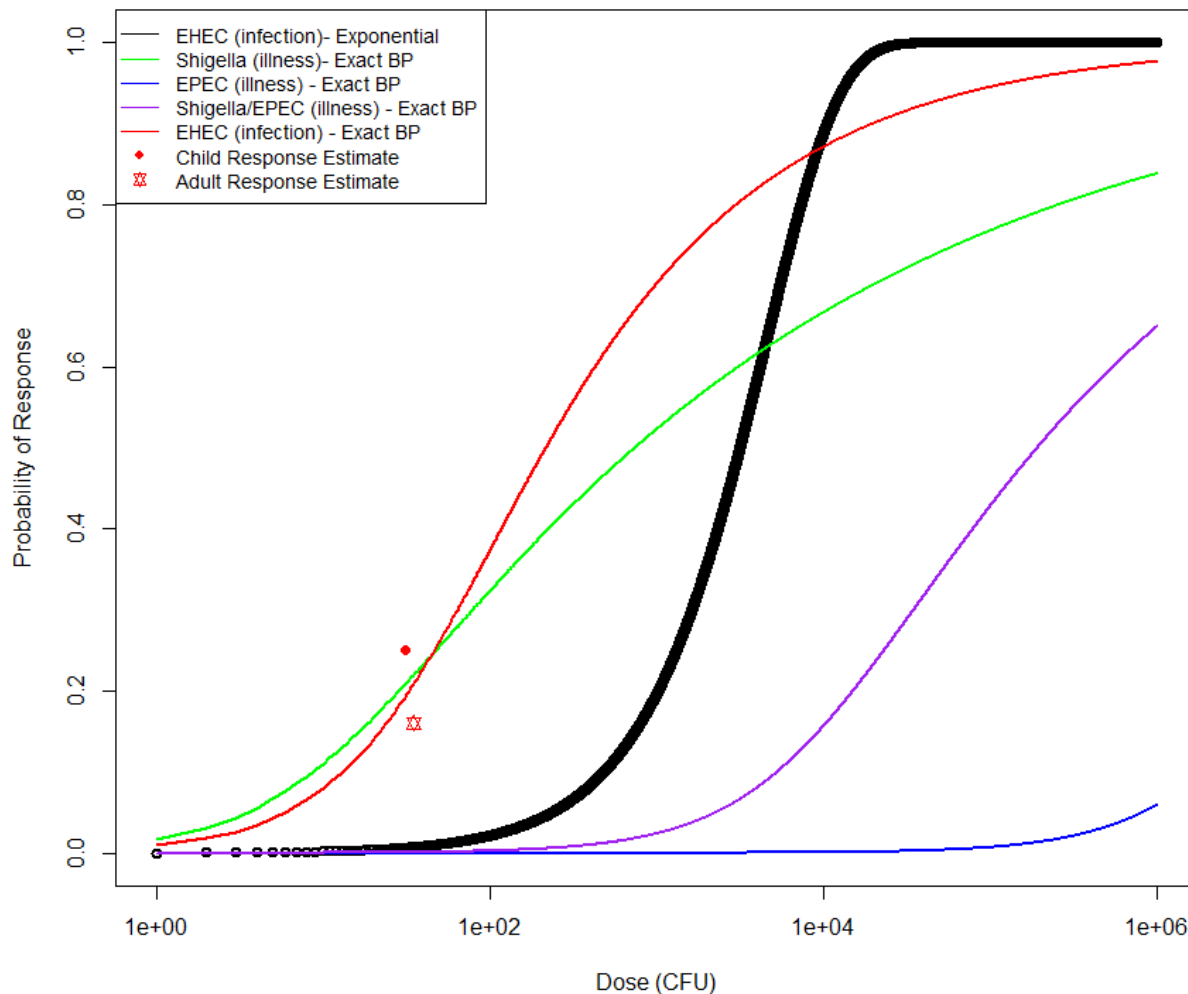


Figure 5.2: Comparison of dose-response models for EHEC available in the literature and outbreak response data; Resources: exponential model in black for EHEC from Wiki (2003), exact beta-Poisson models for Shigella in green, EPEC in blue, and Shigella/EPEC in purple from Powell et al. (2000), exact beta-Poisson in red for EHEC from Teunis et al. (2008), and outbreak responses in red from Teunis et al. (2004)

Only one study identified in the literature fit a dose response model for an oral exposure route to HAV. Teunis et al. (2016) used a hierarchical framework to fit HAV dose response models for ingestion, inhalation, intranasal, and intraocular routes, considering variation in infectivity and uncertainty due to limited data. The studies evaluated the infection response in human hosts, and the beta-Poisson model provided the best fit to the data. Point estimates for  $\alpha$



and  $\beta$  are shown in Table 5.1. Teunis et al. (2016) used similar methodology to calculate conditional illness parameters, such that the morbidity rate is dependent on the dose and can be calculated with Equation 5, where  $\eta$  and  $r$  are the parameters for the gamma distribution describing the length of the infection period. The impact of assuming illness is conditional on dose is illustrated in Figure 5.3. Point estimates for  $\eta$  and  $r$  are shown in Table 5.1.

$$P_{ill|inf}(cV | r, n) = 1 - \left(1 + \frac{cV}{\eta}\right)^{-r} \quad \text{Eq. 5}$$

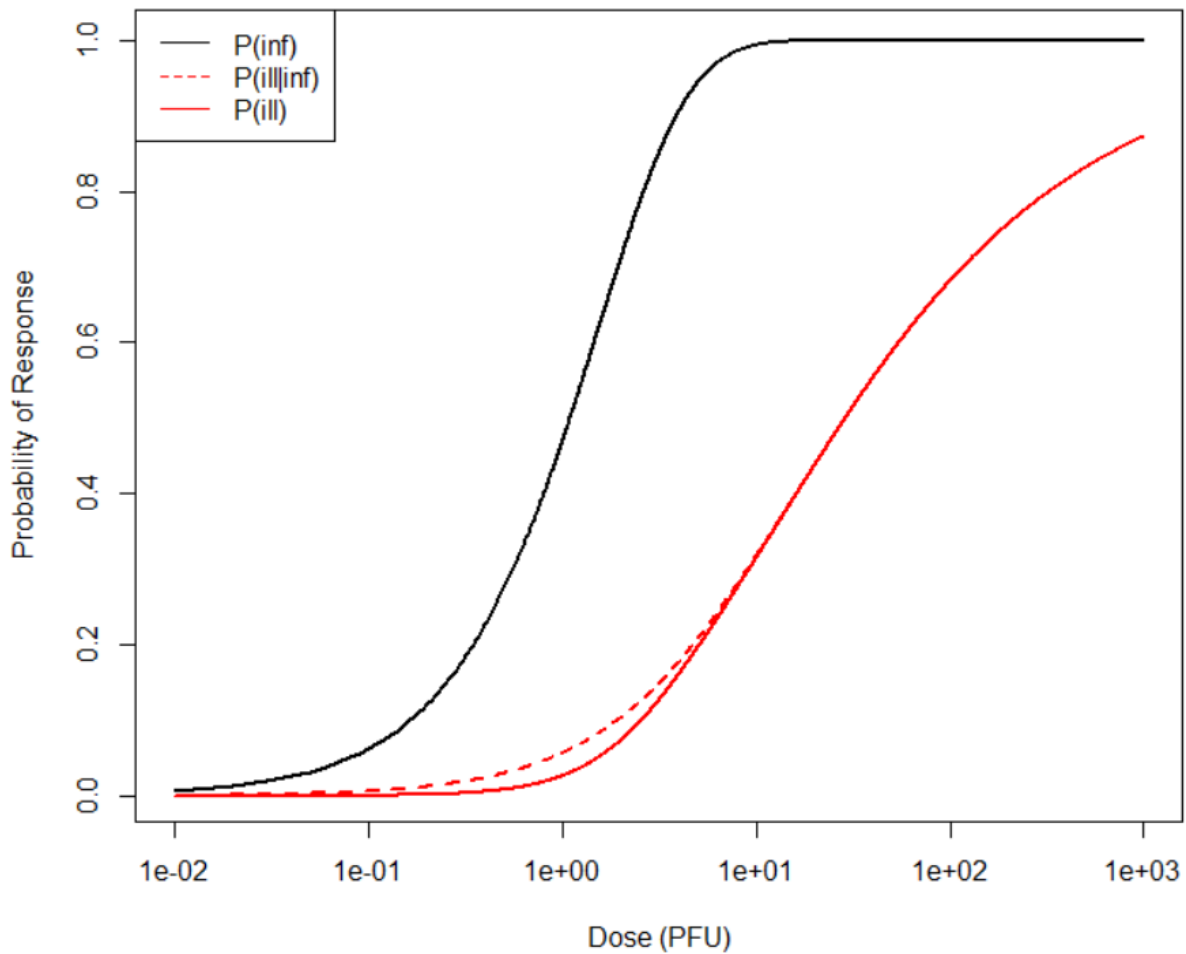


Figure 5.3: Illustration of the HAV dose response model developed for the oral exposure route and the dose-dependent morbidity rate from Teunis et al. (2016)

The risk assessment was simulated in a Monte Carlo analysis with 10,000 runs and a set seed of 123. Risk of illness was calculated at several timepoints after the spill, including (1) the initial time point (no decay); (2) the time required for cENT to decay to the water quality criteria standard threshold value (130 CFU/100 mL) and the geometric mean value (35 CFU/100 mL); and (3) the time required for cENT to decay to the water quality criteria values including the uncertainty factors evaluated in Chapter 4.

