

Bayesian estimation of seasonal and between year variability of norovirus infection risks for workers in agricultural water reuse using epidemiological data

Seis, Wolfgang; Rouault, Pascale; Miehe, Ulf; ten Veldhuis, Marie Claire; Medema, Gertjan

DOI

[10.1016/j.watres.2022.119079](https://doi.org/10.1016/j.watres.2022.119079)

Publication date

2022

Document Version

Final published version

Published in

Water Research

Citation (APA)

Seis, W., Rouault, P., Miehe, U., ten Veldhuis, M. C., & Medema, G. (2022). Bayesian estimation of seasonal and between year variability of norovirus infection risks for workers in agricultural water reuse using epidemiological data. *Water Research*, 224, Article 119079. <https://doi.org/10.1016/j.watres.2022.119079>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Bayesian estimation of seasonal and between year variability of norovirus infection risks for workers in agricultural water reuse using epidemiological data

Wolfgang Seis^{a,b,*}, Pascale Rouault^a, Ulf Miehe^a, Marie-Claire ten Veldhuis^b, Gertjan Medema^b

^a Kompetenzzentrum Wasser Berlin gGmbH, Cicerostaße 24, Berlin 10709, Germany

^b Department of Water Management, Faculty of Civil Engineering & Geosciences, Delft University of Technology, Delft, GA 2600, the Netherlands

ARTICLE INFO

Keywords:

QMRA
Water reuse
Bayesian modeling
Risk assessment
Norovirus

ABSTRACT

Norovirus infections are among the major causes of acute gastroenteritis worldwide. In Germany, norovirus infections are the most frequently reported cause of gastroenteritis, although only laboratory confirmed cases are officially counted. The high infectivity and environmental persistence of norovirus, makes the virus a relevant pathogen for water related infections. In the 2017 guidelines for potable water reuse, the World Health Organization proposes Norovirus as a reference pathogen for viral pathogens for quantitative microbial risk assessment (QMRA). A challenge for QMRA is, that norovirus data are rarely available over long monitoring periods to assess inter-annual variability of the associated health risk, raising the question about the relevance of this source of variability regarding potential risk management alternatives. Moreover, norovirus infections show high prevalence during winter and early spring and lower incidence during summer. Therefore, our objective is to derive risk scenarios for assessing the potential relevance of the within and between year variability of norovirus concentrations in municipal wastewater for the assessment of health risks of fieldworkers, if treated wastewater is used for irrigation in agriculture. To this end, we use the correlation between norovirus influent concentration and reported epidemiological incidence ($R^2=0.93$), found at a large city in Germany. Risk scenarios are subsequently derived from long-term reported epidemiological data, by applying a Bayesian regression approach. For assessing the practical relevance for wastewater reuse we apply the risk scenarios to different irrigation patterns under various treatment options, namely “status-quo” and “irrigation on demand”. While status-quo refers to an almost all-year irrigation, the latter assumes that irrigation only takes place during the vegetation period from May - September. Our results indicate that the log-difference of infection risks between scenarios may vary between 0.8 and 1.7 log given the same level of pre-treatment. They also indicate that under the same exposure scenario the between-year variability of norovirus infection risk may be > 1 log, which makes it a relevant factor to consider in future QMRA studies and studies which aim at evaluating safe water reuse applications. The predictive power and wider use of epidemiological data as a suitable predictor variable should be further validated with paired multi-year data.

1. Introduction

Quantitative microbial risk assessment (QMRA) has been developed to support risk-based decision making by quantifying health risks resulting from the exposure to pathogenic microorganisms (WHO, 2016). QMRA supports the assessment of the performance of a system to achieve pre-defined health targets, and can be used to explore various

risk management alternatives. In QMRA, the source water concentration of pathogens is an essential model input, because source water concentrations and exposure scenarios determine operational health targets in terms of required log-removal values (LRV). Therefore, the better the understanding of source water pathogen concentrations, including its variability (e.g. seasonal) and uncertainty (e.g. parameter uncertainty), the better risk reduction measures can be tailored to the local

Abbreviations: WHO, World Health Organization; QMRA, quantitative microbial risk assessment; RKI, Robert-Koch Institute; PPD, posterior predictive distribution; AGI, acute gastro-enteric illness; NoV GII, norovirus genogroup II; DALY, disability adjusted life years; (R)MSE, (root) mean squared error.

* Corresponding author at: Kompetenzzentrum Wasser Berlin gGmbH, Cicerostaße 24, Berlin 10709, Germany.

E-mail address: wolfgang.seis@kompetenz-wasser.de (W. Seis).

<https://doi.org/10.1016/j.watres.2022.119079>

Received 22 April 2022; Received in revised form 15 August 2022; Accepted 6 September 2022

Available online 8 September 2022

0043-1354/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

circumstances.

In Germany, norovirus (NV) infections are the leading cause of reported cases of viral gastroenteritis among all age groups. Pathogenic noroviruses are subdivided into three groups (GI, GII, and GIV) with NV GII.4 being currently responsible for 70–80% of all reported gastroenteritis cases (Bernard et al., 2014; de Graaf et al., 2015). NV GII are highly infective, and resistant in the environment, are considered relevant pathogens for water related infections, and are included in multiple studies related to water reuse applications (Sales-Ortells et al., 2015; Soller et al., 2018, 2017). Despite criticism towards using NV as a reference pathogen in QMRA, which originates from the fact that Norovirus is not readily culturable from environmental samples (Nappier et al., 2018), Norovirus was included as a reference pathogen in the 2017 WHO guidelines for potable water reuse, which underlines its relevance for water reuse applications (WHO, 2017).

In probabilistic QMRA studies, sources water pathogen concentration are described as random variables. Distributions fitted to pathogen data in the source water, aim at representing the existing variability of these pathogens over a specific period, usually one year. In this context, a potential source of variability, whose relevance has rarely been addressed in water related QMRA studies, is the within and between-year-variability of NV concentrations in wastewater and how these variations might affect risk management alternatives. In most studies, NV concentration are measured over periods of ≤ 1 year (Huang et al., 2022). Therefore, the question arises about how inter-annual NV variations might affect risk assessment results and thus risk management alternatives for longer periods. To address this question approaches for deriving plausible risk scenarios are needed.

Additionally to inter-annual variations, in moderate climates, like e.g. in Germany, NV outbreaks and water demands in agriculture, show an inverse seasonality, meaning that NV incidences are lowest in summer (Bernard et al., 2014) where water demands in agriculture is at its peak and vice versa. Seasonal differences, i.e. intra-annual variations, of wastewater concentrations have been confirmed in a review of NV in wastewater by Eftim et al. (2017). Therefore, for the special case of agricultural water reuse, adapting irrigation periods to low incidence periods may be a feasible, low-tech and cheap risk-reduction measure complementing other, more obvious risk reduction measures like wastewater disinfection, protective farming practices or the use of exposure reduced irrigation technologies (e.g. drip irrigation).

For NV, there is a direct causal relationship between infection, and virus shedding (Atmar et al., 2008; Teunis et al., 2015), and therefore the presence of these viruses in municipal wastewater. While virus concentrations in wastewater have been used to estimate the variability of incidence in the population (Hellmer et al., 2014; Medema et al., 2020), long-term reported incidence has not been used to derive plausible risk scenarios for addressing the relevance of intra- and inter annual variations, for irrigation related exposure of fieldworker in agricultural water reuse.

Therefore, the objectives of the present study are to:

- (a) apply this technique to an existing reuse site to assess the potential relevance of inter annual variations of NV concentrations for the assessment of health risks for fieldworkers exposed via sprinkler irrigation,
- (b) derive practical implications of the estimated inter-annual variations of NV GII concentrations for the derivation of local LRV values, and
- (c) assess the risk reduction potential of adapting irrigation periods to low incidence based on the expected intra-annual/within year variability.

2. Methods

The city of Braunschweig is considered a suitable location to investigate the formulated questions, as

- (a.) it is a large water reuse location in Germany (nominal WWTP size: 275,000 p. e., applied load: 350,000 p. e.), so that population incidences of virus infections are potentially represented in the wastewater composition
- (b.) it uses treated water to irrigate agricultural areas almost during all the year, and
- (c.) long-term epidemiological data on reported norovirus cases are available on a weekly basis for the district of Braunschweig (cf. Section 2.1.2) by public epidemiological databases.

