EÖTVÖS LORÁND UNIVERSITY

Faculty of Sciences

Doctoral School of Environmental Sciences

Program of Environmental Biology

VIRUS MONITORING IN ENVIRONMENTAL WATERS – NEW DIRECTIONS IN EPIDEMIOLOGY

Theses of doctoral dissertation

Eszter Róka

Leader of doctoral school: Dr. Tamás Turányi

Program leader: Dr. Erika Tóth

Supervisor: Dr. Márta Vargha

National Public Health Center

Department of Environmental Health Laboratories, Budapest

2022

Introduction

By the mid-19th century it became clear, that many pathogens can be transmitted via water. Identification of the infection source is critical to prevent spread of the epidemic and reduce the risk of further outbreaks. Beyond traditional epidemiological methods (epidemic surveys, investigation the dynamics of the epidemic) microbiological methods also provide assistance: detection of the pathogen from the environmental source is considered a decisive evidence. Environmental investigation is a longstanding practice in the case of easy-to-culture bacterial pathogens. The nucleic acid-based methods and new concentration methods made a breakthrough also in the case of viruses.

Although the number of waterborne infectious diseases is decreasing due the advanced drinking water and wastewater treatment, their importance is still not negligible. The risk of infection is still present especially in case of extreme weather conditions or use of alternative water sources. In addition, pool waters with insufficient treatment or contaminated natural bathing waters can also pose a risk.

However, in wastewater pathogens are not only a health risk: by detecting their presence or monitoring their quantity important information can be obtained about the population's health status. Fast and reliable tracking of SARS-CoV-2 in wastewater gained prime importance during the COVID-19 pandemic.

SARS-CoV-2 is a respiratory pathogen, but it is also shed in feces. Symptomatic and asymptomatic patients shed the virus from the early stages of infection. Thus it can be detected in sewage even before infected people would appear in the healthcare system. From early 2020, several research groups started to work on SARS-CoV-2 detection from wastewater, and most of them found strong correlation with clinical data.

From the second half of 2020, attention focused on new variants of SARS-CoV-2 which were more transmissible and caused more severe disease. Beyond tracking and predicting the number of infections, detection of variants became also a primary objective of wastewater investigations.

Objectives

In this study my objectives were the following:
- Method optimization for detecting viruses form drinking water, surface water and wastewater

- Detection of viruses from water as presumptive source of infection in enteric epidemics
- Through this, reduce the risk of further infections
- Method development for detecting new coronavirus (SARS-CoV-2) form wastewater
    - Develop a method, which is reliably available and can be optimized to adequate throughput
    - Optimize an available and effective nucleic acid extraction method
- Investigation of the potential use of SARS-CoV-2 detection in wastewater for outbreak management
    - Optimization of data processing to obtain reliable and stable quality data for the estimaton of trends in COVID-19 case numbers
    - Assessment of the correlation between the results and COVID-19 case numbers on city level, estimation of predictive potential
    - Assessment of the correlation between the results and COVID-19 case numbers (including hospitalized cases) on national level, estimation of predictive potential
- Comparison of the predictive potential of sewage data in different waves of the COVID-19 pandemic
- Investigation of SARS-CoV-2 variants of concern from wastewater
    - Optimization of a fast and effective method
    - Comparison of results with clinical data

Materials and methods

For investigation of enteric outbreaks, four samplings were carried out in the study period:

- From a small lake in the Great Hungarian Plain on 13. 07. 2015 (Lake1) because of a calicivirus epidemic
- From three therapeutic water wells near a hospital (K1, K2, K3) on 18. 07. 2016. because of a calicivirus epidemic
- From a private well of a family (K4) at 26. 07. 2016. because a family cluster of hepatitis A cases
- From two pools of a spa (M1, M2) at 15. 02. 2017. because of a calicivirus epidemic

Samples were concentrated in the laboratory by skim milk flocculation then nucleic acid was extracted; presence of pathogens was detected with taxon-specific PCR and agarose gel electrophoresis. One positive and one negative control sample was also analyzed with every sample series to check sample processing.

For SARS-CoV-2 detection from wastewater, weekly samples were collected from Budapest and every county seat from the beginning of July 2020. Several cities from Budapest's agglomeration were also sampled weekly from August 2020; while from October 2021 two samples a week were taken from the biggest cities. Data were analyzed until the end of October 2021, but national average data were extended until December 2021 to compare of pandemic waves.