Table 5.1: Parameters and Distribution Types for QMRA Case Study

Parameter		Value	Units	Distribution	Source
Initial Concentration ( $C_{int}$ )	cENT	2.61E+05	MPN/100mL	Point Estimates	This study
	EHEC	3.07E+04	gene copies/100mL		
	HAV	3.47E+05	gene copies/100mL		
Persistence Models	cENT	JM2, $k_1=0.35$ , $k_2=2.13$	NA	Point Estimates	This study
	EHEC	JM2, $k_1=1.05$ , $k_2=1.42$			
	HAV	JM2, $k_1=-7.30$ , $k_2=4.50$			
Ingestion Rates	Women	18	mL	Gamma( $r=0.51$ , $\lambda=35$ )	Schets et al. 2010
	Men	27		Gamma( $r=0.45$ , $\lambda=60$ )	
	Children	37		Gamma( $r=0.64$ , $\lambda=58$ )	
Dose Response Model	EHEC	Beta-Poisson, $\alpha=0.37$ , $\beta=39.1$	NA	Point Estimates	Teunis et al., 2008
	HAV	Beta-Poisson, $\alpha= 5.11$ , $\beta=2.80$ Illness parameters, $\eta=6.53$ and $r=0.41$		Point Estimates	Teunis et al., 2016
Morbidity Rate	EHEC	30	%	Uniform(min=20%, max=40%)	U.S. EPA, 2010

### 5.2.3 General Model Application

In addition to the application of the uncertainty factors, the general model developed in Chapter 4 was also applied to the sampling data to analyze the utility of a general model on inferences for persistence of targets in a contamination event. Targeted sampling efforts for several specific pathogens over time is labor and resource intense so application of a general

model to describe persistence based on pathogen target type (bacteria, virus, etc.) may be a feasible way to reduce uncertainty in risk assessment and decisions.

### 5.3 Results

#### 5.3.1 Indicator and Pathogen Persistence

The best fitting parameter estimates for the JM2 models fit to cENT, mENT, EHEC and HAV are in Table 5.1 and the persistence curves for each of the four targets are shown in Figure 5.4. The decay of cENT and mENT was similar to EHEC for the first log-reduction (T90), but the T90 for HAV was greater than that of the other targets. The decay of EHEC over time tapered off most quickly as shown in Figure 5.4, and cENT and mENT became conservative estimators of HAV persistence past the 3.5 log-reduction time point. As expected from the general model (Chapter 4), HAV had a predominant shoulder in the decay curve. However, in contrast to the general model, the pathogenic bacteria (EHEC) persistence significantly differed from the indicator data past the 2 log-reductions time point.

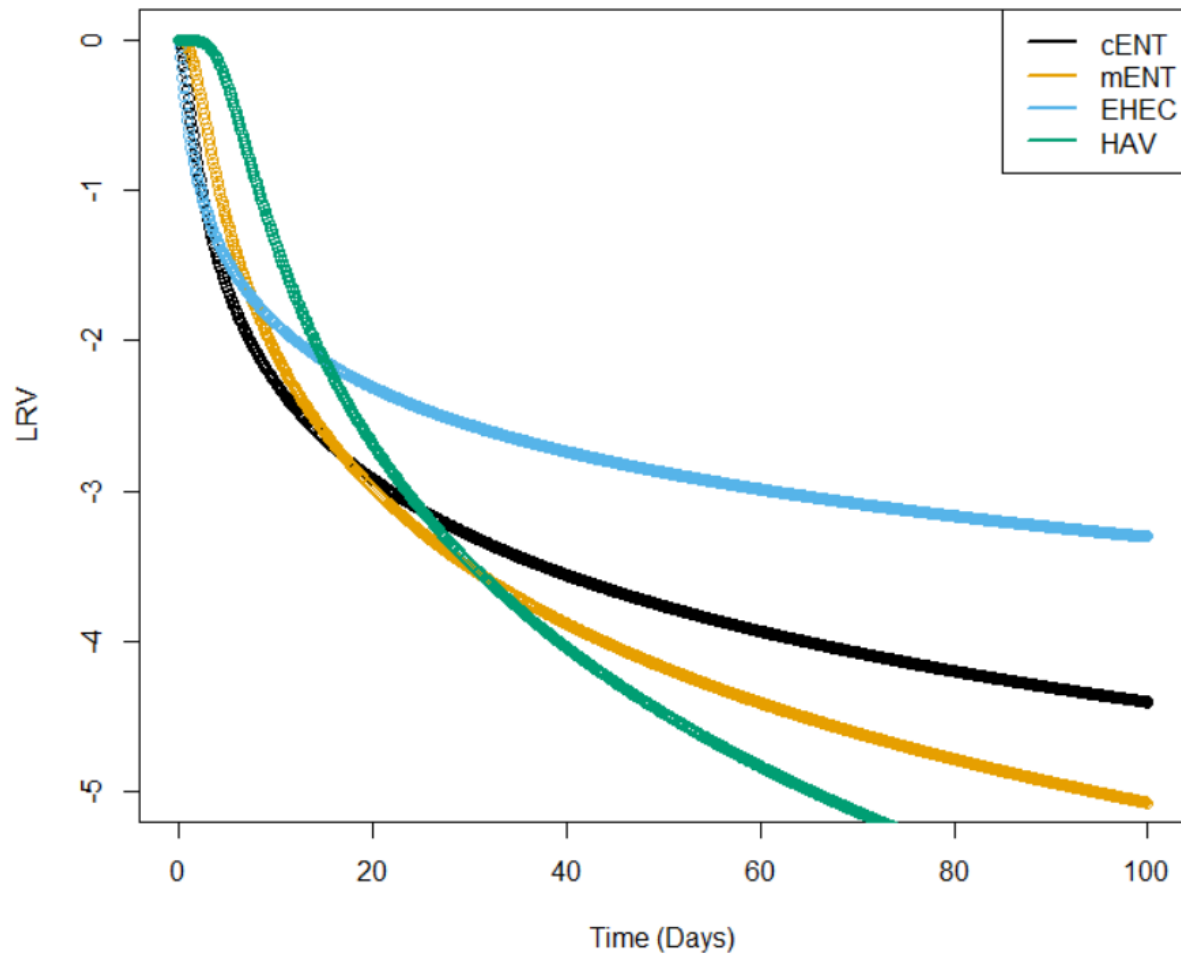


Figure 5.4: JM2 models depicting the persistence of cENT, mENT, EHEC and HAV in the retention pond

The initial concentration of cENT was 261,000 MPN/100 mL. Therefore, a 3.9 log-reduction was required to meet the geometric mean (GM) criteria of 35 CFU/100 mL and a 3.3 log-reduction was required to meet the standard threshold value (STV) of 130 CFU/100 mL. Using the JM2 model fit to the cENT data (Table 5.1), it was estimated that the 3.3 and 3.9 log-reduction would be achieved after 30 and 58 days, respectively. Using their respective JM2 models, the concentrations of EHEC after 30 and 58 days were estimated to be 83 CFU/100 mL and 33 CFU/100 mL, and the concentrations of HAV after 30 and 58 days were estimated to be 11 PFU/100 mL and 0.60 PFU/100 mL.

### 5.3.2 Uncertainty Factors for Pathogen Persistence

The development of the general model (Chapter 4) indicated that in 20°C surface water the average decay of bacteria may be 1-1.1x slower than the average decay of FIB and the average decay of viruses may be 1-1.9x slower. Thus, the risk of illness for EHEC was also evaluated at 33 days (73 CFU/100 mL) and 64 days (29 CFU/100 mL) and the risk of illness for HAV was evaluated at 57 days (0.62 PFU/100 mL) and 110 days (0.03 PFU/100 mL).