Wastewater is treated by standard primary treatment with subsequent activated sludge treatment. To this day, no disinfection unit is implemented. Therefore, irrigation is restricted to energy crops or products which are consumed only after further industrial processing. However, current research activities explore the potential benefits of upgrading water quality to “Class A” according to European standards for water reuse, which would allow to diversify the products grown on the area. In Braunschweig, irrigation takes place all year round except for a short period between December and January. Irrigation is conducted via sprinkler irrigation and is organized by the *Wastewater Association Braunschweig (AVB)*, which matches the water demands of farmers with the quantities provided by the wastewater treatment plant. Irrigation itself is conducted by irrigation managers and workers associated with the AVB. They are responsible for placing and operating irrigation machines on the specific sectors of the agriculture areas. These workers represent the group of people, which is at risk of being regularly and directly exposed to pathogens via spray of treated wastewater (cf. Section 2.3.2).

2.1. Data collection

2.1.1. Influent data of Norovirus GII

Norovirus samples were collected from October 15th to December 3rd, 2014. Manual grab samples were taken from the influent of the wastewater treatment plant once per week on Wednesdays during morning peak flow. Samples were collected into sterile plastic bottles of 1L. After sample collection, samples were frozen and sent to the University of Barcelona for subsequent NV RNA-analysis. NV have been shown to keep intact for several freezing and thawing cycles (Richards et al., 2012), which justifies freezing as a sample conservation strategy. NV in water samples were concentrated with the skimmed milk flocculation protocol (Calgua et al., 2013a, 2013b). The viral RNA concentrations were quantified by specific qPCR assays for NoV GII (Kageyama et al., 2003; Loisy et al., 2005).

2.1.2. Collecting epidemiological data

Epidemiological data were collected from the publicly available epidemiological database *SurvStat*, which summarizes illness caused by any NV genogroup and –type as *norovirus gastroenteritis*. Only cases which fulfill the reference definition were included in the query, which includes cases confirmed by either PCR test, antigen tests or electron microscope detection. This definition and the comparatively mild illness of NV infections are known to cause high numbers of underreporting caused by asymptomatic infections of NV, differences in tolerance levels before consulting a physician, and difference in testing and reporting practices between physicians when gastroenteritis is diagnosed (Bernard et al., 2014).

From *SurvStat* we queried weekly data on reported NV cases for the district of Braunschweig on a NUTS 2 level (Nomenclature des unités territoriales statistiques). The NUTS-2 level includes the city of Braunschweig and surrounding districts. It is larger than the catchment of the WWTP, whose geographic boundaries would better match the boundaries at a NUTS-3 level. We prefer the NUTS-2 level of geographic boundary over the more detailed NUTS-3 level, because of the high expected underreporting for NV. Increasing the population size makes the observed incidence more robust against the detection of single

outbreaks. Moreover, as Braunschweig is a regional economic center commuting back and forth the city from outer districts is common, so that inhabitants of these districts are likely to contribute to the city's wastewater composition and incidence dynamic. In their annual seasonal reports about NV infections, health authorities of Lower Saxony underline that annual spatial differences of reported incidence between neighboring districts are more likely to be caused by differences in specific diagnose regime and reporting behavior rather than other causes like e.g. age structure. Therefore, staying at the NUTS-2 level is considered the best choice for balancing the robustness of the incidence estimate while accounting for regional behavior of norovirus infections. The specific query is attached to the supporting information (SI). For detailed evaluation (cf. Section 2.2) we only consider the years between 2001 (first year of reporting) and 2019, as 2020 and 2021 are influenced by the corona pandemic, and thus a potentially overburdened health reporting system.

2.2. Data analysis

2.2.1. Data preparation

For an initial quality check, the collected epidemiological data were plotted for detecting any obvious, visual irregularities. Moreover, the dataset was checked for completeness by counting the number of observations in each calendar year. We applied an inclusion criterion of maximum 5% (2 data points) of missing data for each year. If missing values existed, values were imputed by linear interpolation, which we prefer over alternative imputation methods like mean or median imputation due to the expected auto-correlation in the data set. The years 2004, 2009, and 2015 have 53 calendar weeks. To align the periodicity between years, and ease computations, incidence values were added to the values of week 52, and week 53 was removed from the dataset. This adjustment has no effect on the risk calculation as December and January are not included in any risk modelling scenario.

2.2.2. Descriptive analysis

The reported norovirus incidence is grouped by year and calendar week, which, after the application of the inclusion criterion (cf. Section 5) leads to 14 observations. For each week the summary mean, median, minimum, and maximum are calculated. Additionally, negative binomial distributions are fitted to each dataset, and 95% credible and prediction intervals are simulated from the posterior distribution. For all statistical analyses, we apply a Bayesian approach to statistical inference using the programming languages *R* (R Development Core Team, 2008), *Stan* (StanDevelopmentTeam, 2017), and the interface between *R* and *Stan*, *brms* (Bürkner, 2017). For the negative binomial regression, *brms* implements the following parameterization, which allows the negative binomial to be interpreted as a Poisson distribution whose variance parameter μ is complemented by an additional factor for overdispersion ϕ scaled by μ^2 .

$$\text{NegBinomial}(y|\mu, \phi) = \binom{y + \phi - 1}{y} \left(\frac{\mu}{\mu + \phi}\right)^y \left(\frac{\phi}{\mu + \phi}\right)^\phi \quad (1)$$

$$E[Y] = \mu \text{ and } \text{Var}[Y] = \mu + \frac{\mu^2}{\phi}. \quad (2)$$

2.2.3. Bayesian regression modelling for developing risk scenarios

In the present study, we use reported incidence as a relative index to create *plausible risk scenarios* for the variability of NoV GII concentrations in municipal wastewater using statistical regression models. Due to the continuous evolvments of NV genotypes and existing under-reporting, it is not likely that the correlation between wastewater concentrations and reported incidence observed in one year remains exactly constant over time. Thus, using the reported incidence to predict NV wastewater concentration for a specific point in time may not always be accurate. However, we argue that the longer-term variability in reported

incidence reflects the longer-term variability of norovirus infections and illness within the population. For example, Stegmaier et al. (2020) compared differences in spread dynamics of influenza and NV in the population based on data provided by the SurvStat database. Therefore, we further argue that generating *scenario predictions* under the assumption of a constant correlation, leads to estimates which might not necessarily be correct for each specific year, but which can be considered *realistic scenarios* of unobserved situations. The latter is a key element of risk assessment studies. To emphasize this distinction, we will write e.g. 2008 if we refer to the year and its related data, and use the term *incidence scenario (IS08)* if we refer to the risk scenario derived from reported incidence.

In the literature, different information can be found regarding whether wastewater concentrations correlate best with the incidence reported in the same week (Kazama et al. (2017) or whether there is a delay in reporting rates. Hellmer et al. (2014) reported peak concentrations of NV, 2 weeks before an increase in incidence could be observed. Therefore, we checked the correlation and goodness-of-fit (cf. Section 3.1) for both the incidence of the current reporting week as well as to a 2-week rolling average, which aggregates of the incidence of the present and the 2 subsequent weeks.

For the paired data, first, the Pearson's correlation coefficient is calculated. Second, three regression models, which make different distributional assumptions, are fitted and compared. Since both incidence and NoV GII concentrations are all positive quantities, we fit the relation on a logarithmic scale. As candidate models a lognormal, negative binomial and gamma regression are fitted to the paired data. The parameterization of the negative binomial distribution follows the example above Eqs. (1) and (2) so that the location and scale parameters can be addressed separately. The one for the gamma regression is shown in Eqs. (10) and (11). For the scale parameters ϕ , σ , and φ default prior distribution provided by the *brms* package were chosen, which uses heavily tailed distributions, to only introduce minimal information (see Eqs. (5), (8), (13)).