Previously methodology was only available in the laboratory for concentration of non-enveloped viruses. Therefore several methods were compared for SARS-CoV-2 concentration (skim milk flocculation, ultrafiltration with commercial available membranes, ultrafiltration with custom-developed flat sheet membrane, concentration on electronegative membrane, virus recovery from sedimented sewage and other commercially available methods). In the beginning of the investigation, availability of nucleic acid extraction kits was also limited. Thus two general kits (e.g. recommended for serum) and two kits recommended for fecal and soil samples were tested. Quantification of SARS-CoV-2 was carried out by RT-qPCR, with N1 primer sett recommended by CDC.

The quality of wastewater can show significant fluctuations. The degree of dilution was followed by quantifying a fecal indicator (fecal *Enterococci*) and through the collection of data on the daily volume of sewage influent in wastewater treatment plant.

The number of COVID-19 cases was obtained from the official database of the National Public Health Center (OSZIR – Országos Szakmai Információs Rendszer – National Professional Information System) broken down to city level. The number of hospitalized cases was obtained from official reports of the Operative Board (aggregated on national level).

For statistical analysis of the results, Microsoft Excel (version: 2110) and TIBCO Statistica software (version:14.0.0.15) were used. Two-sample paired t-test was used to method comparison, while for examination of the correlation between wastewater SARS-CoV-2

concentrations and COVID-19 cases linear regression and Almon polynomial distributed lag analysis was used.

Results and discussion

1. Detection of waterborne enteric viruses did not give an unambiguous result in all cases. No PCR product was detected in the sample from the lake in the Great Hungarian Plane, thus other transmission medium can be presumed. In case of pool waters, PCR product was detected, however, not exactly in the expected length. Laboratory results and epidemiological evidences were both poor in this case; thus water origin can neither be confirmed nor ruled out. PCR product was detectable in the therapeutic water wells in the calicivirus outbreak investigation, even in ten-fold dilution for two out of three samples. In this case, epidemiological evidences were also clear; thus the source of infections was undoubtedly the therapeutic water well. Water origin was not confirmed in the case of familiar hepatitis A cluster, other infection routes can be presumed among family members.

2. Detection of waterborne enteric viruses is an important method in source identification of outbreaks; therefore, further development of the tests is necessary primarily for quantitative analysis.

3. During method development, several virus concentration and nucleic acid extraction methods were tested for detection of SARS-CoV-2 from wastewater. Skim milk flocculation method, which was used previously for non-enveloped enteric viruses in the laboratory, was not found to be suitable for the detection of the enveloped SARS-CoV-2. Only two parallel concentrations was carried out with this method, these indicated 0.8 and 12% recovery compared to the positive control sample inoculated with the same amount of virus after the concentration step. Best results were obtained with ultrafiltration methods. The commercial Centriprep (Merck KGaA, Darmstadt, Germany) and the custom-developed flat sheet membrane (SUEZ WTS, Tatabánya, Hungary) showed comparable good result with 80% and 96% recovery, respectively. However, there were significant differences in the availability of the devices and the work demand of sample processing for the advantage of the sheet membrane.

Concentration on electronegative membrane and virus recovery from sedimented wastewater were found to be simple and rapid methods, however, their recovery was significantly worse for real samples, than that of the flat sheet ultrafiltration method, which was routinely used by the time these methods were tested. The commercial virus concentrating kit

gave similar results to ultrafiltration in the most (12/18) samples, but in 6 samples more than one order of magnitude difference was observed. Based on long term availability, recovery and optimization of laboratory working processes, flat sheet ultrafiltration method was selected for virus concentration.

Choosing of nucleic-acid extraction method was also difficult, because in the beginning of the project, these kits were also difficult to obtain. Two general purpose kits (recommended e.g. for serum) and two kits recommended for fecal and soil samples were tested. Though kits for fecal and soil sample processing theoretically have better inhibitor removal efficiency, worse recovery was observed compared to general kits. In addition, the time demand of sample processing was also significantly higher for fecal kits. Therefore one of the general kits (Viral Mini kit, Qiagen, Hilden, Germany) was used for further measurements.

The developed methodology was found to be reliably available and suitable for large number of samples. The selected ultrafiltration method has adequate recovery (96%) and sensitivity (theoretical limit of detection: 1710 GC/L) for epidemiological utilization of the data.