### 5.3.3 Risk of Illness from Swimming

The risk of illness for EHEC and adenovirus were evaluated for women, men, and children assumed to be swimming in the retention pond at the aforementioned timepoints. Children had the greatest risk of illness for each time point, as they were estimated to ingest the most water while swimming. The median risk of illness for EHEC for the three populations was 28-31% immediately after the spill event compared to median risks of 92-94% for HAV. For an acceptable risk of 36 illnesses in 1,000 people (3.6%), the STV is 130 CFU/100 mL for cENT, and this concentration was predicted to occur 30 days after the spill was contained. The EHEC median risk of illness ranged from 2-5% after 30 days, and the HAV median risk ranged from 2-9% for the evaluated populations. Applying the bacteria and virus uncertainty factors from Chapter 4 reduced the median risk of EHEC illness to 2-4% and the median risk of HAV illness to 0.01-0.06%.

Table 5.2: Estimated Risk of Illness for EHEC and HAV for a Swimming Exposure

Time (Days)	Population	EHEC Risk of Illness		HAV Risk of Illness	
		Median	[2.50%, 97.50%]	Median	[2.50%, 97.50%]
Initial (T=0)	Women	0.28	[0.02, 0.50]	0.92	[0.25, 0.97]
	Men	0.29	[0.01, 0.51]	0.93	[0.19, 0.97]
	Children	0.31	[0.09, 0.52]	0.94	[0.61, 0.98]
FIB Levels Decay to STV Criteria Level (T=30)	Women	0.02	[5.49E-05, 0.14]	0.02	[2.01E-07, 0.31]
	Men	0.03	[3.96E-05, 0.18]	0.04	[1.03E-07, 0.40]
	Children	0.05	[4.70E-04, 0.20]	0.09	[1.42E-05, 0.43]

Table 5.2 (cont'd)

FIB Levels Decay to STV Criteria with UF* (T <sub>EHEC</sub> =33, T <sub>HAV</sub> =57)	Women	0.02	[4.79E-05, 0.13]	9.07E-05	[6.21E-10, 0.01]
	Men	0.02	[3.46E-05, 0.17]	1.90E-04	[3.18E-10, 0.02]
	Children	0.04	[4.11E-04, 0.19]	4.43E-08	[6.40E-04, 0.03]
FIB Levels Decay to GM Criteria (T=58)	Women	0.01	[2.18E-05, 0.08]	8.50E-05	[5.82E-10, 0.01]
	Men	0.01	[1.57E-05, 0.11]	1.78E-04	[2.98E-10, 0.02]
	Children	0.02	[1.87E-04, 0.12]	6.01E-04	[4.15E-08, 0.03]
FIB Levels Decay to GM Criteria with UF* (T <sub>EHEC</sub> =64, T <sub>HAV</sub> =110)	Women	0.01	[1.90E-05, 0.07]	2.68E-07	[1.80E-12, 3.50E-05]
	Men	0.01	[1.37E-05, 0.10]	5.66E-07	[9.21E-13, 9.43E-05]
	Children	0.02	[1.63E-04, 0.11]	1.96E-06	[1.28E-10, 1.24E-04]

\*UF: Uncertainty Factor

The distributions of risk associated with the STV criteria level are illustrated in Figures 5.5 and 5.6. The initial risk of illness for HAV ingestion is greater than the risk of illness for EHEC ingestion immediately after the spill, and although HAV decayed more quickly than EHEC, with more than 3.5 log-reductions observed at 30 days compared to approximately 2.5 log-reductions for EHEC, the risk of illness after 30 days was still higher for HAV ingestion than EHEC. Applying the FIB-bacteria uncertainty factor of 1.1 still resulted in median risks greater than the acceptable 3.6% for EHEC ingestion for children. When the FIB-virus uncertainty factor of 1.9 was applied, the risk associated with HAV ingestion was two orders of magnitude lower than the acceptable level for all populations. Although the preceding meta-analysis and general model identified minimal differences between FIB and bacteria, the 30-day timepoint was associated with a 3.3 log-reduction of cENT and only a 2.5 log-reduction of EHEC. An uncertainty factor of 1.1 lessened this gap to only 0.7 log-reductions, still resulting in median risks above the acceptable level for some of the populations of concern (Figure 5.5).

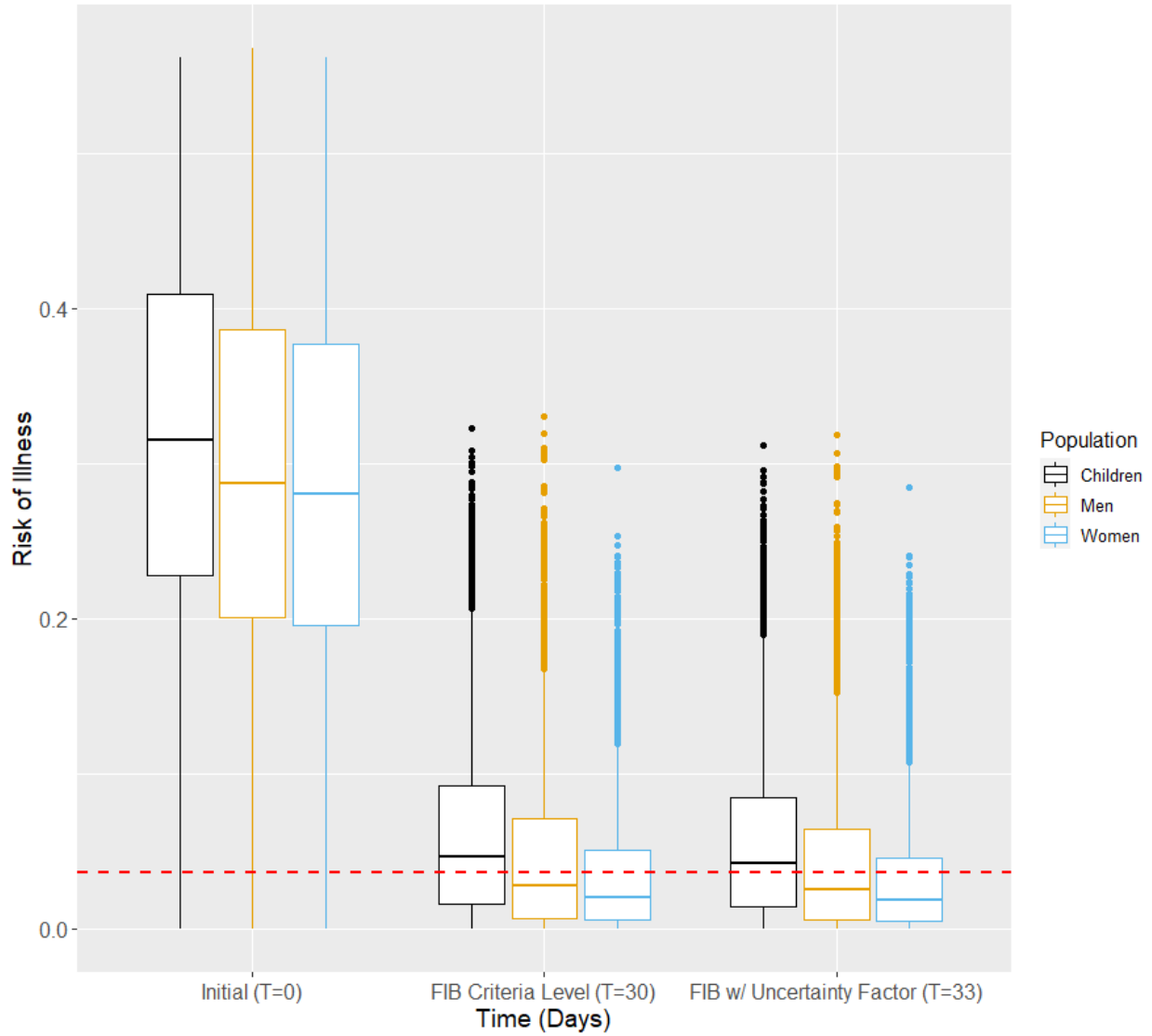


Figure 5.5: Box plots of risk of illness for EHEC in recreational waters immediately after the spill (T=0), when FIB levels decay to the water quality criteria STV levels (T=30 days), and when considering the FIB-bacteria uncertainty factor of 1.1 (T=33 days)

