**QMRA Batch Processing Application**

Technical User Guide & Demonstration

New Zealand Institute of Water and Atmospheric Research (NIWA)  
Earth Sciences Division

*October 2025*

Version: 2.0 (Simplified Three-File Approach)  
Language: Python 3.8+  
Framework: Streamlit Web Application  
Status: Production Ready

# Table of Contents

* 1. Executive Summary
* 2. Application Overview
* 3. System Architecture
* 4. Installation & Setup
* 5. The Three-File Input Approach
* 6. Batch Processing Workflow
* 7. Configuration & Parameters
* 8. Running Your First Assessment
* 9. Interpreting Results
* 10. Advanced Features
* 11. Troubleshooting & Best Practices
* 12. Example Scenarios
* 13. Frequently Asked Questions

# 1. Executive Summary

The QMRA Batch Processing Application is a professional-grade web-based tool for conducting Quantitative Microbial Risk Assessments (QMRA). It enables water quality managers and environmental scientists to evaluate health risks from waterborne pathogenic contamination across multiple scenarios, locations, and treatment options.  
  
This application integrates advanced statistical methods including Monte Carlo simulation, empirical distributions, and dose-response modeling to provide comprehensive uncertainty quantification. The simplified three-file input approach makes it accessible to users without advanced programming skills.  
  
Key capabilities include batch processing of 100+ scenarios, multi-site spatial analysis, temporal trend assessment, treatment comparison, and professional visualization and reporting.

# 2. Application Overview

## 2.1 Purpose & Scope

The QMRA Batch Processing App automates Quantitative Microbial Risk Assessment for water quality scenarios. It answers the question: "What is the probability of waterborne disease infection based on water quality, treatment, and exposure parameters?"

## 2.2 Core Features

* Batch Scenario Processing - Run 100+ scenarios automatically
* Empirical Distributions - ECDF for dilution, Hockey Stick for pathogens
* Monte Carlo Simulation - 10,000 iterations with uncertainty quantification
* Multiple Assessment Modes - Batch, spatial, temporal, treatment comparison
* Interactive Web Dashboard - Streamlit-based user interface
* Professional Visualizations - Risk overview, compliance, impact assessment
* Multiple Export Formats - CSV, Excel, PDF, PNG images
* Standalone Design - No external dependencies, portable

## 2.3 Supported Pathogens

* Norovirus
* Campylobacter
* Cryptosporidium
* E. coli
* Rotavirus
* Salmonella

# 3. System Architecture

## 3.1 Main Components

|  |  |
| --- | --- |
| Component | Description |
| web\_app.py | Streamlit web interface with dashboard |
| batch\_processor.py | Core QMRA processing engine |
| qmra\_core/ | Monte Carlo, dose-response, pathogen database |
| input\_data/ | CSV data files and examples |

# 4. Installation & Setup

## 4.1 System Requirements

|  |  |
| --- | --- |
| Requirement | Specification |
| Python | 3.8 or higher |
| RAM | 4 GB minimum, 16 GB recommended |
| Disk Space | 200 MB for application |
| OS | Windows, macOS, or Linux |
| Browser | Any modern browser |

## 4.2 Quick Installation

Step 1: Navigate to Application Directory  
cd "path/to/Batch\_Processing\_App"  
  
Step 2: Create Virtual Environment (recommended)  
python -m venv venv  
source venv/bin/activate  
  
Step 3: Install Dependencies  
pip install -r requirements.txt  
  
Step 4: Launch Application  
streamlit run web\_app.py  
  
The application will open at <http://localhost:8502>

# 5. The Three-File Input Approach

The application uses three separate CSV files: Dilution Data, Pathogen Data, and Scenario Definitions. This separation enables flexible scenario creation while using all available data for uncertainty analysis.

## 5.1 File 1: Dilution Data

### Purpose

Contains raw dilution factors from hydrodynamic models. For each location, observations create an empirical distribution (ECDF).

### Columns

Time: When measurement occurred (YYYY-MM-DD format)  
Location: Site identifier (must match scenarios.csv)  
Dilution\_Factor: Wastewater dilution ratio (e.g., 115 = 1:115 dilution)

### Example

Time,Location,Dilution\_Factor  
2024-01-01,Site\_A,120  
2024-01-02,Site\_A,95  
2024-01-03,Site\_A,130

## 5.2 File 2: Pathogen Data

### Purpose

Defines pathogen concentrations using Hockey Stick distribution. You specify minimum, median, and maximum concentrations.

