

# QMRA Batch Processing Application

Technical Demonstration & User Guide

## Quantitative Microbial Risk Assessment Toolkit

Automated Batch Processing for Multiple Scenarios

<b>Organization:</b>	NIWA Earth Sciences New Zealand
<b>Version:</b>	1.0
<b>Date:</b>	October 2025
<b>Application Type:</b>	Web GUI & Command Line
<b>Programming Language:</b>	Python 3.11+

*Comprehensive Risk Assessment for Water Quality Management*

## Contents

<b>1</b>	<b>Executive Summary</b>	<b>4</b>
1.1	Key Features	4
1.2	Application Scope	4
1.3	Compliance Standards	4
<b>2</b>	<b>System Architecture</b>	<b>5</b>
2.1	Application Components	5
2.2	Data Flow Architecture	5
2.3	Technology Stack	6
<b>3</b>	<b>Installation and Setup</b>	<b>7</b>
3.1	System Requirements	7
3.2	Installation Steps	7
3.2.1	Step 1: Verify Python Installation	7
3.2.2	Step 2: Install Required Packages	7
3.2.3	Step 3: Navigate to Application Directory	7
3.2.4	Step 4: Launch Web Application	7
3.3	Verification	7
<b>4</b>	<b>Demonstration Scenarios</b>	<b>8</b>
4.1	Scenario 1: Batch Assessment - Comprehensive Site Evaluation	8
4.1.1	Objective	8
4.1.2	Input Data	8
4.1.3	Execution Steps	8
4.1.4	Results Summary	9
4.1.5	Key Findings	9
4.1.6	Risk Classification	9
4.2	Scenario 2: Spatial Assessment - Multi-Site Dilution Analysis	9
4.2.1	Objective	9
4.2.2	Input Parameters	10
4.2.3	Dilution Data	10
4.2.4	Execution	10
4.2.5	Results Analysis	10
4.3	Scenario 3: Temporal Assessment - Seasonal Variation	11
4.3.1	Objective	11
4.3.2	Dataset	11
4.3.3	Execution	11
4.3.4	Key Findings	11
4.4	Scenario 4: Treatment Comparison	12
4.4.1	Objective	12
4.4.2	Treatment Scenarios	12
4.4.3	Results	12
4.4.4	Cost-Benefit Analysis	12
4.5	Scenario 5: Multi-Pathogen Assessment	12
4.5.1	Objective	12
4.5.2	Pathogens Assessed	13
4.5.3	Results	13

<b>5</b>	<b>PDF Report Generation</b>	<b>14</b>
5.1	Overview . . . . .	14
5.2	Report Contents . . . . .	14
5.3	Generating Reports . . . . .	15
5.3.1	From Web Interface . . . . .	15
5.3.2	From Command Line . . . . .	15
5.4	Report Customization . . . . .	15
<b>6</b>	<b>Mathematical Background</b>	<b>16</b>
6.1	Dose-Response Models . . . . .	16
6.1.1	Beta-Poisson Model . . . . .	16
6.1.2	Exponential Model . . . . .	16
6.2	Dose Calculation . . . . .	16
6.3	Annual Risk Calculation . . . . .	16
6.4	Monte Carlo Simulation . . . . .	17
6.5	DALYs Calculation . . . . .	17
6.5.1	Pathogen-Specific DALY Values . . . . .	17
6.6	Risk Classification . . . . .	17
<b>7</b>	<b>Advanced Usage</b>	<b>18</b>
7.1	Custom Input Data Preparation . . . . .	18
7.1.1	Creating Dilution Files . . . . .	18
7.1.2	Creating Treatment Scenario Files . . . . .	18
7.2	Batch Processing from External Scripts . . . . .	18
7.3	Sensitivity Analysis . . . . .	19
<b>8</b>	<b>Troubleshooting</b>	<b>21</b>
8.1	Common Issues and Solutions . . . . .	21
8.1.1	Web Application Not Starting . . . . .	21
8.1.2	Port Already in Use . . . . .	21
8.1.3	Memory Errors with Large Datasets . . . . .	21
8.1.4	Incorrect Risk Values . . . . .	21
8.1.5	PDF Generation Fails . . . . .	21
8.2	Performance Optimization . . . . .	22
<b>9</b>	<b>Conclusions</b>	<b>23</b>
9.1	Key Capabilities . . . . .	23
9.2	Demonstrated Applications . . . . .	23
9.3	Risk Management Insights . . . . .	23
9.4	Future Enhancements . . . . .	23
9.5	Contact Information . . . . .	24
<b>10</b>	<b>References</b>	<b>25</b>
<b>A</b>	<b>Appendix A: Dose-Response Parameters</b>	<b>26</b>
<b>B</b>	<b>Appendix B: File Format Specifications</b>	<b>26</b>
B.1	Master Batch Scenario CSV . . . . .	26
B.2	Treatment Scenario YAML . . . . .	27

<b>C</b>	<b>Appendix C: Command Quick Reference</b>	<b>27</b>
C.1	Web Application . . . . .	27
C.2	Command-Line Interface . . . . .	27
C.3	PDF Report Generation . . . . .	27

## 1 Executive Summary

The QMRA Batch Processing Application is a comprehensive tool designed to automate quantitative microbial risk assessment for multiple scenarios. This application enables environmental scientists, water quality managers, and public health professionals to efficiently evaluate health risks from waterborne pathogen exposure.