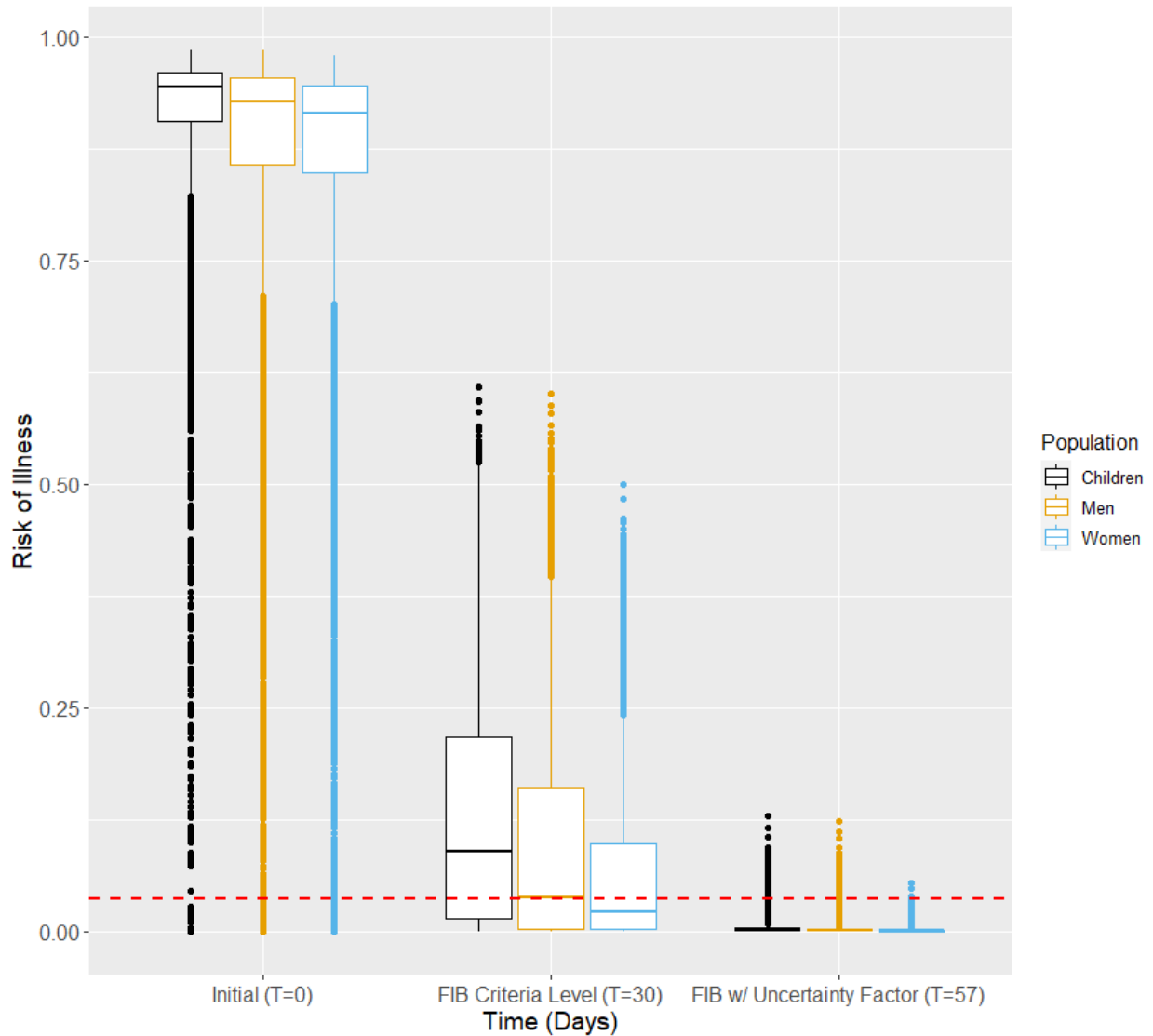


Figure 5.6: Box plots of risk of illness for HAV in recreational waters immediately after the spill (T=0), when FIB levels decay to the water quality criteria STV levels (T=30 days), and when considering the FIB-virus uncertainty factor of 1.9 (T=57 days)

cENT decayed to the GM criteria value in freshwater (35 CFU/100mL) 58 days after the spill was contained. The median risk of illness for EHEC 58 days after the spill ranged from 1%-2% for the three populations, and the median risk of illness for adenovirus ranged from <0.01% to 0.06%. When the FIB-bacteria and FIB-virus uncertainty factors were applied to the estimated persistence of cENT, the median risk of illness remained at 1%-2% for EHEC and fell below 1 in



100,000 for adenovirus. Figures 5.7 and 5.8 illustrate the change in risk associated with the time required for cENT to reach the established RWQC GM values (U.S. EPA, 2012). Applying the FIB-virus uncertainty factor resulted in increased confidence that the HAV risk of illness were below the acceptable level, with the upper quantiles of the distribution of risk ranging from 0.004-0.01% (Figure 5.8).

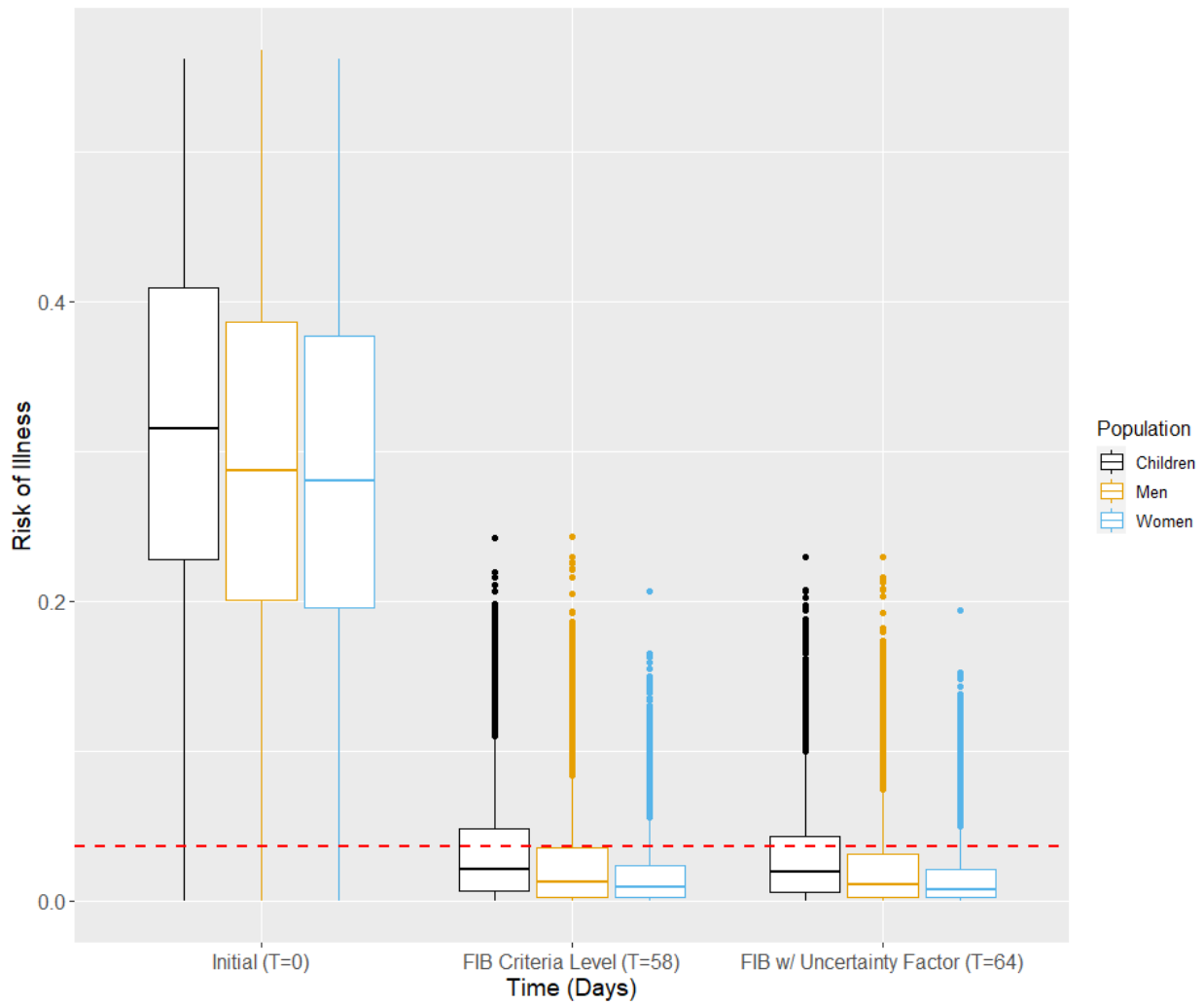


Figure 5.7: Box plots of risk of illness for EHEC in recreational waters immediately after the spill (T=0), when FIB levels decay to the water quality criteria GM levels (T=58 days) associated with a risk of 36/1000 (dashed red line), and when considering the FIB-bacteria uncertainty factor of 1.1 (T=64 days)