Lognormal regression

$$c_{NoV} \sim \text{Lognormal}(\mu, \sigma) \quad (3)$$

$$\log(\mu) = \beta_0 + \beta_1 \log(\text{incidence}) \quad (4)$$

$$\begin{aligned} \sigma &\sim \text{half-t}(\nu = 3, \lambda = 0, \zeta = 2.6) \\ \nu &= \text{degrees of freedom} \\ \lambda &= \text{location} \\ \zeta &= \text{scale} \end{aligned} \quad (5)$$

Negative binomial regression

$$c_{NoV} \sim \text{negBinomial}(\mu, \phi) \quad (6)$$

$$\log(\mu) = \beta_0 + \beta_1 \log(\text{incidence}) \quad (7)$$

$$\phi \sim \text{Gamma}(0.1, 0.1) \quad (8)$$

Gamma regression

$$c_{NoV} \sim \text{Gamma}(\alpha, \beta) \quad (9)$$

$$\alpha = \frac{\mu^2}{\phi} \quad (10)$$

$$\beta = \frac{\mu}{\phi} \quad (11)$$

$$\log(\mu) = \beta_0 + \beta_1 \log(\text{incidence}) \quad (12)$$

$$\phi \sim \text{Gamma}(0.1, 0.1) \quad (13)$$

For fitting the models we use Bayesian estimation methods based on Markov Chain Monte Carlo (MCMC). For every model we ran four independent Markov Chains, with 10,000 iterations and a warm-up phase

of 5000. Thereby, 20,000 independent posterior samples were created for further evaluation. We chose this high number of samples as the posterior samples are not only used for estimation but also for subsequent risk simulation (cf. Section 2.3.2). Convergence of the Markov Chains were checked by inspecting whether the traceplots of the four chains were well-mixed and by checking whether the Gelman-Rubin diagnostic statistic (\hat{R}) has converged to 1. The traceplot and convergence diagnostics of the selected model is shown in the SI.

2.2.4. Model selection

For selecting between models the expected log-predictive density (elpd) is used. The out-of-sample elpd is calculated using leave-one-out cross-validation, following the approach published by Vehtari et al. (2017). To this end, the implemented function “loo” of the *brms* package is used. The elpd is preferred over other methods for model comparison like e.g. MSE or RMSE as it explicitly accounts for different distributional assumptions, by calculating the probability density of the held out data point and not only assesses the distance to the estimated mean. The higher a model’s elpd the better it is expected to predict new data. Additionally, the Bayesian approximation of R^2 is simulated according to Gelman et al. (2019). The model with the highest elpd is selected.

2.3. Risk modeling

2.3.1. Problem formulation

Risk modeling focuses on quantifying the risk reduction potential for fieldworkers of an “irrigation on demand” scenario in comparison to a “status quo” scenario. Under the scenario status quo irrigation takes place from February to the end of November, thus from calendar week 5 to calendar week 47. Under the scenario irrigation on demand, we assume that irrigation to take place from May to September, thus from week 16 to 36. This adaption is expected to reduce the risk in two ways. First, a reduction of irrigation period reduces the risk by a reduction of the number of exposure events per year (cf. Eq. (13)), given an average weekly exposure rate. Secondly, the risk of infection is reduced by avoiding irrigation during the peak season of NV infections. To compare the practical relevance of the difference between scenarios we assess both the relative risk between scenarios given the same wastewater treatment performance, and the required log-reduction to achieve the same level of health protection (cf. Section 2.3.4). For the latter, a health benchmark is needed, for which we apply a value of 10^{-2} per person per year (pppy). This value is arbitrary, as no level of acceptable risk exists for fieldworkers in Germany. However, it follows the rational that water related exposures should be considered safe and should not contribute to the overall disease burden. The latter was quantified for acute gastrointestinal illness (AGI) in Germany among adults in a representative survey by Wilking et al. (2013). They found that that 9 out of 10 adults in Germany experience a case of AGI per year. Therefore, the applied threshold of 10^{-2} is approximately 2 orders of magnitude below the existing disease burden and thus, considered to be sufficiently safe, or at least hardly detectable. Applying a disease per infection ratio of 0.7 and a disease burden of 5×10^{-4} disability adjusted life years (DALYs) per case (WHO, 2017), results in a tolerable disease burden of 3.5×10^{-6} DALYs pppy, which is close to the threshold of 10^{-6} DALYs pppy applied for drinking water.

2.3.2. Exposure assessment

For modelling infection risks for the two irrigation and 14 incidence scenarios we use the weekly incidence and the model selected in Section 3.2 to generate scenario predictions for NV concentrations for every week of every incidence scenario. To this end, we sample from the posterior predictive distribution (PPD) for every week during this period using the obtained parameter distributions. From the simulated PPD, the population distribution of the expected dose per exposure event is calculated by:

$$d = 10^{c_{\text{influent}} - \sum_{T=1}^n LRV_T} * V_{\text{ingested}} \quad (14)$$

Where d is the dose per exposure event, c_{influent} is the PPD of NoV GII in lg GC/L, LRV is the log-removal value for treatment step T , n is the number of treatment steps, and V_{ingested} is the ingested volume in L.

To isolate the effect of different annual irrigation periods we assume the remaining model inputs to be known without variation. For the log-reduction of the wastewater treatment plant, before irrigation, we run simulations for the range between 1 and 8. This allows to determine the sensitivity of derived risk indicators (cf. Section 4.3.4) between scenarios as a function of the installed level of wastewater treatment. For the volume ingested by fieldworkers (V_{ingested}) per exposure event via sprinkler irrigation and the frequency of exposure events, we assume an ingestion of 0.1mL of treated wastewater to happen once a week. This leads to 42 exposure events per year for the “status quo” scenario and 20 for the “irrigation on demand” scenario. To assess the sensitivity of our result to these assumptions we additionally ran simulations for:

- (a.) A volume 1mL, once per week
- (b.) A volume of 0.1, twice per week

In brief, for risk simulation we follow the following algorithm:

- 1 For each incidence scenario, filter the scenario-specific irrigation periods (status-quo, irrigation on demand) from the generated array of PPDs.
- 2 sample a random sample from each distribution of weekly NV GII simulation (42 for "status-quo", 20 for "irrigation on-demand")
- 3 for each sample apply the assumed log-reduction and assumed volume per exposure event to estimate the dose per exposure event (Eq. (14))
- 4 calculate the infection probability per exposure event by applying the dose-response of NoV GII for disaggregated viruses (cf. section 0) (Eq. (15))
- 5 aggregate the calculated event probabilities into an estimate for the annual risk (Eq. (16)), and
- 6 repeat steps 2-5 for 1000 times to generate an uncertainty estimate for the average annual risk.

To further estimate the generated Monte-Carlo error we repeated this simulation under the initial assumptions (0.1mL, 1/week) for 25 times.

2.3.3. Health effect assessment

The dose-response model for norovirus is subject of ongoing discussion. Until recently, QMRA studies which assessed the risk of norovirus infection relied on the dose-response relation published by Teunis et al. (2008), which was simplified by Messner et al. (2014) and criticized to be highly uncertain and overly conservative for low concentrations by Schmidt (2015). The dose-response model was derived only for norovirus of genogroup I, so that subsequent QMRA studies had to make the inherent assumption that this dose-response relationship would apply for all norovirus genogroups and genotypes. A summary of the discussion and practical recommendations was provided by Van Abel et al. (2017). Recently, Teunis et al. (2020), summarized the information from challenge studies and outbreaks, using a Bayesian hierarchical modelling approach to derive an updated generalized dose-response relationship for norovirus of genogroups I and II, using the secretor status of the exposed person as an additional predictor variable. We consider the latter study as the most comprehensive study on this topic so far, so that we base our risk estimate on the parameters reported there. We also chose this study as it provides specific estimates for the dose-response parameters for NoV GII, which our study focuses on. The dose-repose model is based on gene copies which is in line with the data measured at the WWTP. As secretor negative (SE-) persons are

widely protected against norovirus infections, we focus our analysis on secretor positive people and assume all fieldworkers to be SE+. Thus, we calculate the population distribution of risk outcomes per exposure event as the probability of infection P_{inf} from the population distribution of simulated doses d by:

$$P_{inf|exposure} = 1 - {}_1F_1(a, a + b, -d) \tag{15}$$

As parameter values we use the median of their posterior distributions as reported in Table A13 of Teunis et al. (2020), namely $a = 0.23$, and $b = 5.04$.