4. For comparison of SARS-CoV-2 concentration in wastewater and clinical data, time period between July 2020 and October 2021 was chosen. In this period, 1738 samples were processed in the laboratory. Further one and a half months of data were used for a more accurate evaluation of the 4th pandemic wave (national average based on 209 samples).

a. Since results can be variable due to the inaccuracy of sampling (especially in case of grab samples), the inhomogeneity of emission and the uncertainty of the measurement, selection of a suitable data processing method for reducing of this biased was necessary as first step. To reduce the effect of wastewater inhomogeneity and dilution, data normalized using daily wastewater volumes or calculating fecal load by measuring fecal *Enterococci* count. Three point rolling average was calculated to reduce sampling and measurement uncertainty. After that, correlation (linear regression) was calculated between wastewater SARS-CoV-2 genome copies and the number of new COVID-19 cases registered in the service area of the wastewater treatment plant using several data processing methods. Strongest correlation was observed using 3 point rolling average of the logarithmic transformed and fecal *Enterococci* normalized wastewater results.

Using this normalization and smoothing method, wastewater results showed strong correlation with new and cumulated COVID-19 cases in most cities. (Pearson-correlation, $r^2=0.75 - 0.42$; value below 0.5 was only observed for three cities. Correlation is significant in all cases ($p<0.05$). However, clear differences were observed in the strength of correlation in the individual county seats. Looking for the reasons behind the variability, good results were

observed in all wastewater treatment plants using composite auto-sampler (Pearson-correlation, r^2=0.75 – 0.58), however, similar good results were observed in many cases where grab sampling was used in the wastewater treatment plant. (The correlation falls within the above range in 10 out of 15 wastewater treatment plants with grab sampling method.) Comparing the other characteristics of wastewater treatment plants, the size of supplied population was not a significant factor: higher uncertainty was also observed in some of the largest cities (e.g. Debrecen). The potential impact of the exact sampling point was also investigated, but according to the information obtained from the operators, the same point, raw sewage after the coarse grid was sampled in every plant. Probable reason for the observed differences can be the presence of inhibitors, periodic industrial emission or other unknown factors. Further investigations are necessary to identify the confounding factors.

b. Almon polynomial distributed lag analysis was used to investigate the lag time between wastewater SARS-CoV-2 copy number and clinical data. This model is used to determine the effect of different lag periods on the dependent variable. Looking at the national average, strongest correlation was observed with new COVID-19 cases 1 week, while hospitalized cases 2 weeks ahead (Figure 1). In the case of the county seats, reliable data were only available about the new registered cases. For most cities, the correlation was also strongest with case numbers one week ahead, however, in many cases the correlation is also similar with two weeks lead.