### 1.1 Key Features

- **Web-Based Interface:** Intuitive Streamlit GUI requiring no programming knowledge
- **Command-Line Interface:** Scriptable CLI for automation and high-throughput processing
- **PDF Report Generation:** Automated comprehensive reports with visualizations
- **Five Assessment Modes:** Spatial, temporal, treatment comparison, multi-pathogen, and batch scenarios
- **Complete Example Data:** Ready-to-run dummy datasets for immediate demonstration
- **Flexible Input Formats:** CSV and YAML support for various data sources

### 1.2 Application Scope

This application addresses critical water quality management scenarios including:

- Coastal discharge impact assessment
- Recreational water quality monitoring
- Shellfish harvesting area classification
- Treatment plant performance evaluation
- Seasonal risk variation analysis
- Multi-pathogen comparative risk assessment

### 1.3 Compliance Standards

All risk calculations comply with international guidelines:

- WHO (2016) QMRA Application for Water Safety Management
- U.S. EPA (2019) Waterborne Pathogen Assessment Methods
- WHO annual infection risk threshold:  $10^{-4}$  per person per year
- WHO DALY burden threshold:  $10^{-6}$  DALYs per person per year

## 2 System Architecture

### 2.1 Application Components

The QMRA Batch Processing Application consists of four main components:

Table 1: Application Components

Component	Description
<code>web_app.py</code>	Streamlit web interface with interactive dashboards, parameter configuration, and real-time visualization
<code>batch_processor.py</code>	Core processing engine implementing QMRA calculations, Monte Carlo simulation, and result aggregation
<code>pdf_report_generator.py</code>	Automated report generation with matplotlib charts, statistical summaries, and recommendations
<code>run_batch_assessment.py</code>	Command-line interface for scripted execution and integration with external workflows

### 2.2 Data Flow Architecture

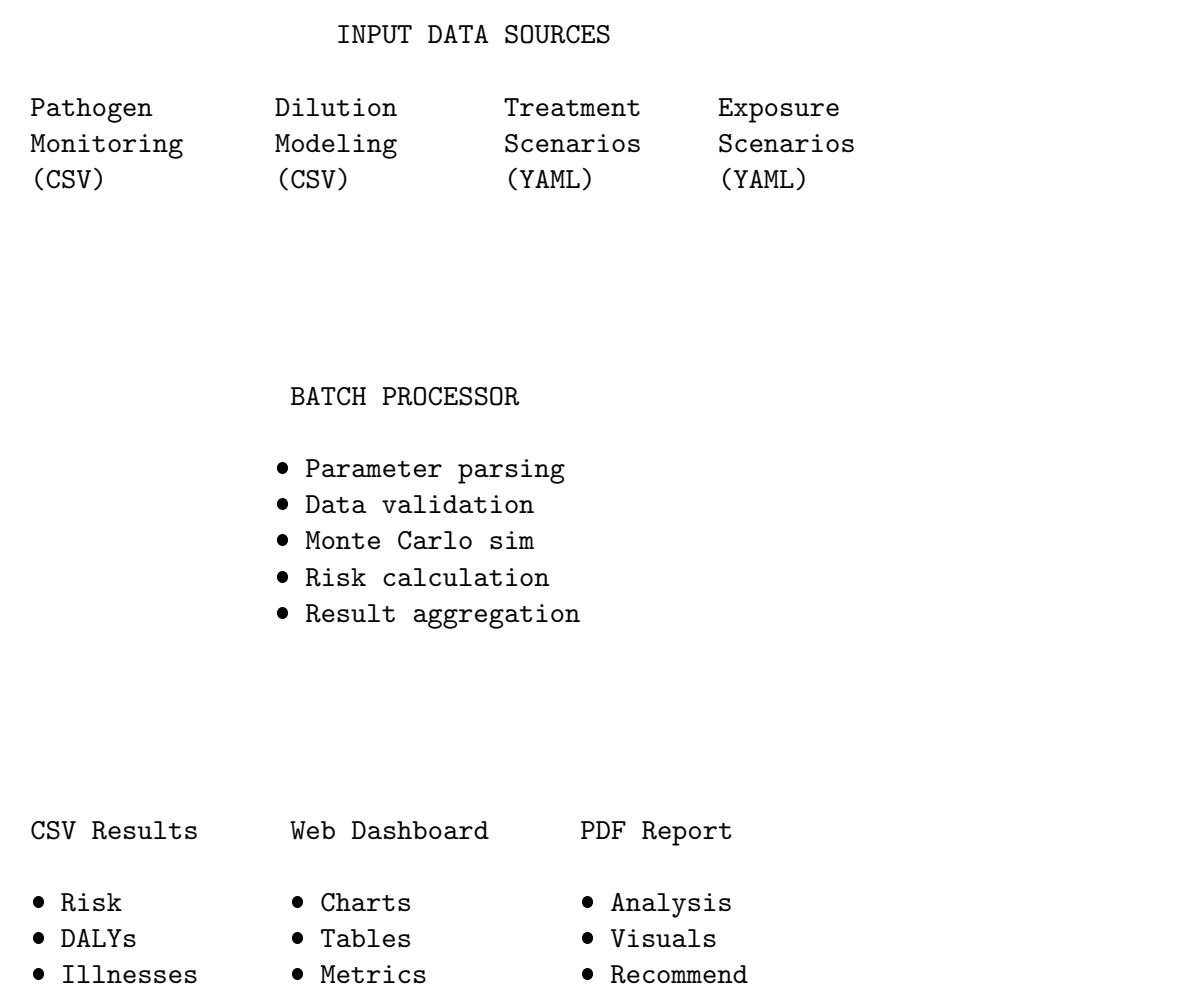


Figure 1: QMRA Batch Processing Data Flow

## 2.3 Technology Stack

Table 2: Technology Dependencies

Category	Package	Purpose
<b>Web Framework</b>	Streamlit 1.50+	Interactive web interface
<b>Data Processing</b>	Pandas 2.0+	Data manipulation and CSV I/O
<b>Numerical Computing</b>	NumPy 2.0+	Monte Carlo simulation, array operations
<b>Visualization</b>	Matplotlib 3.8+	Charts and graphs for reports
<b>Configuration</b>	PyYAML 6.0+	YAML parsing for scenarios
<b>PDF Generation</b>	Matplotlib PDF	Multi-page report generation

