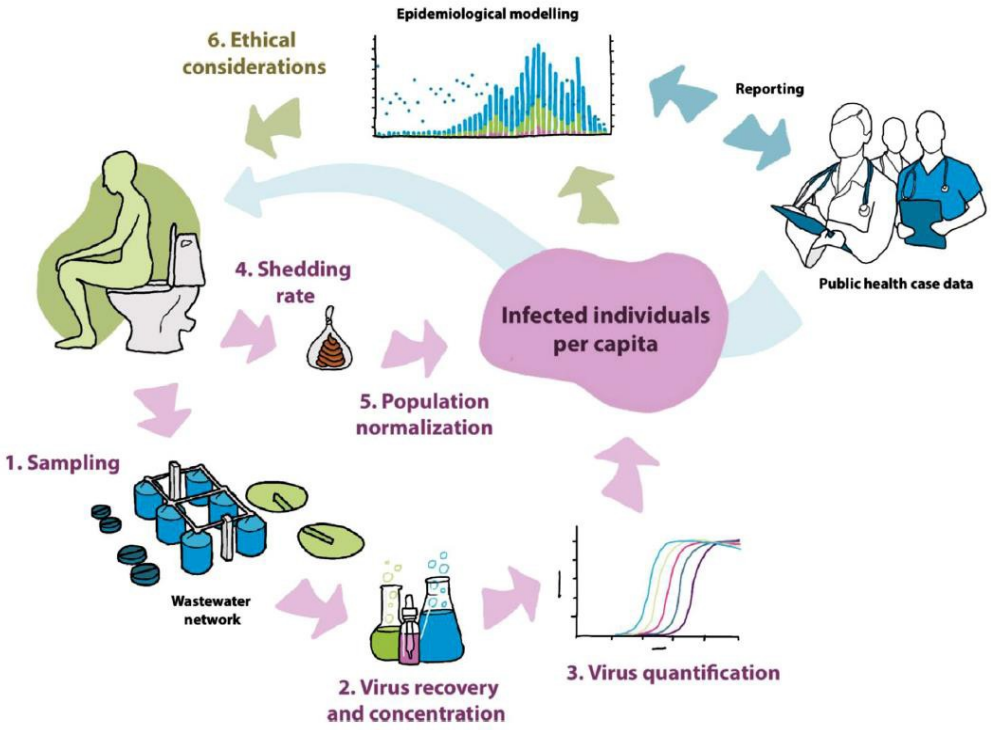
Pierādījumos balstītas rekomendācijas (ieteikumi) SARS-CoV-2 izplatības monitoringam un ierobežošanai, ņemot vērā tā klātbūtni apkārtējā vidē

2021

Pierādījumos balstītas rekomendācijas (ieteikumi) SARS-CoV-2 izplatības monitoringam un ierobežošanai, ņemot vērā tā klātbūtni apkārtējā vidē, ir sagatavotas Valsts pētījumu programmas "COVID-19 seku mazināšanai" **projekta Nr. VPP-COVID-2020/1-0008 "Multidisciplināra pieeja COVID19 un citu nākotnes epidēmiju monitorēšanai, kontrolei un ierobežošanai Latvijā"** ietvaros laika periodā no 2020.gada 1.jūlija līdz 2021. gada 31.martam. Projekta WP4 darba pakas mērķis bija izstrādāt kompleksus risinājumus koronavīrusu, t.sk. SARS-CoV-2, monitoringam vidē (notekūdeņos), kur viens no uzdevumiem bija rekomendāciju izstrāde tālākai to izmantošanai notekūdeņu monitoringa ieviešanai Latvijā esošās COVID-19 pandēmijas laikā, kā arī citu nākotnes epidēmiju monitorēšanai, kontrolei un ierobežošanai Latvijā. Rekomendāciju izstrādē piedalījās Rīgas Tehniskās universitātes (RTU), Latvijas Biomedicīnas pētījumu un studiju centra un Pārtikas drošības, dzīvnieku veselības un vides zinātniskā institūta "BIOR" eksperti.