c. Comparing the second, third and fourth pandemic waves in Hungary, the correlation between wastewater and clinical data is influenced by multiple factors. Average age of infected people was much lower in the beginning of second wave than later. The younger individuals are more likely to develop asymptomatic infection or milder symptoms, therefore, they are more rarely diagnosed. The characteristics of the dominant SARS-CoV-2 variant can affect the average virus shedding per infected people and also the severity of symptoms. Both affect the correlation between the number of cases and the wastewater data. Testing strategy had also changed in the investigated one and a half years. In the autumn of 2020, all test were carried out using the PCR method, which is sensitive, but requires laboratory equipment, and thus it is lengthy. From the beginning of 2021, the proportion of antigen rapid test gradually increased. These tests are much faster, and could significantly decrease the time between the symptom onset and diagnosis. By the end of third wave, 60% of the population was already vaccinated: this slowed down the disease transmission, and also reduced the symptoms of the COVID-19 cases. However, investigating the lag period between wastewater data and new and hospitalized cases, no significant shift was observed during the three pandemic waves.
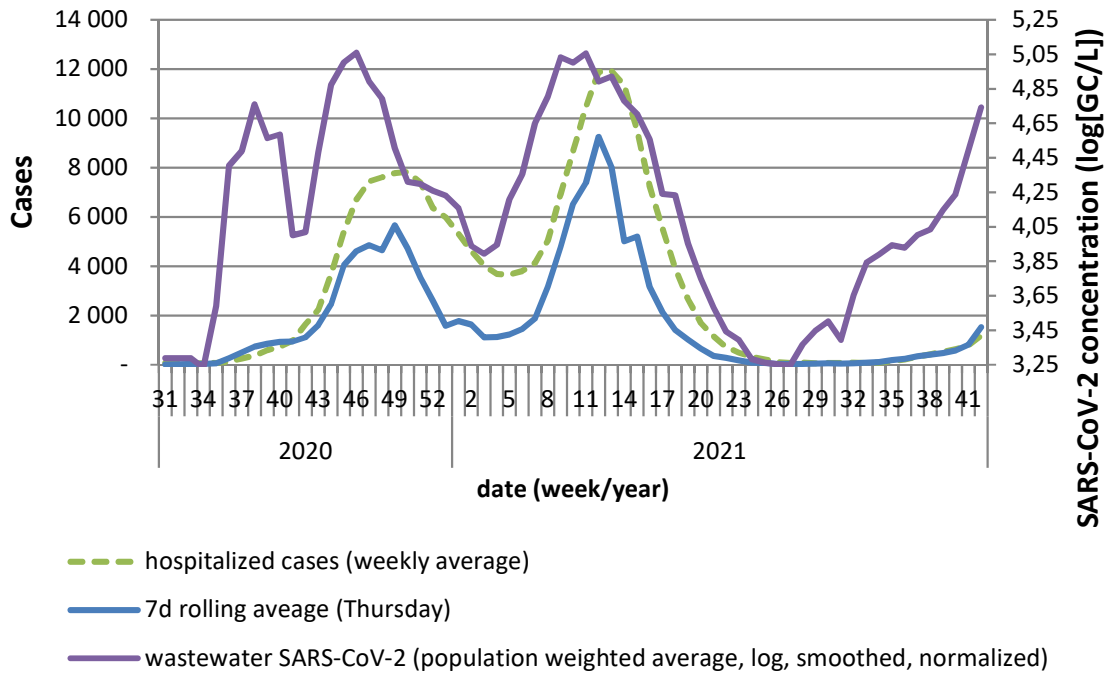
Figure 1: Population weighted average of wastewater SARS-CoV-2 results (normalized to fecal Enterococci, 3 point rolling average), the new and hospitalized COVID-19 cases between 01. 08. 2020. and 31. 10. 2021.

d.   Though lag periods between wastewater and clinical data were similar in the investigated pandemic waves, the same SARS-CoV-2 genome copy number in wastewater corresponded to significantly different registered case numbers in different pandemic waves. As a consequence, neither the exact number of the new COVID-19 cases nor the number of hospitalized cases can be estimated based solely on wastewater data, since these indicators are also affected by other conditions of the pandemic (e.g. average age of infected individuals, the characteristic virus variant and the vaccination status of the population).

5.   The identification of the circulating SARS-CoV-2 variants was carried out using ddPCR. The advantage of this method is that it generates fast and quantitative result about the most characteristic mutations of the investigated variant. Detailed measurements were carried out in case of the alpha variant: from the beginning of December 2020 on the samples from Budapest and from the middle of January for all other sampling sites, until the middle of March. (The first known alpha case in Hungary was identified in mid-January.) Weak positive results were found in some of the samples from 2020 for the tested mutations (N501Y and delH69/V70), however, these mutation can be also found in other lineages, therefore this is not an unequivocal evidence for the presence of alpha variant in 2020. From January, 2021, results show strong correlation with the ratio of the alpha variant in the clinical data (Pearson-correlation, R=0.921, p<0.05; Figure 2). Disadvantage of the method that it can only detect known variants, it needs special consumables for every new variant and their procurement can