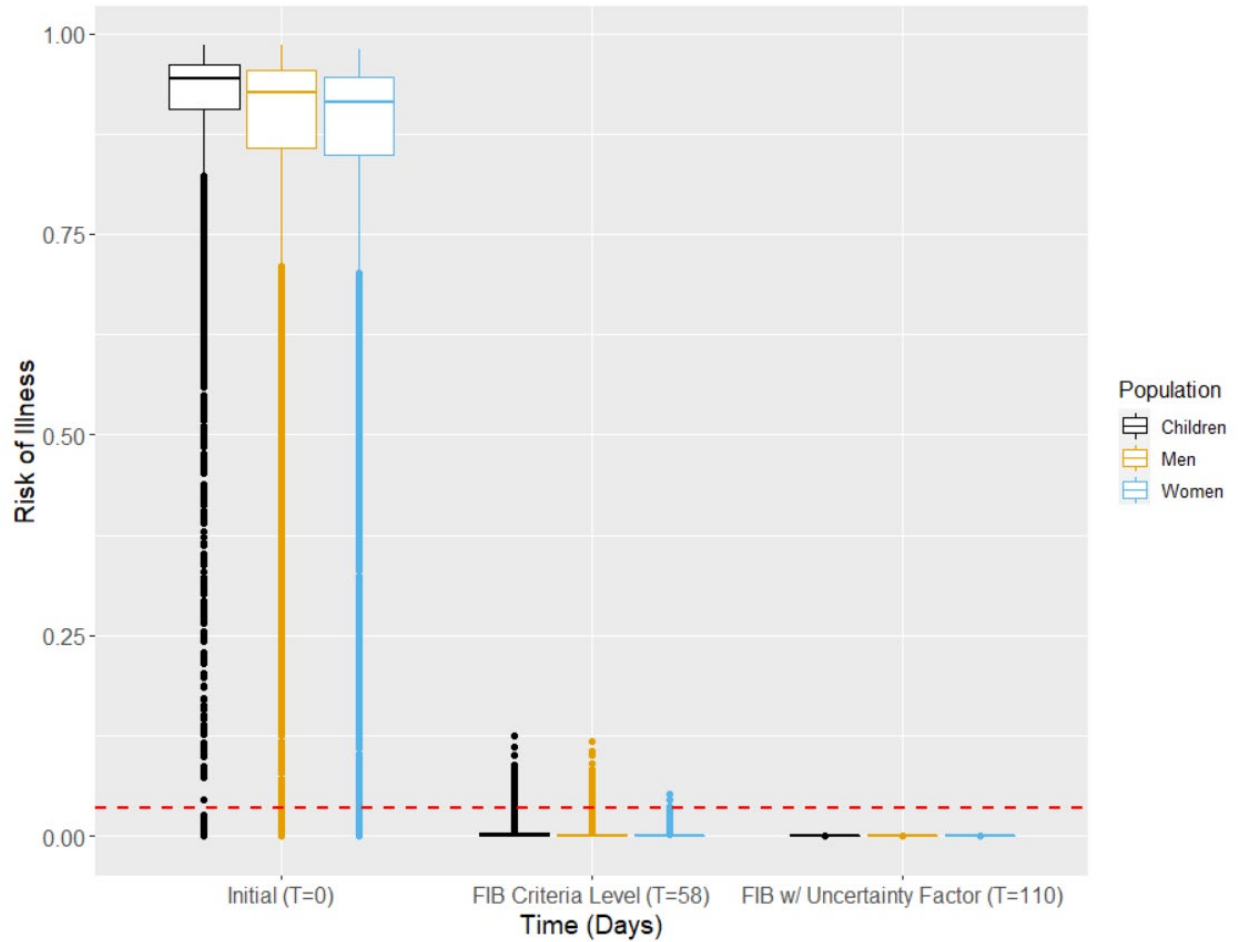


Figure 5.8: Box plots of risk of illness for HAV in recreational waters immediately after the spill (T=0), when FIB levels decay to the water quality criteria GM levels (T=58 days) associated with a risk of 36/1000 (dashed red line), and when considering the FIB-virus uncertainty factor of 1.9 (T=110 days)

#### 5.3.4 Application of the General Model

The general models developed for the persistence of FIB, bacteria, and viruses (Chapter 4) were compared to the cENT, EHEC, and HAV sampling data from the spill event. As shown in Figure 5.9, the majority of the observed log-reductions of cENT, EHEC, and HAV fell within the 95% confidence bounds of the general models. The EHEC observations excluded from the general model for pathogenic bacteria were below the limit of detection (BL0D). Figure 5.9 indicates that in lieu of additional sampling data, the upper bound of the general models would

have provided relatively conservative estimates for the persistence of the targets of interest over time.

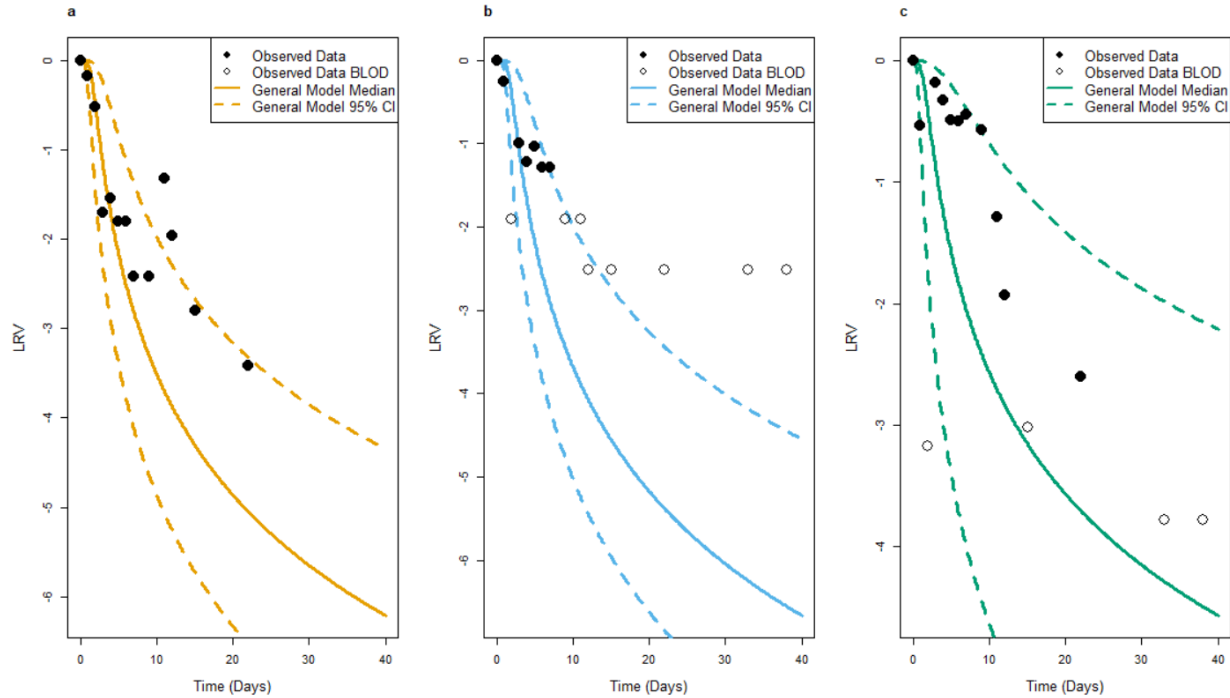


Figure 5.9: Comparison of the a) FIB, b) bacteria, and c) virus general model performance on the sampling data for a) cENT, b) EHEC, and c) HAV

#### 5.4 Discussion

A QMRA case study was completed for a sewage spill event affecting a retention pond known to be used for recreation. Concentrations of cENT, EHEC, and HAV were monitored for 40 days after the containment of the spill, and the JM2 model was fit to the persistence data for each target. Using the best fitting JM2 models and the uncertainty factors developed in Chapter 4, the risk of illness for EHEC and HAV ingestion while swimming was characterized immediately after the spill, when cENT levels returned to recreational water quality criteria values, and when the uncertainty factors were applied to the cENT data. Children had the highest risk of illness for all scenarios, as they were expected to ingest the most water while swimming (Schets et al., 2011). HAV was associated with the highest risk of illness immediately after the

spill, however the risk of illness associated with EHEC ingestion was higher than HAV after 58 days. HAV decayed more quickly than EHEC, achieving 4.8 log-reductions within 58 days compared to only 3 log-reductions for EHEC. The difference in risk between the pathogens at later time-points may also have been driven by HAV's dose dependent morbidity rate; the concentration of HAV immediately after the spill was associated with a morbidity rate of 93%, whereas the concentration after 30 days was associated with a morbidity rate of less than 1% (Teunis et al., 2016).