For solving the confluent hypergeometric function we use the function *hyperg.1F* of the *gsl*-package of the R-programming language. We aggregate the individual event probabilities ($P_1 \dots P_N$) to an estimate of the annual infection risk (P_{annual}) by:

$$P_{annual} = 1 - \prod_{i=1}^N (1 - P_{inf|exposure[i]}) \tag{16}$$

2.3.4. Risk characterization

To characterize the risk we assess both the statistical and practical significance of the obtained results. To assess, whether the difference between the simulated risk distributions are statistically significant, we derive a Bayesian p-value from the simulated average annual risk. To this end, the average risk distributions for the status-quo scenario are subtracted from the distribution of the irrigation of demand scenario for each incidence scenario. The proportion of samples of the resulting difference distribution > 0 , is used as an approximation for the probability of the two distributions coming from the same underlying population. If the calculated probability is < 0.05 , the two distributions are considered statistically different. Since we simulated 1000 annual risk samples of for each incidence scenario, the lower limit for this approach is 0.001^{**} , which is sufficiently low.

For an assessment of the practical significance we calculate the risk ratio and log difference in risk between scenarios for each incidence scenario and each applied log-removal.

$$Risk\ ratio = \frac{Risk_{annual, status\ quo}}{Risk_{annual, irrigation\ on\ demand}} \tag{14}$$

This corresponds to the differences in risk, assuming the same level of wastewater treatment for both irrigation scenarios.

Additionally, we deduce the log-removal needed to obtain the same level of health protection given the applied exposure scenario. For each

scenario the required log-removal for achieving a value $P_{inf} \leq 1\%$ is deduced by linear interpolation between of the closest points of the discrete evaluations of the resulting risk function in the range 1–8 LRV. Subsequently, the difference between scenarios is calculated as measure of additional health benefit due to changes in irrigation patterns.

3. Results

The results of the initial quality control are shown in Fig. 1-A in the SI. From the data no obvious trends in the development of reported norovirus incidence can be determined, which would indicate an underlying systematic effect. Instead, peaks of norovirus outbreaks during the winter months reoccur annually with differences in amplitude. The quality check for data completeness revealed complete datasets from 2009-2019. 2001-2003 and to a lesser extent 2004 and 2005 showed relevant data gaps, and were removed. The years 2006 (2 missing), 2007, 2008 (1 missing each) stayed within acceptable limits.

3.1. Descriptive analysis

Fig. 1 shows the distribution of weekly reported norovirus of incidence in the district of Braunschweig from 2006-2019. Incidences show a pronounced seasonality. They start rising in early winter around weeks 47-48 (December) and reach maxima during weeks 1-8 (January-February). Lower numbers of reported cases during week 52 are considered artifacts caused by Christmas and New Year's vacations. Results show that during winter the median incidence varies between 4 and 7 per week. Single peak observations reach values of up to 37. In summer, the reported incidence stays at lower numbers between 0.4 and 2. The peak incidences of 37 were reported outside both irrigation scenarios and thus do not affect risk calculations.

Table 1 shows the raw NV measurements, paired number of reported cases, the incidence, the rolling two week incidence, the calculated Pearson correlation factors, the posterior mode of the distribution of the Bayesian R^2 , and the values for the calculated log-predictive densities for the lognormal, gamma, and negative binomial regression, respectively. Results show that during the sampling period between week 42 and 49 2014, an increase in both the incidence, ranging from 0.37 to 6.7 and norovirus concentrations ranging from $10^{4.3}$ - 10^7 GC/L can be observed. Results show a strong positive correlation with a Pearson correlation factor of 0.96. Combined with the known causality between the number of infected people and norovirus wastewater concentration,

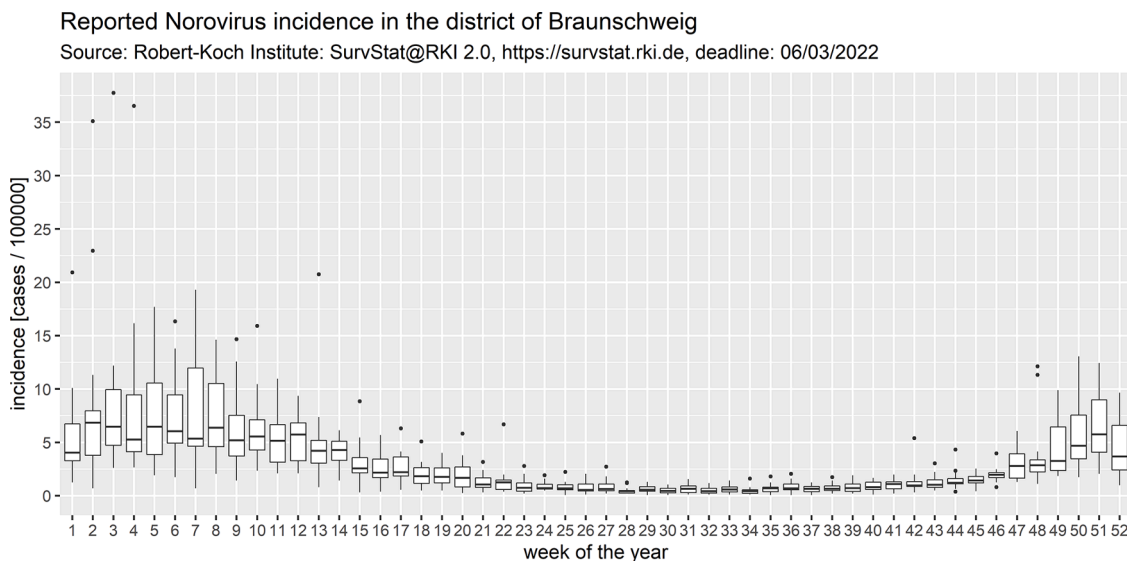


Fig. 1. Overview over reported norovirus cases in the district of Braunschweig. Individual boxplots include the weekly data from the selected 14 years between 2006-2019.

Table 1

Overview over measurements, reported cases, incidence, correlation factor and model performance indicators.

Reporting week	# cases / incidence / rolling mean incidence	Sampling Date	NoVGII [GC/L]	Pearson correlation	R ² / elpd log-normal	R ² / elpd gamma	R ² / elpd negative binomial
42	15 / 0.93 / 0.9	2014-10-15	1.8×10^4	0.96	0.5 / 0.5 -109.1/-107.6	0.9/0.93 /-108.4/-106.2	0.9/0.93 / -108.5/-106.1
43	14 / 0.87 / 0.62	2014-10-22	1.19×10^5				
44	6 / 0.37 / 0.71	2014-10-29	3.45×10^5				
45	17 / 1.05 / 1.43	2014-11-05	7.82×10^5				
46	29 / 1.80 / 2.45	2014-11-12	6.93×10^5				
47	50 / 3.09 / 3.62	2014-11-19	6.66×10^6				
49	108 / 6.69 / 5.89	2014-12-03	1.09×10^7				

this observation provides support for the hypothesis that the variability of reported epidemiological provides information for generating plausible risk scenarios for the potential variability of wastewater concentrations.

3.2. Model selection

Regarding the candidate models, the negative binomial approach indicates a slightly better (higher elpd) predictive performance than the lognormal, and gamma approach. Since the negative binomial distribution is also the most logical choice for modeling over-dispersed discrete data, the negative binomial model is selected for simulating Norovirus influent concentrations. The models developed with the rolling 2-day mean performed slightly better than the models developed with the incidence reported in the same week. Although the difference is very small the negative binomial model based on the 2-day rolling mean incidence is used for subsequent estimations. Fig. 1-D in the SI shows model predictions against NV data.