be time-consuming. The most reliable method for the variant identification is next generation sequencing, but this requires further method development in the case of complex environmental samples.
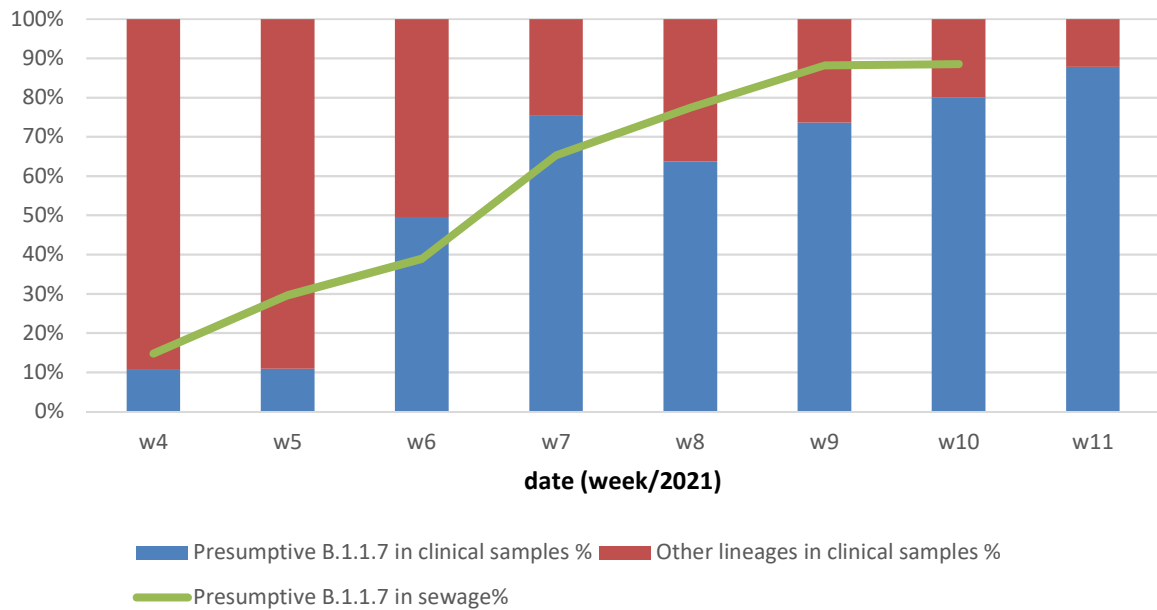

Figure 2: Percentage of presumptive lineage B.1.1.7 (based on N501Y and del H69/V70 mutations screening) in clinical samples and the same mutations measured in wastewater (national average).

6. For the practical utilization of the results, effective communication of the data to the decision makers and to the public is essential. The Hungarian wastewater surveillance data was primarily used for to predict the expected burden on the health care system: the two reports per week were extremely important information in the most severe pandemic waves. The public communication reached a wide audience, and was frequently cited in the media: this played an important role in raising public awareness.

Publication:

The theses based on the following publications:

Róka, Eszter, Bernadett Khayer, Zoltán Kis, Luca Bella Kovács, Eszter Schuler, Nóra Magyar, Tibor Málnási, és mtsai. 2021. „Ahead of the Second Wave: Early Warning for COVID-19 by Wastewater Surveillance in Hungary". *Science of The Total Environment* 786 (szeptember): 147398. https://doi.org/10.1016/j.scitotenv.2021.147398.

Róka, Eszter, Dániel Déri, Bernadett Khayer, Zoltán Kis, Eszter Schuler, Nóra Magyar, Bernadett Pályi, Tamás Pándics, és Márta Vargha. 2022. „SARS-CoV-2 Variant Detection from Wastewater: Rapid Spread of B.1.1.7 Lineage in Hungary". *Journal of Water and Health*, január, jwh2022179. https://doi.org/10.2166/wh.2022.179.

Róka Eszter, Khayer Bernadett, Kis Zoltán, Kovács Luca Bella, Magyar Nóra, Málnási Tibor, Oravecz Orsolya, Pályi Bernadett, Pándics Tamás, Schuler Eszter, Vargha Márta. 2021. „A járványkezelés új eszköze: a szennyvíz alapú epidemiológia". Vízügyi Közlemények 103 (2) 187-206.

Navarro, Anna, Livia Gómez, Isabella Sanseverino, Magdalena Niegowska, Eszter Roka, Rosalba Pedraccini, Marta Vargha, és Teresa Lettieri. 2021. „SARS-CoV-2 Detection in Wastewater Using Multiplex Quantitative PCR". *Science of The Total Environment* 797 (november): 148890. https://doi.org/10.1016/j.scitotenv.2021.148890.

Pándics, Tamás, Eszter Róka, Bernadett Khayer, Zoltán Kis, Luca Bella Kovács, Nóra Magyar, Tibor Málnási, és mtsai. 2021. „A szennyvíz alapú epidemiológia jelentősége a COVID–19 járványban és azon túl". *Scientia et Securitas* 2 (1): 30–37. https://doi.org/10.1556/112.2021.00005.

Pándics Tamás, Róka Eszter, Henczkó Judit, Khayer Bernadett, Kis Zoltán, Málnási Tibor, Pályi Bernadett, Schuler Eszter, Vargha Márta. „A hazai szennyvíz alapú COVID-19 előrejelző rendszer - másfél év tanulságai" Közlésre elfogadva: *Népegészségügy*