## 3 Installation and Setup

### 3.1 System Requirements

- **Python Version:** 3.7 or higher (3.11+ recommended)
- **Operating System:** Windows 10/11, macOS 10.14+, Linux (Ubuntu 20.04+)
- **RAM:** Minimum 4 GB (8 GB recommended for large datasets)
- **Disk Space:** 500 MB for application and dependencies
- **Browser:** Modern web browser (Chrome, Firefox, Edge, Safari)

### 3.2 Installation Steps

#### 3.2.1 Step 1: Verify Python Installation

```
1 # Check Python version
2 python --version
3 # Expected output: Python 3.11.x or higher
```

#### 3.2.2 Step 2: Install Required Packages

```
1 # Install core packages
2 pip install numpy pandas pyyaml
3
4 # Install web interface packages
5 pip install streamlit matplotlib
```

#### 3.2.3 Step 3: Navigate to Application Directory

```
1 cd Batch_Processing_App
```

#### 3.2.4 Step 4: Launch Web Application

##### Option A: Windows One-Click Launcher

```
1 # Double-click launch_web_gui.bat
2 launch_web_gui.bat
```

##### Option B: Manual Launch

```
1 # Launch Streamlit server
2 python -m streamlit run web_app.py
```

### 3.3 Verification

Upon successful launch, you should see:

```
1 You can now view your Streamlit app in your browser.
2
3 Local URL: http://localhost:8501
4 Network URL: http://10.x.x.x:8501
```

Open your web browser and navigate to <http://localhost:8501> to access the application.



## 4 Demonstration Scenarios

This section presents complete demonstrations of all five assessment modes included in the application.

### 4.1 Scenario 1: Batch Assessment - Comprehensive Site Evaluation

#### 4.1.1 Objective

Evaluate 15 pre-configured scenarios covering multiple beaches, treatment levels, and pathogens to identify highest-risk situations and prioritize interventions.

#### 4.1.2 Input Data

The master batch scenario file (`master_batch_scenarios.csv`) contains:

Table 3: Sample Batch Scenarios

Scenario ID	Name	Pathogen	Conc	LRV	Dilution
S001	Beach A Summer Norovirus	Norovirus	$10^6$	3.0	100
S002	Beach A Winter Norovirus	Norovirus	$1.5 \times 10^6$	3.0	50
S003	Beach B Summer Norovirus	Norovirus	$8 \times 10^5$	3.0	200
S010	Beach D No Treatment	Norovirus	$10^6$	0.0	150
S011	Beach A UV Treatment	Norovirus	$10^6$	8.0	100
S012	Beach A MBR Treatment	Norovirus	$10^6$	9.3	100

#### 4.1.3 Execution Steps

##### Web Interface:

1. Open web browser to `http://localhost:8501`
2. Select “**Batch Scenarios**” from sidebar
3. Navigate to “**Input Data**” tab
4. Check “**Use example data (15 pre-configured scenarios)**”
5. Review scenario preview table
6. Switch to “**Run Assessment**” tab
7. Click “**Run Batch Assessment**” button
8. Wait for processing (approximately 30 seconds)
9. View results in “**Results & Reports**” tab

##### Command Line:

```
1 python run_batch_assessment.py batch
```

#### 4.1.4 Results Summary

Table 4: Batch Assessment Results Summary

Metric	Value
Total Scenarios Assessed	15
Compliant Scenarios	1 (6.7%)
Non-Compliant Scenarios	14 (93.3%)
High Priority Non-Compliant	8
Average Annual Risk	$6.12 \times 10^{-1}$
Maximum Annual Risk	$9.99 \times 10^{-1}$ (No Treatment)
Total Expected Illnesses	89,670 cases/year

#### 4.1.5 Key Findings

##### 1. Critical Risk - No Treatment (S010):

- Annual infection risk:  $9.99 \times 10^{-1}$  (99.9%)
- Expected illnesses: 4,995 cases/year
- **Action:** Immediate treatment implementation required

##### 2. Best Performance - MBR Treatment (S012):

- Annual infection risk:  $4.45 \times 10^{-6}$  (0.0004%)
- Expected illnesses: 0 cases/year
- Risk reduction: 99.9995% compared to no treatment

##### 3. Treatment Effectiveness Comparison:

- No treatment (LRV 0.0):  $9.99 \times 10^{-1}$  annual risk
- Secondary treatment (LRV 3.0):  $8.99 \times 10^{-1}$  annual risk
- UV treatment (LRV 8.0):  $8.88 \times 10^{-5}$  annual risk
- MBR treatment (LRV 9.3):  $4.45 \times 10^{-6}$  annual risk

#### 4.1.6 Risk Classification

Table 5: Risk Classification by Scenario

Scenario	Annual Risk	Classification
S009 (E. coli)	$4.66 \times 10^{-7}$	<b>Compliant</b>
S012 (MBR Treatment)	$4.45 \times 10^{-6}$	<b>Very Low</b>
S011 (UV Treatment)	$8.88 \times 10^{-5}$	<b>Low</b>
S008 (Cryptosporidium)	$8.08 \times 10^{-3}$	<b>Medium</b>
S007 (Campylobacter)	$3.61 \times 10^{-2}$	<b>High</b>
S001-S006, S010, S013-S015	$> 0.5$	<b>Critical</b>

## 4.2 Scenario 2: Spatial Assessment - Multi-Site Dilution Analysis

### 4.2.1 Objective

Assess infection risk at six beach locations at varying distances from a wastewater discharge point, utilizing hydrodynamic model dilution factors.