3.3. Simulated influent concentrations

Results of the predicted weekly norovirus concentrations based on the mean 2-day rolling average incidence are shown in Fig. 2. Results for the individual incidence scenarios are shown in the SI. The results show that following the dynamic of norovirus incidence, the highest concentrations are expected to occur during winter (week 48-12). Prediction intervals cover peak concentrations of up to 10^9 genome copies per liter. The highest measured concentration during the sampling period was 1.1×10^7 genome copies / Liter. Norovirus concentrations of up to 10^9 GC/L have been reported for noroviruses in European wastewater treatment plants before (Eftim et al., 2017; Seis et al., 2020). Thus, the upper range of expected unobserved observations is considered plausible and in line with existing knowledge about norovirus concentrations in wastewater.

Average concentrations during the peak season are expected to lie at 10^7 GC/L, while during summer average concentrations are expected to lie 1.5-2 order of magnitude below, with minima at 10^3 GC/L in for scenario IS06, and maxima at 10^6 e.g. in summer of scenario IS14.

3.4. Annual risk for different irrigation scenarios

Fig. 3 (upper graph) shows exemplarily the distribution of the average annual risk simulated for the two irrigation scenarios and 14 incidence scenarios assuming an installed pretreatment of 6 LRV. The results indicate that within the same irrigation scenario result may vary by more than 1 order of magnitude. For example, with a LRV of 6 the mean annual risk of the “status quo” scenario ranges between 2.1×10^{-4} for IS06 and 2.8×10^{-3} for IS08, for the irrigation on demand scenario between 1.7×10^{-5} for IS17 and 1.9×10^{-4} for IS10. The minimum and maximum incidence scenarios do not correspond between irrigation scenarios because winter peaks, dominating risk simulations under status quo conditions are cut off from the risk simulation in the irrigation on demand scenario. This leads to different risk patterns. Fig. 3 (lower graph) shows exemplarily the simulated annual risk as a function of the installed log-removal for the incidence scenario IS08. Since the graphs from the individual incidence scenarios look very similar, the remaining ones are provided in the SI. The calculated Bayesian p-values are statistically significant (Bayesian p-value < 0.001**) for every single incidence scenario and installed pretreatment.

Regarding the practical significance, Fig. 5 shows that the effectiveness of a change in irrigation patterns depends on the installed pretreatment. For the assumed exposure volume and frequency, the calculated risk indicators (risk ratio and log-difference) increase until the pre-treatment reaches a value between 3 and 4 LRV and stay at an approximately constant level afterwards. The sensitivity assay (cf. SI Fig. 1-G, 1-H) reveals that the level of required pre-treatment to reach stable conditions increases with the assumed exposure. Note, however,

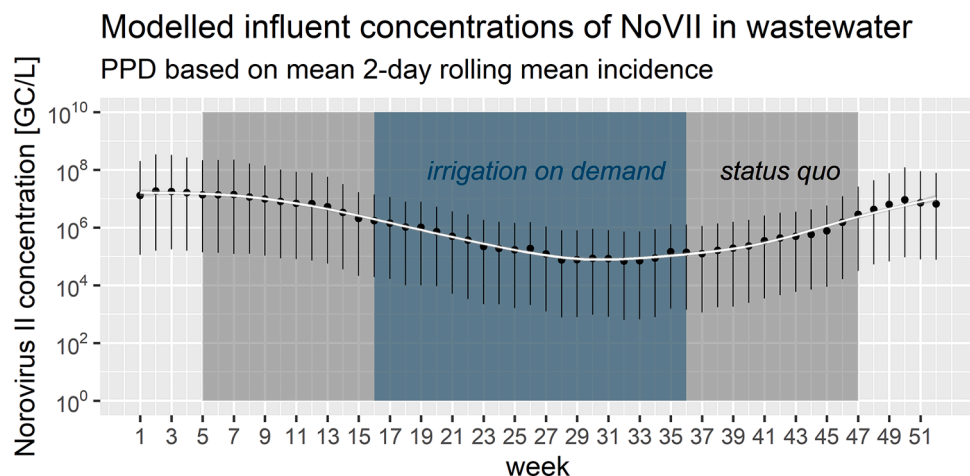


Fig. 2. Predicted norovirus concentration and uncertainty intervals for calendar week 1-52 for based on median weekly reported incidence over all incidence scenarios. Shaded areas indicate irrigation scenarios.

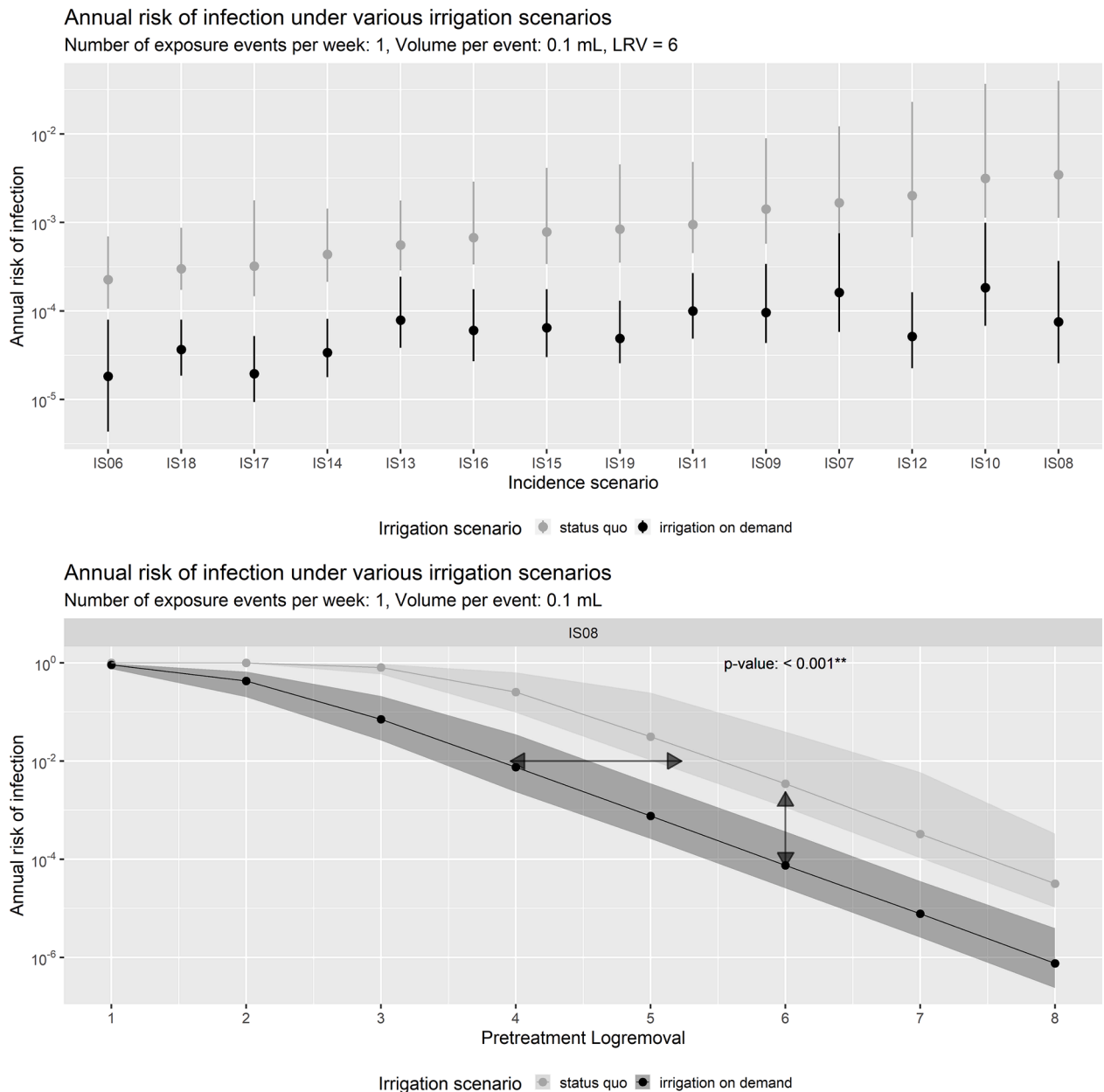


Fig. 3. Risk of infection. The upper graph shows the between incidence scenario comparison at a LRV value of 6. The lower graph shows the differences between scenarios as a function of installed pre-treatment. Vertical arrow: values used for calculating the risk ratio at a LRV = 6. Horizontal arrow: difference of necessary pre-treatment for achieving the same level of health protection.

that once stable conditions are reached, risk indicators between sensitivity scenarios stay within the margin of Monte Carlo error of the original exposure scenario. This indicates that overall conclusions about the magnitude of expected differences/ratios and thus the relevance of the inter- and intra-annual variability of NV remain valid against varying exposure assumptions.