### Columns

Pathogen\_ID: Unique identifier  
Pathogen\_Name: Descriptive name  
Pathogen\_Type: Type (norovirus, campylobacter, etc.)  
Min\_Concentration: Minimum concentration (organisms per 100 mL)  
Median\_Concentration: Typical concentration  
Max\_Concentration: Maximum concentration

### Example

Pathogen\_ID,Pathogen\_Name,Pathogen\_Type,Min\_Concentration,Median\_Concentration,Max\_Concentration  
PATH001,Norovirus\_Summer,norovirus,500000,1000000,2000000

## 5.3 File 3: Scenario Definitions

### Purpose

Combines pathogen and location data with exposure parameters. Each row is one assessment scenario.

### Key Columns

Scenario\_ID, Pathogen\_ID, Location, Exposure\_Route, Treatment\_LRV, Ingestion\_Volume\_mL, Exposure\_Frequency\_per\_Year, Exposed\_Population, Priority, Notes

# 6. Batch Processing Workflow

## 6.1 Step-by-Step Process

1. PREPARE DATA - Create three CSV files with your data  
  
2. LAUNCH APPLICATION - Windows: Double-click launch\_web\_gui.bat  
  
3. UPLOAD FILES - Select "Batch Scenario Processing" from sidebar  
  
4. PREVIEW DATA - Verify columns, locations, and pathogens match  
  
5. CONFIGURE - Set Monte Carlo iterations (10,000 default)  
  
6. RUN ASSESSMENT - Click "Run Batch Assessment" and wait 2-3 minutes  
  
7. VIEW RESULTS - See all scenarios in results table  
  
8. VIEW VISUALIZATIONS - Risk overview, compliance, distribution, impact  
  
9. DOWNLOAD RESULTS - Choose CSV, PDF, or ZIP format  
  
10. ANALYZE & REPORT - Open results in Excel or share visualizations

# 7. Configuration & Parameters

## 7.1 Monte Carlo Iterations

More iterations provide higher precision but increase runtime. Standard practice uses 10,000 iterations. For quick testing, use 5,000. For critical decisions requiring maximum precision, use 50,000.

## 7.2 Treatment Effectiveness (LRV)

|  |  |
| --- | --- |
| LRV Value | Treatment Type |
| 0 | No treatment |
| 1-2 | Primary treatment |
| 2-3 | Secondary treatment |
| 3-4 | Secondary + chlorination |
| 4-6 | Tertiary treatment |
| 6-8 | UV or membrane filtration |
| 8+ | Advanced treatment (MBR, RO, ozone) |

# 8. Running Your First Assessment

## 8.1 5-Minute Quick Start

1. Launch the application (1 min)  
2. Web browser opens to http://localhost:8502  
3. Select "Batch Scenario Processing" (30 seconds)  
4. Load Example Data (30 seconds)  
5. Review Data Preview (1 minute)  
6. Run Assessment (2-3 minutes)  
7. View Results (1 minute)  
8. Explore Visualizations (1 minute)  
9. Download Results (optional)

# 9. Interpreting Results

## 9.1 Key Output Columns

Annual\_Risk\_Median  
 Annual probability of infection per person  
 Example: 0.8660 = 86.6% chance of infection in one year  
 Decision criterion: < 0.0001 (WHO guideline) = COMPLIANT  
  
Annual\_Risk\_5th, Annual\_Risk\_95th  
 Uncertainty bounds (5th and 95th percentiles)  
  
Population\_Impact  
 Expected annual illnesses in exposed population  
 Formula: Annual\_Risk\_Median × Exposed\_Population  
  
Compliance\_Status  
 COMPLIANT: Annual\_Risk\_Median < 0.0001  
 NON-COMPLIANT: Annual\_Risk\_Median >= 0.0001

## 9.2 Risk Classification

|  |  |  |
| --- | --- | --- |
| Annual Risk | Classification | Action |
| < 0.00001 | Negligible | No action |
| 0.00001-0.0001 | Very Low | Monitor |
| 0.0001-0.001 | Low | Monitor closely |
| 0.001-0.01 | Medium | Implement controls |
| > 0.01 | High | Immediate action |

# 10. Advanced Features

## 10.1 Spatial Assessment

Compare risks across multiple sites. Select "Spatial Assessment" from sidebar to analyze the same pathogen scenario across different locations with varying dilution factors.