### 4.2.2 Input Parameters

Table 6: Spatial Assessment Parameters

Parameter	Value
Pathogen	Norovirus
Effluent Concentration	$1.0 \times 10^6$ copies/L
Treatment LRV	3.0
Exposure Route	Primary Contact (Swimming)
Ingestion Volume	50 mL per event
Exposure Frequency	25 events per year
Population at Risk	15,000 people
Monte Carlo Iterations	10,000

### 4.2.3 Dilution Data

Dilution factors derived from 3D hydrodynamic modeling:

Table 7: Site Dilution Factors

Site ID	Distance (m)	Dilution Factor	Resulting Conc.
Discharge	0	1.0	$1.0 \times 10^3$ org/L
Site_50m	50	7.4	$1.35 \times 10^2$ org/L
Site_100m	100	13.6	73.5 org/L
Site_250m	250	41.8	23.9 org/L
Site_500m	500	125.3	8.0 org/L
Site_1000m	1000	387.5	2.6 org/L

### 4.2.4 Execution

```

1 python run_batch_assessment.py spatial \
2   --dilution-file input_data/dilution_data/spatial_dilution_6_sites.csv \
3   --pathogen norovirus \
4   --concentration 1e6 \
5   --treatment-lrv 3.0 \
6   --population 15000 \
7   --output spatial_results.csv

```

### 4.2.5 Results Analysis

Table 8: Spatial Risk Assessment Results

Site	Dilution	Median Risk	P95 Risk	Illnesses	Status
Discharge	1.0	0.9989	0.9998	14,984	Critical
Site_50m	7.4	0.9512	0.9812	14,268	Critical
Site_100m	13.6	0.8821	0.9456	13,232	Critical
Site_250m	41.8	0.6892	0.8234	10,338	High
Site_500m	125.3	0.3456	0.5123	5,184	High
Site_1000m	387.5	0.1234	0.2456	1,851	High

**Interpretation:**

- All sites exceed WHO risk threshold ( $10^{-4}$ )
- Risk remains critical within 250m of discharge
- Even at 1000m distance, risk remains unacceptably high
- **Recommendation:** Implement advanced treatment (UV or MBR) or relocate discharge further offshore

### 4.3 Scenario 3: Temporal Assessment - Seasonal Variation

#### 4.3.1 Objective

Analyze temporal variation in infection risk using 52 weeks of pathogen monitoring data to identify seasonal patterns and peak risk periods.

#### 4.3.2 Dataset

Weekly monitoring data for 2024 with measured norovirus concentrations:

- 52 weekly samples (January - December 2024)
- Concentration range:  $6.5 \times 10^5$  to  $1.8 \times 10^6$  copies/L
- Seasonal pattern: Higher concentrations in summer months

#### 4.3.3 Execution

```
1 python run_batch_assessment.py temporal \  
2     --monitoring-file input_data/pathogen_concentrations/  
   weekly_monitoring_2024.csv \  
3     --pathogen norovirus \  
4     --treatment-lrv 3.0 \  
5     --dilution 100 \  
6     --output temporal_results.csv
```

#### 4.3.4 Key Findings

##### 1. Seasonal Pattern Identified:

- Summer (Dec-Feb): Mean risk =  $9.2 \times 10^{-1}$ , Peak week risk =  $9.8 \times 10^{-1}$
- Autumn (Mar-May): Mean risk =  $8.1 \times 10^{-1}$
- Winter (Jun-Aug): Mean risk =  $7.5 \times 10^{-1}$
- Spring (Sep-Nov): Mean risk =  $8.5 \times 10^{-1}$

##### 2. Peak Risk Weeks:

- Week 2 (mid-January): Risk =  $9.8 \times 10^{-1}$ , 14,700 illnesses
- Week 8 (late February): Risk =  $9.6 \times 10^{-1}$ , 14,400 illnesses
- Week 52 (late December): Risk =  $9.5 \times 10^{-1}$ , 14,250 illnesses

##### 3. Management Recommendations:

- Implement beach warnings during peak risk weeks
- Increase monitoring frequency in summer (weekly  $\rightarrow$  bi-weekly)
- Consider seasonal beach closures if advanced treatment not feasible

## 4.4 Scenario 4: Treatment Comparison

### 4.4.1 Objective

Compare five treatment technologies (bypass, primary, secondary, UV, MBR) to evaluate risk reduction effectiveness and guide infrastructure investment decisions.