Lower risk indicators at lower pre-treatment are caused by the fact that if the pretreatment is too low, the simulated annual risk is very close to 1 for both scenarios. If the constant level is reached the highest ratio is calculated for the IS08 and IS12 with risk ratios between 35 and 50, which corresponds to a log-difference of 1.5-1.7. The lowest risk ratio and log-difference is observed for IS13 with values of 6-7 and 0.8, respectively. The mean and median of the median log-difference are 1.15 and 1.1, respectively, making in it on average 10-15 times more likely to become infected under the status quo scenario in comparison to the irrigation on demand scenario once stable condition are reached.

Regarding the required log-removal to achieve the same level of health protection of 1% we calculate median LRV between 0.9 and 1.9 between incidence scenarios with a median of 1.2 LRV (Fig. 4). Note, that while the absolute required LRV values increase with increasing exposure in the sensitivity assay, the derived *difference* between irrigation scenarios stay within the Monte Carlo error and thus do not affect the overall conclusion of the analysis. That means that the between year variability of the required treatment performance may be up to 1 order of magnitude and that an on average an additional order of magnitude can be achieved by adapting irrigation patterns to an irrigation in the demand scenario. As a side note, our results also illustrate the difference between statistical and practical significance, as even the difference between scenarios at a pre-treatment level of LRV equal to 1 is statistically significant, while from a practical perspective this difference is irrelevant.

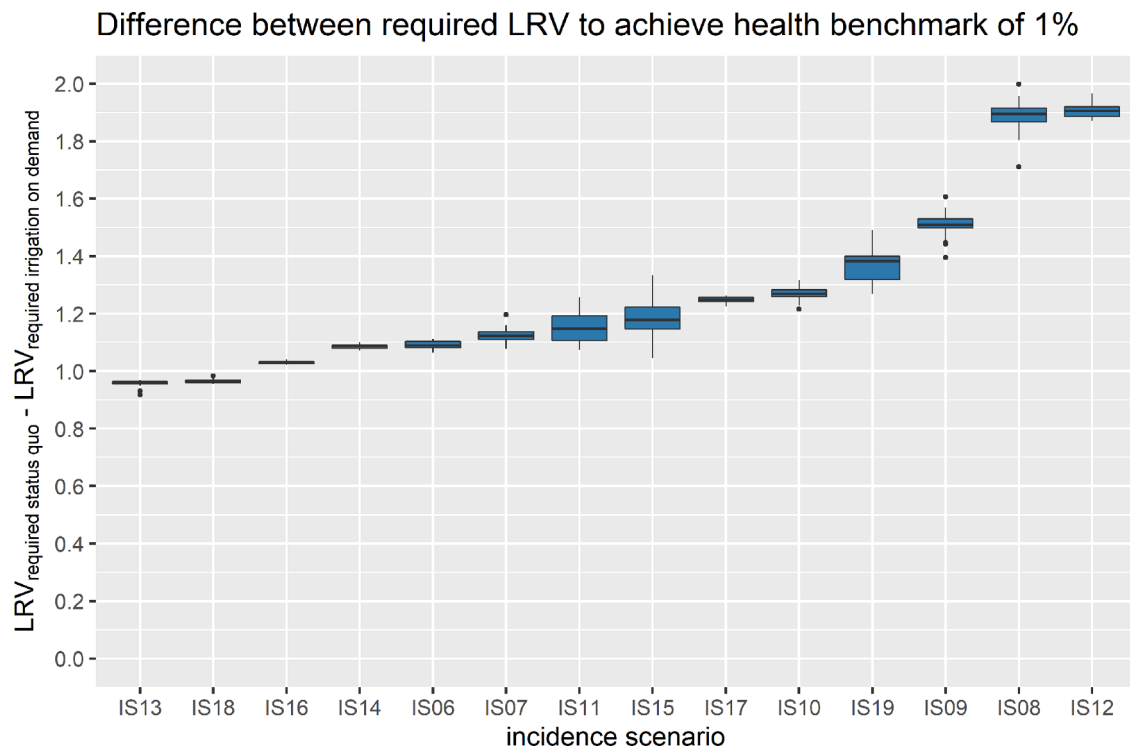


Fig. 4. Difference in required LRV for achieving a risk level of 1% per person per year. Boxplots indicate the Monte Carlo uncertainty derived from 25 independent risk scenario evaluations.

4. Discussion

In the present study, we used epidemiological data to assess the potential relevance of seasonal and between-year variability of NV concentration in WW for risk assessment and management of field-workers in agriculture. To this end, we derived plausible risk scenarios from weekly incidence data on reported NV cases in the district of Braunschweig. Several studies before confirmed both seasonal variations of NoV GII concentrations as well as the positive correlations between wastewater concentrations and the number of reported NoV GII cases / incidence (Eftim et al., 2017; Hellmer et al., 2014; Huang et al., 2022; Kazama et al., 2017; Miura et al., 2016). In previous studies, the relation between pathogen concentrations in wastewater and incidence in the population was used to estimate the incidence in the population, proposing that regular pathogen monitoring might function as an early warning system for diseases spreading in the population (Hellmer et al., 2014; Medema et al., 2020). To the best of our knowledge, our study is the first attempt to use this relationship to estimate the relevance of the between-year variability of NV risk in water-related QMRA studies, and eventually its potential practical implications regarding risk reduction potentials and performance requirements.

Our results suggest that within and between year-variability is likely to have a relevant impact on risk assessment results and thus should be considered a relevant factor to consider when evaluating local data for local QMRA studies. Thereby, our approach addresses a very common problem in QMRA studies, namely an assessment of the potential impact and relevance of the lack of multi-year data for local risk management.

Due to the low sample size ($n=7$) of norovirus our own study has to be seen as a first attempt to use epidemiological data to estimate the relevance of inter-annual variability of Norovirus infections. Further validation of this approach with multi-year data (including seasonality) on NV in wastewater versus reported NV infections is warranted. Low sample sizes are a common cause of uncertainty in QMRA studies. Sales-Ortells et al. (2015) based a risk estimate for NV from consumption of lettuce on a sample size similar to ours ($n = 8$). Others based risk

estimates on even smaller dataset, e. g. Amoueyan et al. (2019) or McBride et al. (2013) ($n = 4$). Complementing NoV GII observations with additional information, like long-term reported incidence, most likely leads to a more realistic estimate of the existing variability of NoV GII risk in comparison to deriving input distribution for source water concentration solely from the data collected over a limited period of time. Using incidence to estimate variation of NoV GII concentrations certainly constitutes an interesting alternative to only grouping norovirus data by season, as seasons may be quite different, as our results indicate. In one of the single studies, which measured NV over three years and correlated NV concentrations to reported AGI cases, Kazama et al. (2017) also found a significant positive correlation, which indicates that correlations may hold for more than one year and gives some support for generating plausible risk scenarios from reported incidence. Therefore, if epidemiological information is available it might be beneficial to check whether collected NV wastewater data come from a low, average, or high incidence year. If data have been collected under- low incidence conditions only, an additional safety factor could be considered.

However, our result also underline that there is a need for more long-term monitoring data, to better characterize the between year variability as a source of variation. Local multi-year wastewater data are still the preferred source of information for risk assessment in comparison to deriving scenarios from epidemiological data and are also needed for further validating existing modelling approaches, like ours. In this context, it is important to note, that the relation, i.e. regression parameters, between wastewater concentrations and reported incidence cannot be transferred to other locations as reporting systems and practices may be different between countries. Even within Germany, the federal governmental structure of the reporting system, makes the transfer of derived regression coefficients questionable.