## 10.2 Temporal Assessment

Analyze seasonal and time-series trends. Select "Temporal Assessment" to determine which seasons pose highest risk.

## 10.3 Treatment Comparison

Evaluate different treatment technologies. Create scenarios with varying LRV values and compare outcomes to determine cost-effective treatment strategies.

## 10.4 PDF Report Generation

Generate professional PDF reports for stakeholder communication. Reports include executive summary, visualizations, data tables, and methodology.

# 11. Troubleshooting & Best Practices

## 11.1 Common Issues

### Application wont start

Verify Python installed. Check correct directory. Run: python -m streamlit run web\_app.py

### File upload fails

Ensure .csv format. Check column names match exactly. Verify no blank rows.

### Results look wrong

Verify pathogen concentrations realistic. Check dilution values in expected range (50-300).

### Application runs slowly

Reduce iterations (5,000 instead of 10,000). Process fewer scenarios.

## 11.2 Best Practices

Start with baseline scenarios before testing treatments  
Test one parameter at a time for understanding  
Use defensible parameter values (document sources)  
Report uncertainty bounds (5th-95th percentiles)  
Include seasonal and spatial variation  
Keep scenario matrices organized and documented

# 12. Example Scenarios

## 12.1 Multi-Site Summer Beach Assessment

OBJECTIVE: Assess bathing water safety across 3 beaches during summer  
  
CONTEXT:  
 Location: Coastal region, summer season  
 Pathogen: Norovirus  
 Exposure: Swimming (primary contact)  
 At-risk population: Beach visitors  
  
EXPECTED FINDINGS:  
 North Beach shows highest risk (lowest dilution)  
 South Beach shows lowest risk (highest dilution)  
 Treatment effectiveness reduces annual risk proportionally

## 12.2 Treatment Technology Evaluation

OBJECTIVE: Determine required treatment level for WHO compliance  
  
BASELINE SITUATION:  
 Current treatment: Primary only (LRV=1.5)  
 Annual risk: 0.015 (1.5%) - NON-COMPLIANT  
 Population at risk: 500 shellfish consumers  
  
ALTERNATIVES TESTED:  
 1. Current: Primary (LRV=1.5)  
 2. Secondary treatment (LRV=3.0)  
 3. Secondary + UV (LRV=5.0)  
 4. Tertiary MBR (LRV=7.0)  
  
RECOMMENDATION:  
 Select LRV=5.0 (Secondary + UV) as optimal  
 Cost-effective balance between safety and costs

# 13. Frequently Asked Questions

### Q1: How long does an assessment take?

Typical batch assessment (10-20 scenarios) with 10,000 MC iterations takes 2-4 minutes.

### Q2: Can I modify pathogen parameters?

Built-in parameters are fixed. Edit qmra\_core/pathogen\_database.py to modify.

### Q3: What if I do not have exact pathogen concentration data?

Use expert judgment based on literature. Hockey Stick requires only min/median/max.

### Q4: Difference between infection risk and annual risk?

Infection risk = per single exposure. Annual risk = across all yearly exposures.

### Q5: How do I know results are reasonable?

Compare to literature. Conduct sensitivity analysis. Verify higher LRV gives lower risk.

### Q6: Can I process hundreds of scenarios?

Yes, scales well. 100+ scenarios takes 5-10 minutes.

### Q7: What file formats are supported?

CSV only. Excel files must be exported as CSV first.

### Q8: How do I export results for other software?

Download full CSV results for use in R, MATLAB, Python, Excel.

### Q9: Is this tool validated?

Yes. Includes automated tests comparing to published QMRA methodologies.

### Q10: Can multiple users run assessments simultaneously?

Yes, each runs separate streamlit instance on different port.

# Conclusion

The QMRA Batch Processing Application provides a powerful, user-friendly platform for conducting quantitative microbial risk assessments. By combining advanced statistical methods with an intuitive interface, it enables water quality professionals to make science-based decisions about pathogenic contamination risks.  
  
Key strengths:  
• Accessibility: Three-file CSV format  
• Rigor: Empirical distributions and proper uncertainty quantification  
• Flexibility: Multiple assessment modes  
• Productivity: Batch process 100+ scenarios rapidly  
• Communication: Professional visualizations and reports  
  
This user guide provides the foundation for effective use. Integrate this tool into your organization's water quality management workflows.

QMRA Batch Processing Application - Technical User Guide  
NIWA Earth Sciences Division  
October 2025