Eiropas Komisija ir izstrādājusi vadlīnijas WBE dalībvalstīm, kuras paredzēts ieviest līdz 2021. gada 1. oktobrim. Šis rekomendācijas ir pamats nacionālo vadlīniju izstrādei.

Notekūdeņu monitoringā balstīto epidemioloģiju (*Wastewater-based epidemiology, WBE*) pielieto kā agrīnās brīdināšanas sistēmu, kas sniedz iespēju savlaicīgi apzināt COVID-19 uzliesmojuma iestāšanos, tā apmērus konkrētā populācijā vai reģionā un sekot līdzi populācijā izplatīto vīrusa variantu klāstam. Mūsu pētījuma rezultāti parāda, ka, pielietojot WBE, infekcijas uzliesmojumu apdzīvotās vietās (ar centralizētu kanalizāciju) var identificēt līdz nedēļu ātrāk, nekā izmantojot tikai klīniskās epimeioloģijas pieeju. (1. attēls)



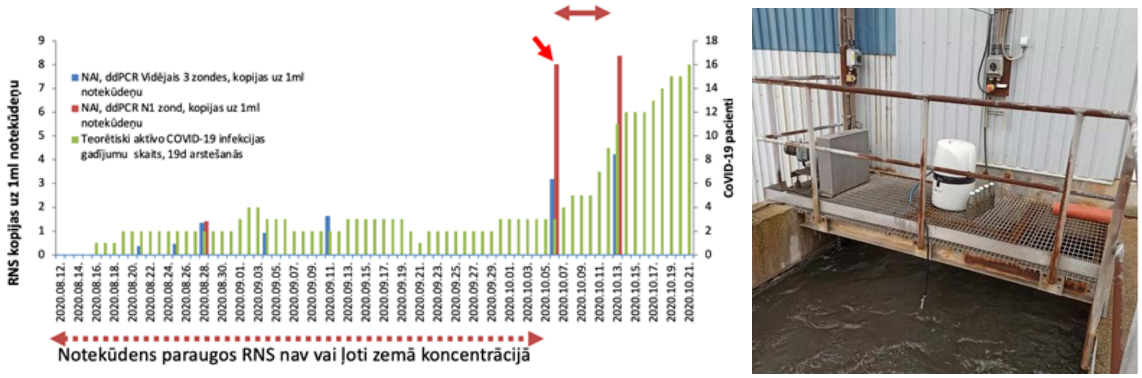
1. attēls. Covid-19 uzraudzība pašvaldības/kopienas līmenī ar notekūdeņu izmeklējumos balstītu strukturētu pieeju epidemioloģijā. Avots: Polo D., Quintela-Baluja M., et al. 2020 (In: *Water Research*)

** WBE izmanto kā metodi infekcijas izplatībastendenču noteikšanai un var tik lietota kopā ar precīzākām metodēm.*

SARS-CoV-2 klātbūtnes noteikšana tiek veikta, izmantojot kvantitatīvo PĶR (kPĶR), ko pielieto arī šīs saslimšanas diagnozes apstiprināšanai.

WBE sekmīgai pieejas izmantošanai ir nepieciešams mazāks paraugu skaits un ir iespējams būtiski samazināt nepieciešamos darbaspēka un finanšu resursus epidemioloģiskās situācijas raksturošanai un izvērtēšanai, kā arī iegūtā informācija sniedz pilnvērtīgāku ieskatu konkrētās teritorijas empidemioloģiskajā situācijā, jo nav atkarīga medicīniskā personāla noslodzes, indivīdu motivācijas ziņot par saslimšanu vai pašvaldības iesaistes līmeņa situācijas kontrolē. Analizējot SARS-CoV-2 RNS klātbūtni WBE ietvaros, nenotiek tieša speciālistu un slimnieku saskarsme ar slimību izraisošo vīrusu, līdz ar to tiek būtiski samazināti riski, kas saistīti ar iespējamo speciālistu inficēšanos. Pirmās valstis, kurās šogad tika ziņots par iespējām izmantot notekūdeņus epidemioloģiskiem nolūkiem, bija Nīderlande, Austrālija, Itālija ASV u.c. Pieejamā notekūdeņu kPĶR metodoloģija balstīta uz novērojumu, ka vidēji 15-83 % no ar SARS-CoV- 2 inficēto pacientu vīrusa RNS ir nosakāms fekālijās (Foladari et al 2020), pat ja netiek novēroti gastrointestinālie simptomi. Turklāt fekāliju paraugi var uzrādīt pozitīvu rezultātu pat tad ja no elpceļiem ņemtajos paraugos vīrusa klātbūtne nav nosakāma.