In the context of the assessed sewage spill, the sampling efforts were critical for evaluating the changing water quality of the retention pond to protect potential water users. The QMRA case study completed in this analysis suggests that the time required for cENT to decay to the STV (30 days) was not sufficient for the risk of illness from EHEC or HAV ingestion to be below 36 in 1,000 for children swimming. The time required for cENT to decay to the GM (58 days), however, was associated with median risks below 3.6%. This is not unexpected, as the EPA recommends that the GM and STV be considered together as the criteria magnitude.

The general model developed in Chapter 4 indicated that pathogenic bacteria in 20°C freshwater may be 1-1.1x slower than FIB, and that virus decay may be up to 1-1.9x slower than FIB. When these uncertainty factors were applied to the time required for cENT to return to GM and STV criteria levels, the 95% confidence interval of risk of illness from HAV ingestion was below 3.6% for both time points. Applying the uncertainty factor of 1.1 to the time required for cENT to return to the GM criteria, ensured the median risks for EHEC ingestion were below 3.6%, however the upper quantiles of risk were still as high as 11%. These results reinforce the need to dually consider GM and STV when monitoring water quality and demonstrate the ability of the uncertainty factors developed in Chapter 4 to add value to classically documented

indicator data, by providing more accurate estimates of the time required for the risk of illness associated with EHEC and HAV to fall below 3.6%.

The importance of characterizing persistence within a QMRA is evident in Figures 5.5-5.8, however, pathogen and site-specific persistence data are not always available for contamination events that may warrant risk assessments. To address this gap, the general models for indicator and pathogen persistence developed in Chapter 4, were compared to the sampling data from the spill event analyzed herein. As shown in Figure 5.9, the general models for FIB, bacteria, and virus persistence in surface waters with average temperatures of 20°C capture the trend of most of the sampling data for cENT, EHEC, and HAV. The first observation below 130 CFU/100 mL occurred on day 22 and the upper bound of the FIB general model predicted a return to 130 CFU/ 100 mL on day 22 (Figure 5.8a).

There were limitations to the QMRA case study developed herein. The dose response model for oral exposure to EHEC selection has varied in previous risk assessments (U.S. EPA, 2010; Westrell et al., 2004; Strachan et al., 2001; Ryan et al., 2014). This case study selected the beta-Poisson model fit to data from eight outbreaks that was found to more closely match the dose response model developed for *S. dysenteriae* than other EHEC or EPEC models (Teunis et al., 2008; Powell et al., 2000). The selection of dose response model, however, has a significant impact on the interpretation of results (Figure A5.1), as a less conservative model choice such as the mixed Shigella and EPEC model analyzed in Powell et al. (2000) indicates risks below 3.6% for time points 30 days and above. The HAV dose response model is also a source of uncertainty in this case study, as there is only one model available in the literature based on minimal observational data (Teunis et al., 2016). The uncertainty associated with the conditional morbidity rate (Teunis et al., 2016) should also be explored in future works, as the morbidity rate

may have different dependencies on dose for different populations of concern such as children or the elderly. There was not sufficient data available in the literature to represent the dose response parameters with appropriate probability distributions, and the parameter pairs for each beta-Poisson model should be acquired in the future to improve the characterization of risks within this case study, as the heterogeneity in host response will impact the presented results as evident in Figure A5.1. Ingestion rates and the EHEC morbidity rate were the only parameters that accounted for uncertainty and variability with the Monte Carlo simulations. The initial concentrations of the FIB and pathogens are another source of uncertainty within the risk assessment, as the data gathered in the sampling efforts were considered point estimates within the case study. The concentrations of EHEC and HAV are also likely over-estimated as 1 gene copy was assumed to be the equivalent of 1 CFU or PFU, in line with the methods implemented for quantification. Future full risk assessments should use probability distributions to capture the uncertainty associated with this conversion.

Despite these limitations, this case study highlights the importance of considering persistence within surface water QMRAs to provide time-relevant information to surface water decision makers. For example, this case study determined that it would take more than 60 days for the water quality to return to acceptable levels for recreational uses. This information allows a water manager to consider the value of potentially applying other mitigation strategies to reduce this time. This case study also demonstrated the ways the uncertainty factors and general models developed in Chapter 4 can be used to improve the utility of the indicator-pathogen paradigm. Applying the uncertainty factor for bacteria and viruses ensured sufficient time for the median risks of illness to fall below 3.6%. Finally, in lieu of sampling data for the retention pond after the initial contamination event, the use of the general models developed in Chapter 4 can

help optimize the use of resources for future contamination events; the upper bound of the general model predicted that the concentration of cENT would return to 130 CFU/100 mL on day 22, and the resources used to sample on days 2-21 could have been conserved.

## REFERENCES

- Ahmed, W., Hamilton, K. A., Lobos, A., Hughes, B., Staley, C., Sadowsky, M. J., & Harwood, V. J. (2018). Quantitative microbial risk assessment of microbial source tracking markers in recreational water contaminated with fresh untreated and secondary treated sewage. *Environment International*, *117*, 243–249. <https://doi.org/10.1016/j.envint.2018.05.012>
- Bielaszewska, M., Janda, J., Bláhová, K., Minaříková, H., Jíková, E., Karmali, M. A., Laubová, J., ŠIKULOVÁ, J., Preston, M. A., Khakhria, R., Karch, H., Klazarová, H., & Nyč, O. (1997). Human *Escherichia coli* O157:H7 infection associated with the consumption of unpasteurized goat's milk. *Epidemiology and Infection*, *119*(3), 299–305. <https://doi.org/10.1017/S0950268897008297>
- Boehm, A. B., Graham, K. E., & Jennings, W. C. (2018). Can We Swim Yet? Systematic Review, Meta-Analysis, and Risk Assessment of Aging Sewage in Surface Waters. *Environmental Science & Technology*, *52*(17), 9634–9645. <https://doi.org/10.1021/acs.est.8b01948>
- CDC. (2014). “*E. coli* (*Escherichia coli*)”. [Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases \(NCEZID\), Division of Foodborne, Waterborne, and Environmental Diseases \(DFWED\)](https://www.cdc.gov/ecoli/index.html). Retrieved from: <https://www.cdc.gov/ecoli/index.html>
- CDC. (2019). “Adenoviruses”. [National Center for Immunization and Respiratory Diseases, Division of Viral Diseases](https://www.cdc.gov/adenovirus/index.html) Retrieved from: <https://www.cdc.gov/adenovirus/index.html>
- Cornick NA & Helgerson AF (2004) Transmission and infectious dose of *Escherichia coli* O157:H7 in swine. *Applied and Environmental Microbiology*. *70*(9), pp.5331-5335.
- Haas, C. N., Thayyar-Madabusi, A., Rose, J. B., & Gerba, C. P. (2000). Development of a dose-response relationship for *Escherichia coli* O157:H7. *International Journal of Food Microbiology*, *56*(2–3), 153–159. [https://doi.org/10.1016/S0168-1605\(99\)00197-X](https://doi.org/10.1016/S0168-1605(99)00197-X)
- Powell, M., Ebel, E., Schlosser, W., Walderhaug, M., & Kause, J. (2000). Dose-Response Envelope for *Escherichia coli* O157:H7. *Quantitative Microbiology*, *2*, 141–163.
- Ryan, M. O., Haas, C. N., Gurian, P. L., Gerba, C. P., Panzl, B. M., & Rose, J. B. (2014). Application of quantitative microbial risk assessment for selection of microbial reduction targets for hard surface disinfectants. *American Journal of Infection Control*, *42*(11), 1165–1172. <https://doi.org/10.1016/j.ajic.2014.07.024>
- 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. <https://doi.org/10.1016/j.watres.2010.06.049>