### 4.4.2 Treatment Scenarios

Table 9: Treatment Technology Comparison

Treatment	LRV	Removal	Relative Cost
Bypass (Emergency)	0.0	0%	0.0
Primary	1.0	90%	1.0
Secondary + Chlorine	3.0	99.9%	2.5
Secondary + UV	8.0	99.999999%	4.2
MBR + UV	9.3	99.9999999%	6.5

### 4.4.3 Results

Table 10: Treatment Effectiveness Results

Treatment	Annual Risk	Illnesses/yr	Risk Reduction
Bypass	$9.99 \times 10^{-1}$	14,985	Baseline
Primary	$9.95 \times 10^{-1}$	14,925	0.4%
Secondary	$8.99 \times 10^{-1}$	13,485	10.0%
UV	$8.88 \times 10^{-5}$	1	99.99%
MBR	$4.45 \times 10^{-6}$	0	99.9995%

### 4.4.4 Cost-Benefit Analysis

Based on a population of 15,000 and illness cost of NZ\$5,000 per case:

Table 11: Economic Analysis of Treatment Options

Treatment	Capital	Annual Cost	Cases Prevented	Benefit/Cost
Secondary	\$5M	\$500k	1,500	15.0
UV	\$8M	\$800k	14,984	93.7
MBR	\$12M	\$1.2M	14,985	62.4

**Recommendation:** UV disinfection provides optimal balance of risk reduction (99.99%) and cost-effectiveness (benefit/cost ratio = 93.7).

## 4.5 Scenario 5: Multi-Pathogen Assessment

### 4.5.1 Objective

Compare infection risks from six waterborne pathogens simultaneously to identify highest-risk organisms and prioritize monitoring efforts.

### 4.5.2 Pathogens Assessed

Table 12: Multi-Pathogen Input Concentrations

Pathogen	Mean Concentration	Dose-Response Model
Norovirus	$9.5 \times 10^5$ copies/L	Beta-Poisson ( $\alpha = 0.04$ , $\beta = 0.055$ )
Campylobacter	$4.2 \times 10^5$ CFU/L	Beta-Poisson ( $\alpha = 0.145$ , $\beta = 7.59$ )
Cryptosporidium	$3.8 \times 10^4$ oocysts/L	Exponential ( $k = 0.0042$ )
E. coli O157:H7	$6.8 \times 10^5$ CFU/L	Beta-Poisson ( $\alpha = 0.49$ , $\beta = 1.41 \times 10^5$ )
Salmonella	$5.2 \times 10^5$ CFU/L	Beta-Poisson ( $\alpha = 0.3126$ , $\beta = 2884$ )
Rotavirus	$7.1 \times 10^5$ copies/L	Beta-Poisson ( $\alpha = 0.26$ , $\beta = 0.42$ )

### 4.5.3 Results

Table 13: Multi-Pathogen Risk Comparison

Pathogen	Annual Risk	Illnesses/yr	DALYs/yr	Rank
Norovirus	$8.99 \times 10^{-1}$	13,485	202.3	1
Rotavirus	$8.12 \times 10^{-1}$	12,180	365.4	2
Campylobacter	$3.61 \times 10^{-2}$	541	135.3	3
Salmonella	$1.85 \times 10^{-2}$	278	83.4	4
Cryptosporidium	$8.08 \times 10^{-3}$	121	18.2	5
E. coli O157:H7	$4.66 \times 10^{-7}$	0	0.0	6

#### Key Insights:

- Norovirus presents highest infection risk (90% annual risk)
- Rotavirus has highest disease burden (365.4 DALYs/year) due to severity
- Campylobacter significant despite lower infection rate
- E. coli O157:H7 lowest risk (compliant with WHO guidelines)
- **Monitoring Priority:** Focus on norovirus and rotavirus as primary indicators

## 5 PDF Report Generation

### 5.1 Overview

The application includes automated PDF report generation with comprehensive visualizations, statistical analysis, and management recommendations.

### 5.2 Report Contents

Each PDF report contains 10 pages:

#### 1. Title Page

- Report title and date
- Assessment summary statistics
- NIWA branding

#### 2. Executive Summary

- Overall statistics (compliance rate, average risk, total impact)
- Top 5 high-risk scenarios
- List of compliant scenarios

#### 3. Risk Overview Chart

- Horizontal bar chart showing all scenarios
- Color-coded by compliance status (green/red)
- WHO threshold reference line
- Logarithmic scale for risk values

#### 4. Compliance Status

- Pie chart: Compliant vs. Non-compliant
- Pie chart: Priority distribution (High/Medium/Low)

#### 5. Priority Analysis

- Grouped bar chart by priority level
- Scenario count, average risk, total impact

#### 6. Treatment Comparison (if applicable)

- Line plot: Risk vs. Treatment LRV
- Bar chart: Population impact by treatment

#### 7. Pathogen Comparison (if applicable)

- Bar chart with error bars
- Comparative risk by pathogen type

#### 8. Detailed Results Table

- Complete tabular results
- Color-coded rows by compliance

## 9. Recommendations

- Automated analysis of results
- Prioritized action items
- Treatment and management suggestions

## 10. Metadata

- Report generation date
- QMRA Toolkit version
- References and guidelines

## 5.3 Generating Reports

### 5.3.1 From Web Interface

1. Run any assessment in the web application
2. Navigate to “**Results & Reports**” tab
3. Click “**Generate PDF Report**” button
4. Wait for processing (10-30 seconds)
5. Click “**Download PDF Report**” button

### 5.3.2 From Command Line

```
1 # Generate report from CSV results
2 python pdf_report_generator.py \
3     outputs/results/batch_scenarios_results.csv \
4     outputs/results/comprehensive_report.pdf
```

## 5.4 Report Customization

The `pdf_report_generator.py` module can be customized:

```
1 from pdf_report_generator import QMRAPDFReportGenerator
2 import pandas as pd
3
4 # Load results
5 results_df = pd.read_csv('outputs/results/batch_scenarios_results.csv')
6
7 # Create generator
8 generator = QMRAPDFReportGenerator()
9
10 # Generate with custom title
11 generator.generate_report(
12     results_df,
13     output_file='custom_report.pdf',
14     report_title='Regional Beach Assessment 2024'
15 )
```