Next to inter-annual variability we also assessed the practical relevance and the risk reduction potential of adapting irrigation periods to low incidence periods taking into account intra-annual variations. Our results suggest that cutting off peak concentrations of NoV GII from the

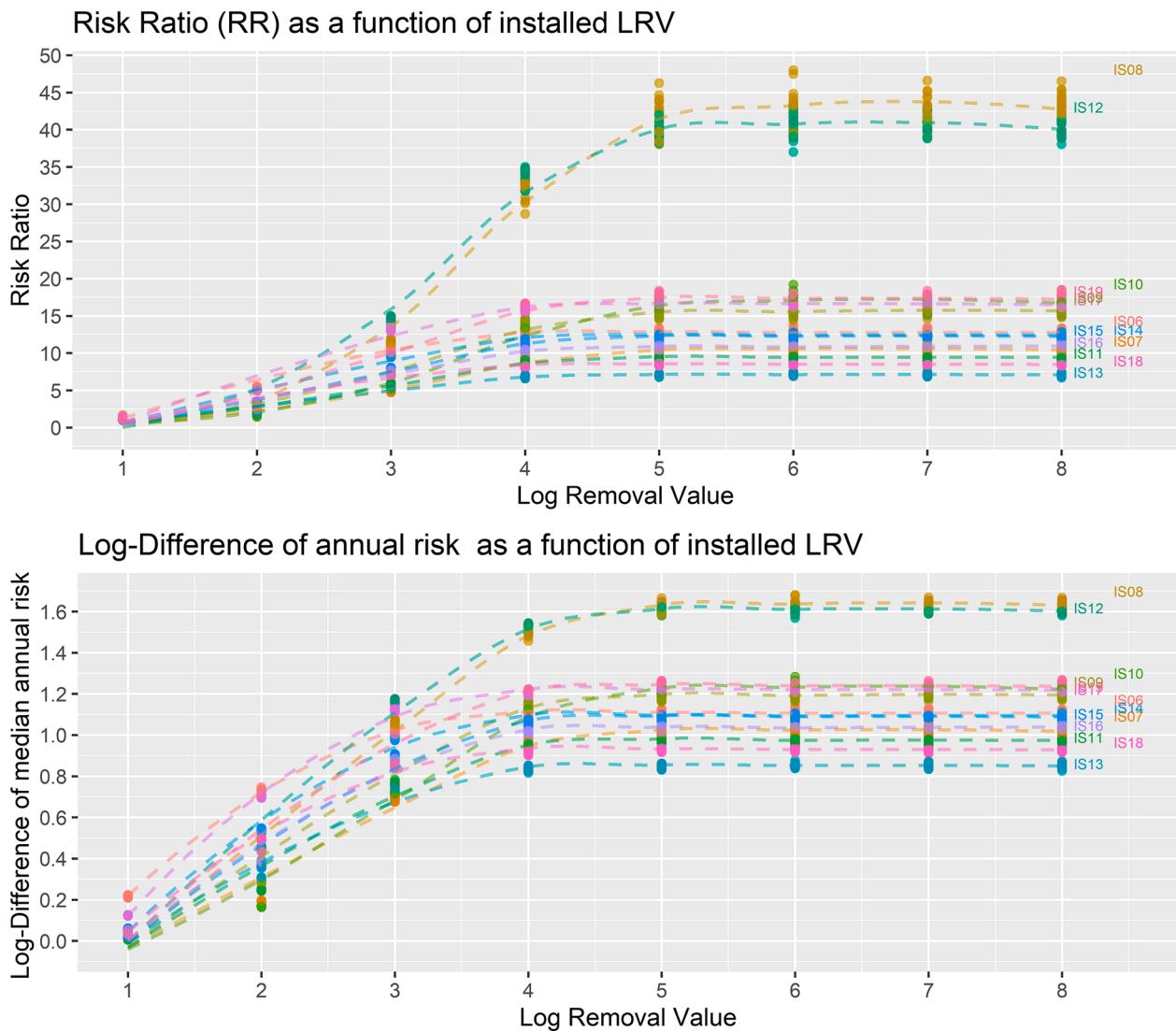


Fig. 5. Risk ratio (upper graph) and log-difference (lower graph) of the simulated median scenario risk. Different colors refer to different incidence scenarios. Individual points refer to the Monte-Carlo error.

exposure period may function as an additional effective risk reduction measure complementing wastewater treatment and disinfection.

However, our result also indicate that, depending on the exposure level, a minimum level of wastewater treatment is necessary in order for this adaption to be most effective. These results may be of practical importance because studies, which aim at deriving required log-reduction values, may account for this additional source of variation. Moreover, while our scenario is tailored to locations in moderate climates, where irrigation is only partially necessary during the vegetation period, the obtained results may also be relevant for seasonal adjustments of pre-treatment options in cases where irrigation is applied all year (e.g. greenhouses), as well as for risk assessments in areas where water use is particularly seasonal, like e.g. bathing water management. Note, however, that for risk management decisions other reference pathogens, like protozoan and bacterial pathogens would need to be considered as well.

Our study makes several simplifying assumptions and disregards several factors. For example, no consideration is given to a potential immune adaption of workers. NoV GII rapidly expresses new variants, which affects immunity, but might also affect infectivity. Indeed, many of the peak incidence seasons can be linked to the occurrence of new variants of Norovirus becoming dominant. For example the high incidence in season 2007/2008 can be attributed the variant 2006b

becoming the dominant norovirus variant; another steep and early rise in Germany is in winter 2016 is connected to a new recombinant variant GII.P16-GII.2 (Niendorf et al., 2017). The effect of new norovirus variants and immune adaption on risk calculations has not been explicitly addressed in water-related QMRA so far and remain subject of future research. A further simplification in our study is that, for isolating the effect of norovirus variation we assumed a constant treatment performance over the whole season. However, log-removal of viruses via biological wastewater treatment may be affected by other factors, like e.g. temperature. Both the under-estimation of large outbreaks by the reporting system, which pre-dominantly occur in winter and the temperature effects are likely to further increase the calculated risk ratios as viruses are known to be more recalcitrant under cold conditions. Thus, while our study already indicated the relevance of seasonal effects, the true effect might be even more pronounced, and should be considered in future research.

5. Conclusion

- Risk scenarios derived from epidemiological data and NoV GII wastewater concentrations indicate that between-year variability is a relevant factor for assessing health risk resulting from NoV GII exposure.

- Annual mean infection risk may vary by > 1 log between years under the same exposure scenario.
- Reducing irrigation to low incidence periods may reduce the risk infection > 1 log units given sufficient pre-treatment.
- The log-removal to achieve the same level of health protection of 1% per period per year is at least 0.9 log between scenarios
- Additional safety factors may be applied in future QMRA studies to account for between-year variability.
- The predictive power and wider use of epidemiological data as a suitable predictor variable should be further validated with paired multi-year data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank Janina Heinze, Heinrich Ripke and the colleagues from the Wastewater Association Braunschweig for their critical feedback of the site description. Local Norovirus data were collected in the research projects and DEMOWARE, funded by the European Union's 7th Framework Programme for Research, Technological Development and Demonstration, theme ENV.2013.WATER INNO&DEMO-1 under grant agreement no 619040.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.119079.