- Strachan, N. J. C., Fenlon, D. R., & Ogden, I. D. (2001). Modelling the vector pathway and infection of humans in an environmental outbreak of *Escherichia coli* O157. *FEMS Microbiology Letters*, 203(1), 69–73. <https://doi.org/10.1111/j.1574-6968.2001.tb10822.x>
- Teunis, P. F. M., Nagelkerke, N. J. D., & Haas, C. N. (1999). Dose Response Models for Infectious Gastroenteritis. *Risk Analysis*, 19(6), 1251–1260. <https://doi.org/10.1111/j.1539-6924.1999.tb01143.x>
- Teunis, P. F. M., Ogden, I. D., & Strachan, N. J. C. (2008). Hierarchical dose response of *E. coli* O157:H7 from human outbreaks incorporating heterogeneity in exposure. *Epidemiology and Infection*, 136(6), 761–770. <https://doi.org/10.1017/S0950268807008771>
- Teunis, P., Schijven, J., & Rutjes, S. (2016). A generalized dose-response relationship for adenovirus infection and illness by exposure pathway. *Epidemiology and Infection*, 144(16), 3461–3473. <https://doi.org/10.1017/S0950268816001862>
- Teunis, P., Takumi, K., & Shinagawa, K. (2004). Dose Response for Infection by *Escherichia coli* O157:H7 from Outbreak Data. *Risk Analysis*, 24(2), 401–407. <https://doi.org/10.1111/j.0272-4332.2004.00441.x>
- U.S. EPA. (2010). *Quantitative Microbial Risk Assessment to Estimate Illness in Freshwater Impacted by*. 456. [https://doi.org/EPA\\_822-R-10-005](https://doi.org/EPA_822-R-10-005)
- U.S. EPA. (2012). Recreational Water Quality Criteria. Office of Water 820-F-12-058. Retrieved from: <https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf>
- Weir, M. (2003). *Escherichia coli* enterohemorrhagic (EHEC). *QMRA Wiki*. Retrieved from: [http://qmrawiki.canr.msu.edu/index.php/Escherichia\\_coli\\_enterohemorrhagic\\_\(EHEC\):\\_Dose\\_Response\\_Models](http://qmrawiki.canr.msu.edu/index.php/Escherichia_coli_enterohemorrhagic_(EHEC):_Dose_Response_Models)
- Westrell, T., Schönning, C., Stenström, T. A., & Ashbolt, N. J. (2004). QMRA (quantitative microbial risk assessment) and HACCP (hazard analysis and critical control points) for management of pathogens in wastewater and sewage sludge treatment and reuse. *Water Science and Technology*, 50(2), 23–30. <https://doi.org/10.2166/wst.2004.0079>
- Worley-Morse, T., Mann, M., Khunjar, W., Olabode, L., & Gonzalez, R. (2019). Evaluating the fate of bacterial indicators, viral indicators, and viruses in water resource recovery facilities. *Water Environment Research*, 91: 830-842. 10.1002/wer.1096
- World Health Organization. (2017). *Guidelines for Drinking Water* (Fourth). World Health Organization. <https://doi.org/10.4060/cb7678en>

## APPENDIX

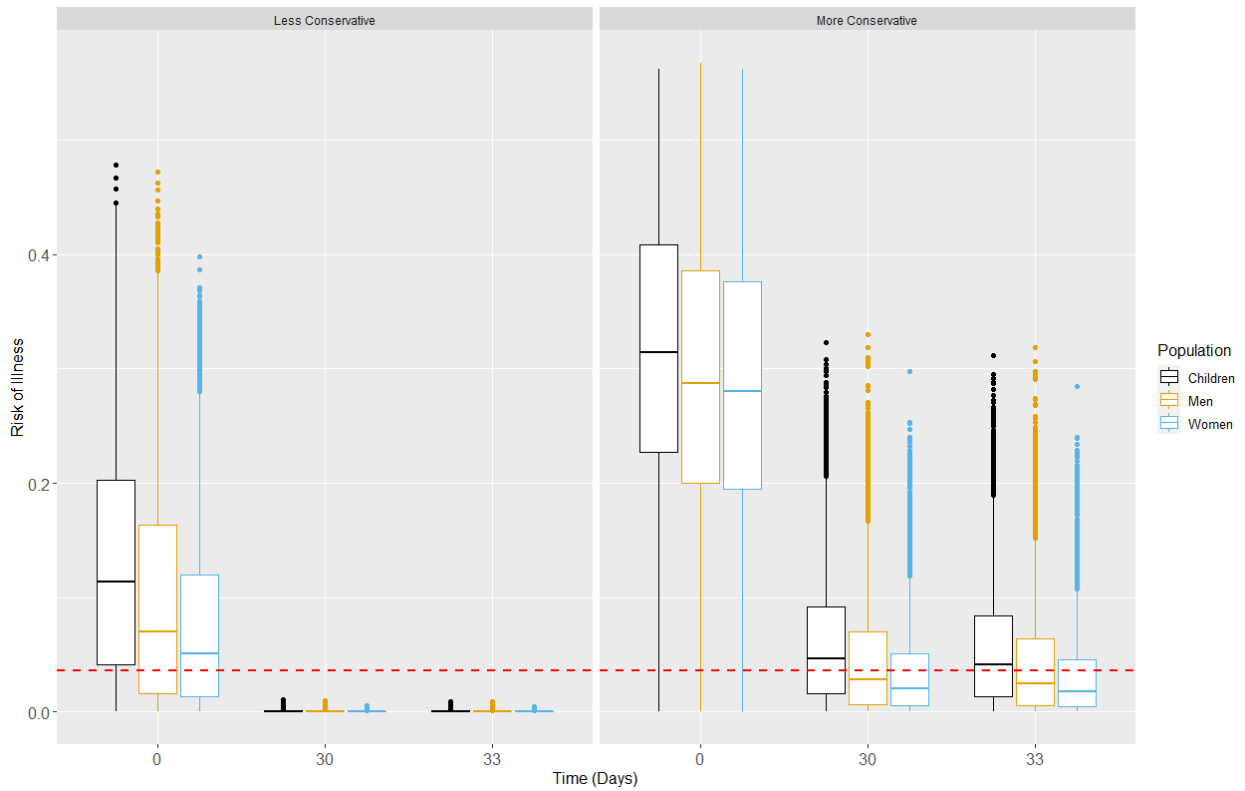


Figure A5.1: The difference in risk of illness estimates over time for EHEC when a) the less conservative Shigella and EPEC dose response model is used in the QMRA compared (Powell et al., 2000); or b) the more conservative EHEC dose response model developed with outbreak data (Teunis et al., 2008)

## CHAPTER 6: CONCLUSIONS

This dissertation has presented: (i) a comprehensive database of persistence experiments for indicators and pathogens in surface waters; (ii) statistically verified relationships between water quality and environmental factors driving persistence; (iii) identified a model form (Juneja and Marks 2) best able to describe target persistence in surface waters; and (iv) quantified the uncertainty between indicators and pathogen persistence in highly varied surface water conditions by evaluating the feasibility of a novel general model. The utility of these results for surface water management and policymaking were further demonstrated in a quantitative microbial risk assessment (QMRA) case study for a sewage spill-impacted waterbody.

In Chapter 2 it was concluded that most available persistence data focuses on fecal indicator bacteria (FIB), freshwater matrices, and culture-based methods of detection. Additional studies are needed that evaluate the persistence of virus and protozoa targets, brackish waters, and molecular-based methods of detection. Interactions between sunlight, water type, and method of detection, as well as predation, temperature, and water type, were identified qualitatively and could be further explored in future experimentation. An unexpected finding from Chapter 2 pertained to data availability; most of the datasets were not readily available in the literature or supplementary materials and needed to be digitized which may have introduced data uncertainty. Comprehensive open and shared databases of funded water pathogen research like those established in other fields could alleviate this source of uncertainty. Chapter 3 identified the Juneja and Marks 2 model as the model form best able to describe indicator and pathogen persistence in surface waters. The best fitting persistence models for over 400 datasets were used to predict dependent variables of interest for series of factor analyses. Random forests methods had the greatest performance and predictive power, highlighting the existence of

nonlinear relationships between factors such as temperature, turbidity, and pH. The random forests also identified significant interactions between temperature and predation as well as sunlight and method of detection. Temperature, water type, and method of detection were identified as the most important variables influencing the evaluated persistence metrics.