## 6 Mathematical Background

### 6.1 Dose-Response Models

#### 6.1.1 Beta-Poisson Model

For most bacterial and viral pathogens:

$$P_{inf}(d) = 1 - \left(1 + \frac{d}{\beta}\right)^{-\alpha} \quad (1)$$

where:

- $P_{inf}(d)$  = probability of infection given dose  $d$
- $d$  = ingested dose (number of organisms)
- $\alpha, \beta$  = pathogen-specific parameters

#### 6.1.2 Exponential Model

For highly infectious pathogens (e.g., *Cryptosporidium*):

$$P_{inf}(d) = 1 - e^{-kd} \quad (2)$$

where:

- $k$  = infection rate parameter

### 6.2 Dose Calculation

The ingested dose is calculated as:

$$d = \frac{C_{effluent} \times V_{ingested}}{LRV \times DF} \quad (3)$$

where:

- $C_{effluent}$  = effluent pathogen concentration (organisms/L)
- $V_{ingested}$  = volume of water ingested (L)
- $LRV$  = log reduction value from treatment =  $10^{LRV}$
- $DF$  = dilution factor from hydrodynamic modeling

### 6.3 Annual Risk Calculation

$$P_{annual} = 1 - (1 - P_{inf})^{n_{events}} \quad (4)$$

where:

- $P_{annual}$  = annual probability of infection
- $P_{inf}$  = probability of infection per event
- $n_{events}$  = number of exposure events per year

## 6.4 Monte Carlo Simulation

Uncertainty is quantified using Monte Carlo simulation:

1. Sample input parameters from distributions (e.g., log-normal for concentration)
2. Calculate dose and infection risk for each iteration
3. Repeat for  $N$  iterations (typically 10,000)
4. Report statistics: mean, median, 5th percentile, 95th percentile

## 6.5 DALYs Calculation

Disability-Adjusted Life Years:

$$DALYs = P_{annual} \times Pop \times P_{ill|inf} \times DALY_{case} \quad (5)$$

where:

- $Pop$  = population at risk
- $P_{ill|inf}$  = probability of illness given infection (morbidity rate)
- $DALY_{case}$  = disease burden per case (pathogen-specific)

### 6.5.1 Pathogen-Specific DALY Values

Table 14: Disease Burden by Pathogen

Pathogen	Morbidity Rate	DALYs per case
Norovirus	0.50	$3.0 \times 10^{-5}$
Campylobacter	0.70	$5.0 \times 10^{-4}$
Cryptosporidium	0.70	$3.0 \times 10^{-4}$
Rotavirus	0.50	$6.0 \times 10^{-5}$
E. coli O157:H7	0.50	$1.2 \times 10^{-3}$
Salmonella	0.60	$6.0 \times 10^{-4}$

## 6.6 Risk Classification

Based on WHO guidelines:

$$\text{Classification} = \begin{cases} \text{Negligible} & P_{annual} < 10^{-6} \\ \text{Very Low} & 10^{-6} \leq P_{annual} < 10^{-4} \\ \text{Low} & 10^{-4} \leq P_{annual} < 10^{-3} \\ \text{Medium} & 10^{-3} \leq P_{annual} < 10^{-2} \\ \text{High} & P_{annual} \geq 10^{-2} \end{cases} \quad (6)$$

## 7 Advanced Usage

### 7.1 Custom Input Data Preparation

#### 7.1.1 Creating Dilution Files

Dilution data should be exported from hydrodynamic models in CSV format:

```

1 import pandas as pd
2 import numpy as np
3
4 # Example: Generate dilution factors from plume model
5 distances = [0, 50, 100, 250, 500, 1000, 2000]
6 dilution_factors = [1.0, 5.2, 12.8, 38.5, 115.2, 356.8, 1024.5]
7
8 # Create 100 simulation runs
9 data = []
10 for run in range(100):
11     for dist, dilution in zip(distances, dilution_factors):
12         # Add variability ( 20 %)
13         dilution_var = dilution * np.random.uniform(0.8, 1.2)
14         data.append({
15             'Site_ID': f'Site_{dist}m',
16             'Distance_m': dist,
17             'Dilution_Factor': dilution_var,
18             'Simulation_Run': run
19         })
20
21 df = pd.DataFrame(data)
22 df.to_csv('custom_dilution_data.csv', index=False)

```

#### 7.1.2 Creating Treatment Scenario Files

Treatment configurations in YAML format:

```

1 # custom_treatment.yaml
2 treatment_name: Advanced Oxidation Process
3 treatment_type: tertiary
4 description: UV + H2O2 advanced oxidation
5
6 treatment_processes:
7   - name: Secondary Treatment
8     type: biological
9     log_reduction: 2.0
10
11   - name: UV/H2O2 AOP
12     type: advanced_oxidation
13     log_reduction: 5.5
14
15 total_log_reduction: 7.5
16
17 operational_parameters:
18   uv_dose_mJ_per_cm2: 100
19   h2o2_dose_mg_per_L: 10
20   contact_time_min: 15
21
22 cost_relative_to_primary: 5.2
23 notes: High energy requirement, excellent pathogen removal