References

- Amoueyan, E., Ahmad, S., Eisenberg, J.N.S., Gerrity, D., 2019. Equivalency of indirect and direct potable reuse paradigms based on a quantitative microbial risk assessment framework. *Microb. Risk Anal.* 12, 60–75.
- Atmar, R.L., Opekun, A.R., Gilger, M.A., Estes, M.K., Crawford, S.E., Neill, F.H., Graham, D.Y., 2008. Norwalk virus shedding after experimental human infection. *Emerg. Infect. Dis.* 14 (10), 1553–1557.
- Bernard, H., Höhne, M., Niendorf, S., Altmann, D., Stark, K., 2014. Epidemiology of norovirus gastroenteritis in Germany 2001–2009: eight seasons of routine surveillance. *Epidemiol. Infect.* 142 (1), 63–74.
- Bürkner, P.C., 2017. brms: an R package for bayesian multilevel models using stan. *J. Stat. Softw.* 1 (1), 2017.
- Calgua, B., Fumian, T., Rusiñol, M., Rodriguez-Manzano, J., Mbayed, V.A., Bofill-Mas, S., Miagostovich, M., Girones, R., 2013a. Detection and quantification of classic and emerging viruses by skimmed-milk flocculation and PCR in river water from two geographical areas. *Water Res.* 47 (8), 2797–2810.
- Calgua, B., Rodriguez-Manzano, J., Hundesa, A., Suñen, E., Calvo, M., Bofill-Mas, S., Girones, R., 2013b. New methods for the concentration of viruses from urban sewage using quantitative PCR. *J. Virol. Methods* 187 (2), 215–221.
- de Graaf, M., van Beek, J., Vennema, H., Podkolzin, A.T., Hewitt, J., Bucardo, F., Templeton, K., Mans, J., Nordgren, J., Reuter, G., Lynch, M., Rasmussen, L.D., Iritani, N., Chan, M.C., Martella, V., Ambert-Balay, K., Vinjé, J., White, P.A., Koopmans, M.P., 2015. Emergence of a novel GII.17 norovirus – end of the GII.4 era? *Euro Surveill.* 20 (26), 21178.
- Eftim, S.E., Hong, T., Soller, J., Boehm, A., Warren, I., Ichida, A., Nappier, S.P., 2017. Occurrence of norovirus in raw sewage – a systematic literature review and meta-analysis. *Water Res.* 111 (Supplement C), 366–374.
- Gelman, A., Goodrich, B., Gabry, J., Vehtari, A., 2019. R-squared for bayesian regression models. *Am. Stat.* 73 (3), 307–309.
- Hellmer, M., Paxeus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergstrom, T., Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis a virus and norovirus outbreaks. *Appl. Environ. Microbiol.* 80 (21), 6771–6781.
- Huang, Y., Zhou, N., Zhang, S., Yi, Y., Han, Y., Liu, M., Han, Y., Shi, N., Yang, L., Wang, Q., Cui, T., Jin, H., 2022. Norovirus detection in wastewater and its correlation with human gastroenteritis: a systematic review and meta-analysis. *Environ. Sci. Pollut. Res.* 1–14.
- Kageyama, T., Kojima, S., Shinohara, M., Uchida, K., Fukushi, S., Hoshino, F.B., Takeda, N., Katayama, K., 2003. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *J. Clin. Microbiol.* 41 (4), 1548–1557.
- Kazama, S., Miura, T., Masago, Y., Konta, Y., Tohma, K., Manaka, T., Liu, X., Nakayama, D., Tanno, T., Saito, M., Oshitani, H., Omura, T., 2017. Environmental surveillance of norovirus genogroups i and ii for sensitive detection of epidemic variants. *Appl. Environ. Microbiol.* 83 (9).
- Loisy, F., Atmar, R.L., Guillon, P., Le Cann, P., Pommepuy, M., Le Guyader, F.S., 2005. Real-time RT-PCR for norovirus screening in shellfish. *J. Virol. Methods* 123 (1), 1–7.
- McBride, G.B., Stott, R., Miller, W., Bambic, D., Wuertz, S., 2013. Discharge-based QMRA for estimation of public health risks from exposure to stormwater-borne pathogens in recreational waters in the United States. *Water Res.* 47 (14), 5282–5297.
- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., Brouwer, A., 2020. Presence of SARS-coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in the Netherlands. *Environ. Sci. Technol. Lett.* 7 (7), 511–516.
- Messner, M.J., Berger, P., Nappier, S.P., 2014. Fractional poisson—a simple dose-response model for human norovirus. *Risk Anal.* 34 (10), 1820–1829.
- Miura, T., Kazama, S., Konta, Y., Masago, Y., Tohma, K., Manaka, T., Liu, X., Saito, M., Oshitani, H., Omura, T., 2016. Applicability of norovirus monitoring in sewage as an early warning system of infectious gastroenteritis. *J. Jpn. Soc. Civ. Eng. Ser. G (Environ. Res.)* 72 (7), III 285–III 294.
- Nappier, S.P., Soller, J.A., Eftim, S.E., 2018. Potable water reuse: what are the microbiological risks? *Curr. Environ. Health Rep.* 5 (2), 283–292.
- Niendorf, S., Jacobsen, S., Faber, M., Eis-Hübinger, A.M., Hofmann, J., Zimmermann, O., Höhne, M., Bock, C.T., 2017. Steep rise in norovirus cases and emergence of a new recombinant strain GII.P16-GII.2, Germany, winter 2016. *Euro Surveill.* 22 (4), 30447.
- R Development Core Team, 2008. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 URL <http://www.R-project.org>.
- Richards, G.P., Watson, M.A., Meade, G.K., Hovan, G.L., Kingsley, D.H., 2012. Resilience of norovirus GII.4 to freezing and thawing: implications for virus infectivity. *Food Environ. Virol.* 4 (4), 192–197.
- Sales-Ortells, H., Fernandez-Cassi, X., Timoneda, N., Dürig, W., Girones, R., Medema, G., 2015. Health risks derived from consumption of lettuces irrigated with tertiary effluent containing norovirus. *Food Res. Int.* 68, 70–77.
- Schmidt, P.J., 2015. Norovirus dose-response: are currently available data informative enough to determine how susceptible humans are to infection from a single virus? *Risk Anal.* 35 (7), 1364–1383.
- Seis, W., Rouault, P., Medema, G., 2020. Addressing and reducing parameter uncertainty in quantitative microbial risk assessment by incorporating external information via Bayesian hierarchical modeling. *Water Res.* 185, 116202.
- Soller, J.A., Eftim, S.E., Nappier, S.P., 2018. Direct potable reuse microbial risk assessment methodology: sensitivity analysis and application to State log credit allocations. *Water Res.* 128, 286–292.
- Soller, J.A., Eftim, S.E., Warren, I., Nappier, S.P., 2017. Evaluation of microbiological risks associated with direct potable reuse. *Microb. Risk Anal.* 5, 3–14.
- StanDevelopmentTeam 2017 *Stan Modeling Language Users Guide and Reference Manual, Version 2.17.0*. <http://mc-stan.org>.
- Stegmaier, T., Oellingrath, E., Himmel, M., Fraas, S., 2020. Differences in epidemic spread patterns of norovirus and influenza seasons of Germany: an application of optical flow analysis in epidemiology. *Sci. Rep.* 10 (1), 14125.
- Teunis, P.F.M., Le Guyader, F.S., Liu, P., Ollivier, J., Moe, C.L., 2020. Noroviruses are highly infectious but there is strong variation in host susceptibility and virus pathogenicity. *Epidemics* 32, 100401.
- Teunis, P.F.M., Moe, C.L., Liu, P., Miller, S.E., Lindesmith, L., Baric, R.S., Le Pendu, J., Calderon, R.L., 2008. Norwalk virus: how infectious is it? *J. Med. Virol.* 80 (8), 1468–1476.
- Teunis, P.F.M., Sukhrie, F.H.A., Vennema, H., Bogerman, J., Beersma, M.F.C., Koopmans, M.P.G., 2015. Shedding of norovirus in symptomatic and asymptomatic infections. *Epidemiol. Infect.* 143 (8), 1710–1717.
- Van Abel, N., Schoen, M.E., Kissel, J.C., Meschke, J.S., 2017. Comparison of risk predicted by multiple norovirus dose-response models and implications for quantitative microbial risk assessment. *Risk analysis: an official publication of the society for Risk Anal.* 37 (2), 245–264.
- Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27 (5), 1413–1432.
- WHO, 2016. *Quantitative Microbial Risk Assessment: Application for Water Safety Management*. World Health Organisation, Geneva, Switzerland.
- WHO, 2017. *Potable Reuse: Guidance for Producing Safe Drinking-Water*. World Health Organization, Geneva.
- Wilking, H., Spitznagel, H., Werber, D., Lange, C., Jansen, A., Stark, K., 2013. Acute gastrointestinal illness in adults in Germany: a population-based telephone survey. *Epidemiol. Infect.* 141 (11), 2365–2375.