Chapter 4 tested the feasibility of a general model for persistence and identified a varying-intercept regression with target-level intercepts and population-level coefficients for temperature, water type, and predation as the optimal evaluated model form. Although pathogen-specific and site-specific persistence models are expected to be more accurate and are preferred, the general model provides information about the average persistence of FIB, bacteriophages, bacteria, viruses, and protozoa in various surface water conditions in lieu of site-specific data. The general model can be used by surface water managers in the future to make inferences about pathogen persistence when relying on indicator data and can be used to maximize resource efficiency; monitoring materials could be conserved during periods of expected minimal inactivation or alternative mitigation strategies could be pursued to minimize the time the water does not meet surface water quality criteria. Potential applications of the general model were demonstrated in Chapter 5, as the application of bacteria and virus uncertainty factors ensured the risk of illness from enterohemorrhagic *E. coli* (EHEC) and adenovirus (HAV) ingestion fell below the Recreational Water Quality Criteria's 36 illnesses in 1,000 swimmers for the predicted time points (U.S. EPA, 2012).

In the U.S. alone, it is estimated that waterborne illnesses incur \$3.33 billion in direct healthcare costs annually (Collier et al., 2021), and the most recent assessments indicate that pathogens are a leading cause of impairments for the nation's rivers, streams, coastal waters, bays and estuaries (U.S. EPA, 2017). Globally, diarrheal diseases have been reported to be

responsible for 57 million disability adjusted life years (DALYs), and 57% of those diseases are attributable to the environment (Prüss-Ustün et al., 2016). The work completed within this dissertation i) reduces the uncertainty associated with how pathogen and indicator persistence is modeled and predicted in surface waters; ii) adds knowledge to the indicator-pathogen paradigm with the calculation of uncertainty factors for fecal indicator bacteria, bacteriophages, pathogenic bacteria, viruses, and protozoa, and iii) facilitates more accurate estimates of indicator and pathogen persistence to help advance the regulation of our waters to better protect human health. The results presented herein emphasize the importance of considering persistence within surface water decision-making and quantitative microbial risk assessments, as a sole reliance on indicator data may lead to erroneous assessments of water quality and risk. Results from this work will inform decision makers advancing water, sanitation, hygiene, water reuse applications, and agricultural practices globally.

## REFERENCES

- Collier, S. A., Deng, L., Adam, E. A., Benedict, K. M., Beshearse, E. M., Blackstock, A. J., Bruce, B. B., Derado, G., Edens, C., Fullerton, K. E., Gargano, J. W., Geissler, A. L., Hall, A. J., Havelaar, A. H., Hill, V. R., Hoekstra, R. M., Reddy, S. C., Scallan, E., Stokes, E. K., ... Beach, M. J. (2021). Estimate of Burden and Direct Healthcare Cost of Infectious Waterborne Disease in the United States. *Emerging Infectious Diseases*, 27(1), 140–149. <https://doi.org/10.3201/eid2701.190676>
- Prüss-Ustün, A., Wolf, J., Corvalán, C., Bos, R., & Neira, M. (2016). *Preventing Disease Through Healthy Environments: A global assessment of the burden of disease from environmental risks*. World Health Organization. <https://apps.who.int/iris/handle/10665/345241>
- U.S. EPA. (2012). Recreational Water Quality Criteria. Office of Water 820-F-12-058. Retrieved from: <https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf>
- U.S. EPA. (2017). National Water Quality Inventory: Report to Congress. Retrieved from: [https://www.epa.gov/sites/default/files/2017-/documents/305brtc\\_finalowow\\_08302017.pdf](https://www.epa.gov/sites/default/files/2017-/documents/305brtc_finalowow_08302017.pdf)

## CHAPTER 7: FUTURE WORK

There are several opportunities for the work presented in this dissertation to be expanded upon in future works. The systematic literature review (Chapter 2) identified several obstacles to the mining of data from the literature for analysis. As such, the sharing of the rich database of experiments and persistence models generated in this work (Chapter 3) is a priority for future work. The MLE estimates and confidence intervals for nearly 500 datasets will be shared on a freely available community portal ([www.qmrawiki.org](http://www.qmrawiki.org)) to allow for their application in a variety of disciplines impacted by this work, including engineering, water treatment, wastewater treatment, microbiology, and risk assessment.

The general model for persistence developed herein (Chapter 4) although novel in nature, has opportunities for improvement. The current model form accurately estimates median values of the training and testing dataset T90s, however consistently underestimates the upper bounds of the observed T90s. To improve upon this performance, the inclusion of untested predictors such as turbidity, pH, or other site-specific characteristics within the general model form should be explored. Additionally, the current general model form was constructed with limited data for viruses and protozoa targets. The Bayesian hierarchical methodology minimizes this limitation by allowing the virus and protozoa datasets to gain knowledge from the fecal indicator bacteria, bacteria, and bacteriophage datasets, however as new data becomes available, the general model should continue to be adapted. This may be particularly relevant for protozoan targets as the systematic literature review suggested that protozoa were minimally affected by the water quality factors influencing other targets; the selection of population-level coefficients for temperature, predation, and water type either suggests that this is not the case, or that there was not enough protozoa data included in the general model to identify this deviation.

Another natural progression of this work is to assess the ability of the Bayesian hierarchical model to improve the fitting of the Juneja and Marks 2 (JM2) model to future datasets. As the hierarchical model incorporates the knowledge of prior datasets, it is expected that the uncertainty associated with parameter and model estimates will be reduced when comparing Bayesian fitting to maximum likelihood estimation (MLE) methods. Some preliminary data for this hypothesis is shown in Figure 7.1, where the parameter and model standard deviations from fitting the JM2 with Bayesian and MLE methods were compared for the culturable enterococci (cENT), enterohemorrhagic *Escherichia coli* (EHEC) and adenovirus (HAV) data from the case study in Chapter 5. Although this example dataset is limited in size, the standard deviations are more often lower with the Bayesian fitting than the MLE.

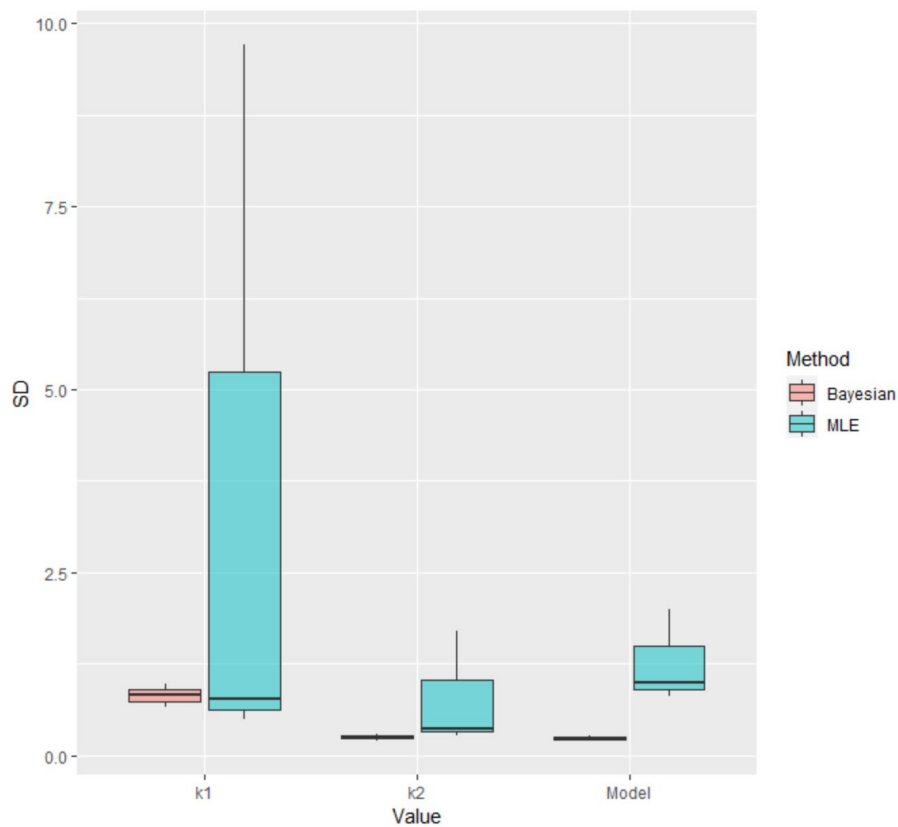


Figure 7.1: Range of parameter and model standard deviations using Bayesian or MLE methods to fit JM2 to cENT, EHEC, and HAV data (Chapter 5)



Finally, microbial source tracking (MST) is another tool that can be used to improve the indicator-pathogen paradigm for surface water decision making. Future works should evaluate the performance of JM2 on MST markers persistence data and a similar analysis evaluating factor-persistence relationships for MST markers could provide critical insights to inform the selection of MST markers for various monitoring efforts. Coupling the results of this dissertation with future MST marker analyses will facilitate the creation of decision support tools that incorporate site-specific considerations and pathogen concerns into the construction of monitoring protocols and practices.