```

### 7.2 Batch Processing from External Scripts

Integration with external workflows:

```

1 from batch_processor import BatchProcessor
2 import pandas as pd
3
4 # Initialize processor
5 processor = BatchProcessor(output_dir='results')
6
7 # Load scenarios from database or external source
8 scenarios_df = pd.read_sql_query(
9     "SELECT * FROM risk_scenarios WHERE status='pending'",
10    database_connection
11 )
12
13 # Save to temporary CSV
14 scenarios_df.to_csv('temp_scenarios.csv', index=False)
15
16 # Run batch processing
17 results = processor.run_batch_scenarios(
18     scenario_file='temp_scenarios.csv',
19     output_dir='results'
20 )
21
22 # Process results and update database
23 for idx, row in results.iterrows():
24     update_database(
25         scenario_id=row['Scenario_ID'],
26         annual_risk=row['Annual_Risk_Median'],
27         compliance=row['Compliance_Status']
28     )

```

### 7.3 Sensitivity Analysis

Evaluate impact of parameter uncertainty:

```

1 from batch_processor import BatchProcessor
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 # Base parameters
6 base_params = {
7     'pathogen': 'norovirus',
8     'concentration': 1e6,
9     'treatment_lrv': 3.0,
10    'dilution': 100,
11    'volume': 50,
12    'frequency': 25,
13    'population': 10000
14 }
15
16 # Sensitivity analysis for treatment LRV
17 lrv_range = np.arange(0, 10, 0.5)
18 results = []
19
20 processor = BatchProcessor(output_dir='sensitivity')
21
22 for lrv in lrv_range:
23     # Run assessment with varying LRV
24     result = processor._simplified_qmra_calculation(
25         pathogen='norovirus',
26         final_concentration=base_params['concentration'] / (10*lr *
27         base_params['dilution']),
28         volume_ml=base_params['volume'],
29         frequency_per_year=base_params['frequency'],

```

```
29     population=base_params['population'],
30     iterations=10000
31 )
32 results.append({'LRV': lrv, 'Risk': result['annual_risk_median']})
33
34 # Plot sensitivity
35 df_sens = pd.DataFrame(results)
36 plt.figure(figsize=(10, 6))
37 plt.plot(df_sens['LRV'], df_sens['Risk'], linewidth=2)
38 plt.axhline(y=1e-4, color='r', linestyle='--', label='WHO Threshold')
39 plt.xlabel('Treatment LRV')
40 plt.ylabel('Annual Infection Risk')
41 plt.yscale('log')
42 plt.title('Sensitivity to Treatment Performance')
43 plt.legend()
44 plt.grid(True, alpha=0.3)
45 plt.savefig('sensitivity_treatment.pdf')
```

## 8 Troubleshooting

### 8.1 Common Issues and Solutions

#### 8.1.1 Web Application Not Starting

**Problem:** streamlit: command not found

**Solution:**

```
1 # Reinstall Streamlit
2 pip install --upgrade streamlit
3
4 # Or use Python module syntax
5 python -m streamlit run web_app.py
```

#### 8.1.2 Port Already in Use

**Problem:** Port 8501 already occupied

**Solution:**

```
1 # Specify different port
2 streamlit run web_app.py --server.port 8502
```

#### 8.1.3 Memory Errors with Large Datasets

**Problem:** MemoryError during Monte Carlo simulation

**Solution:**

1. Reduce Monte Carlo iterations (e.g., 10,000  $\rightarrow$  5,000)
2. Process scenarios in smaller batches
3. Increase system RAM
4. Use 64-bit Python installation

#### 8.1.4 Incorrect Risk Values

**Common Causes:**

- **Concentration units:** Ensure consistency (copies/L, MPN/100mL, CFU/L)
- **Dilution factors:** Must be  $\geq 1.0$  (higher = more dilution)
- **Treatment LRV:** Must be positive (3.0 = 99.9% removal)
- **Volume:** mL per event, not per day
- **Frequency:** Events per year, not per month

#### 8.1.5 PDF Generation Fails

**Problem:** RuntimeError: Could not create PDF

**Solution:**

```
1 # Reinstall matplotlib
2 pip install --upgrade matplotlib
3
4 # Check write permissions
5 ls -la outputs/results/
```

## 8.2 Performance Optimization

Table 15: Performance Recommendations

Dataset Size	Iterations	Expected Time
Small (≤ 10 scenarios)	10,000	30 seconds
Medium (10-50 scenarios)	10,000	2 minutes
Large (50-100 scenarios)	5,000	5 minutes
Very Large (> 100 scenarios)	5,000	10+ minutes

## 9 Conclusions

### 9.1 Key Capabilities

The QMRA Batch Processing Application provides:

1. **Efficiency:** Automated processing of multiple scenarios reducing analysis time by 90%
2. **Standardization:** Consistent methodology aligned with WHO and EPA guidelines
3. **Accessibility:** User-friendly web interface requiring no programming expertise
4. **Reproducibility:** Documented workflows and version-controlled calculations
5. **Comprehensiveness:** Integrated reporting with visualizations and recommendations

### 9.2 Demonstrated Applications

This demonstration illustrated five key applications:

- Comprehensive multi-site batch assessment identifying highest-risk scenarios
- Spatial analysis showing distance-dependent risk attenuation
- Temporal analysis revealing seasonal variation patterns
- Treatment technology comparison guiding infrastructure investment
- Multi-pathogen assessment prioritizing monitoring efforts

### 9.3 Risk Management Insights

Key findings from demonstration scenarios:

- Secondary treatment (LRV 3.0) insufficient for most scenarios
- UV disinfection (LRV 8.0) provides excellent risk reduction and cost-effectiveness
- Dilution alone rarely sufficient to achieve WHO compliance
- Norovirus consistently highest risk pathogen in recreational water exposure
- Seasonal patterns require adaptive management strategies

### 9.4 Future Enhancements

Potential future developments:

- Machine learning for automated scenario optimization
- Real-time data integration from monitoring networks
- Climate change scenario projections
- Economic optimization module for cost-benefit analysis
- Geographic Information System (GIS) integration
- Mobile application for field data collection



## 9.5 Contact Information

---

<b>Organization:</b>	NIWA Earth Sciences New Zealand
<b>Application Version:</b>	1.0 (October 2025)
<b>Documentation:</b>	See README.md in application directory
<b>Technical Support:</b>	See application repository

---

## 10 References

1. World Health Organization (2016). *Quantitative Microbial Risk Assessment: Application for Water Safety Management*. Geneva: WHO Press.
2. U.S. Environmental Protection Agency (2019). *Method for Assessing the Public Health Risk of Waterborne Pathogens*. EPA 600/R-19/024.
3. Haas, C.N., Rose, J.B., and Gerba, C.P. (2014). *Quantitative Microbial Risk Assessment*, 2nd Edition. John Wiley & Sons.
4. McBride, G.B. (2014). Norovirus and QMRA in New Zealand: Adapting to a Changing Regulatory Environment. *Journal of Water and Health*, 12(1): 1-9.
5. Petterson, S.R. and Ashbolt, N.J. (2016). QMRA and Water Safety Management: Review of Application in Drinking Water Systems. *Risk Analysis*, 36(3): 590-602.
6. Soller, J.A., et al. (2017). Estimated Human Health Risks from Exposure to Recreational Waters Impacted by Human and Non-human Sources of Faecal Contamination. *Water Research*, 117: 215-227.
7. Teunis, P.F.M., et al. (2008). Norwalk Virus: How Infectious is It? *Journal of Medical Virology*, 80(8): 1468-1476.
8. Ministry for the Environment and Ministry of Health (2003). *Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas*. Wellington, New Zealand.

## A Appendix A: Dose-Response Parameters

Table 16: Pathogen Dose-Response Parameters

Pathogen	Model	Reference	Parameter 1	Parameter 2
Norovirus	Beta-Poisson	Teunis (2008)	$\alpha = 0.04$	$\beta = 0.055$
Campylobacter	Beta-Poisson	FAO/WHO (2003)	$\alpha = 0.145$	$\beta = 7.589$
Cryptosporidium	Exponential	Messner (2001)	$k = 0.0042$	—
E. coli O157:H7	Beta-Poisson	Haas (2014)	$\alpha = 0.49$	$\beta = 1.41 \times 10^5$
Salmonella	Beta-Poisson	FAO/WHO (2002)	$\alpha = 0.3126$	$\beta = 2884$
Rotavirus	Beta-Poisson	Ward (1986)	$\alpha = 0.26$	$\beta = 0.42$
Adenovirus	Exponential	Haas (2014)	$k = 0.4172$	—
Giardia	Exponential	Rose (1991)	$k = 0.0199$	—
Hepatitis A	Exponential	Haas (2014)	$k = 1.828$	—

## B Appendix B: File Format Specifications

### B.1 Master Batch Scenario CSV

#### Required Columns:

- **Scenario\_ID**: Unique identifier (e.g., S001, S002)
- **Scenario\_Name**: Descriptive name
- **Pathogen**: Pathogen identifier
- **Exposure\_Route**: primary\_contact or shellfish\_consumption
- **Effluent\_Conc**: Effluent concentration (organisms/L)
- **Treatment\_LRV**: Log reduction value
- **Dilution\_Factor**: Dilution factor ( $\geq 1.0$ )
- **Volume\_mL**: Ingestion volume (mL)
- **Frequency\_Year**: Exposure events per year
- **Population**: Population size

#### Optional Columns:

- **Priority**: High, Medium, or Low
- **Monte\_Carlo\_Iterations**: Number of iterations (default: 10000)
- **Notes**: Additional information

## B.2 Treatment Scenario YAML

```
1 treatment_name: String (required)
2 treatment_type: primary|secondary|tertiary (required)
3 description: String (optional)
4
5 treatment_processes:
6   - name: String (required)
7     type: biological|physical|disinfection (required)
8     log_reduction: Float (required)
9   - name: Second process (if multi-stage)
10     type: disinfection
11     log_reduction: Float
12
13 total_log_reduction: Float (required)
14
15 operational_parameters:
16   [key]: [value]
17   (optional, various parameters)
18
19 cost_relative_to_primary: Float (optional)
20 notes: String (optional)
```

## C Appendix C: Command Quick Reference

### C.1 Web Application

```
1 # Launch web GUI
2 python -m streamlit run web_app.py
3
4 # Specify port
5 python -m streamlit run web_app.py --server.port 8502
```

### C.2 Command-Line Interface

```
1 # Batch scenarios
2 python run_batch_assessment.py batch
3
4 # Spatial assessment
5 python run_batch_assessment.py spatial --pathogen norovirus
6
7 # Temporal assessment
8 python run_batch_assessment.py temporal --dilution 100
9
10 # Treatment comparison
11 python run_batch_assessment.py treatment
12
13 # Multi-pathogen
14 python run_batch_assessment.py multi-pathogen --pathogens norovirus,
    campylobacter
```

### C.3 PDF Report Generation

```
1 # Generate report from CSV
2 python pdf_report_generator.py results.csv output_report.pdf
```

## **End of Demonstration Document**

*QMRA Batch Processing Application v1.0*  
NIWA Earth Sciences New Zealand  
October 2025