Šo iemeslu dēļ SARS-CoV-2 klātbūtne inficēto cilvēku izdalījumos, t.sk. fekālijās, notekūdeņu savākšanas un attīrīšanas sistēmās ir globāli piesaistījusi epidemiologu un pētnieku uzmanību. 2020.gada laikā Latvijā tika ievākti vairāk nekā 80 notekūdeņu paraugi no 15 pilsētām, lai izvērtētu SARS-CoV-2 klātbūtni notekūdeņos. Tā kā arī mūsu pētījumā iegūtie rezultāti (2. attēls) apstiptināja, ka vīrusa RNS klātbūtni notekūdeņu attīrīšanas sistēmās var konstatēt pat 5 līdz 7 dienas pirms uzliesmojuma vai epidēmijas konkrētajā pilsētā, šo pieeju var izmantot attiecīgā vīrusa monitoringam ilgstošākā laika periodā (2. attēls).



2. attēls. SARS-CoV-2 vīrusa RNS koncentrācijas dinamika Jelgavas notekūdeņos un to saistība ar saslimšanas gadījumu pieaugumu pilsētā X (pa kreisi) un notekūdeņu paraugu ņemšanas aprīkojums, tā ekspluatācija un diennakts paraugu noņemšanas process konkrētās pilsētas notekūdeņu attīrīšanas stacijā (pa labi).

Pierādīts, ka ir arī iespējams identificēt pilsētas rajonu, kurā var palielināties saslimušo skaits, kas ļauj, savukārt īstenot mērķtiecīgus un lokālus drošības pasākumus konkrētajā pilsētas vai pašvaldības rajonā vai pat institūcijas līmenī (piem., pansionāti, cietumi u.c.)

Vienlaikus VPP projekta realizācijas laikā Institūtā "BIOR" tika veikta hromatogrāfijas un masspektrometrijas parametru optimizācija, kā arī izvēlēta paraugu sagatavošanas procedūra, pielietojot cietās fāzes ekstrakcijas tehnoloģiju. Gala metode iekļauj vairāku biomarķieru (5-HIAA, kotinīns, kofeīns, ibuprofēns) noteikšanu notekūdens paraugos. Analizējot vairāku pilsētu notekūdens paraugus, tika iegūtas 5-HIAA koncentrācijas (8-22 µg/L), kas atbilst literatūras datiem par biomarķieru saturu citās pasaules pilsētās. Tika novērota arī korelācija starp kotinīna un 5-HIAA saturu, kas liecina par analītiskās procedūras piemērotību pētījumiem par populācijas apjomu. Projekta ietvaros izstrādāto noteikšanas metodi ir iespējams papildināt ar farmaceitisko preparātu lietošanas biomarķieriem, iegūstot papildus informāciju par iedzīvotāju sabiedrības veselības stāvokli.

Visnozīmīgākā SARS-CoV-2 kontekstā ir 5-HIAA analīze, jo, izmantojot šo mērījumu, tika veikta metodoloģijas izstrāde populācijas apjoma noteikšanai, kas vienlaikus deva iespēju novērtēt vai novērotās vīrusa daudzuma izmaiņas ir saistāmas ar saslimušo skaita izmaiņām vai kopējā iedzīvotāju skaita izmaiņām apdzīvotajā vietā vai iestādē. Apvienojot vidējo diennakts notekūdeņu plūsmas apjomu un vidēji viena cilvēka izdalīto attiecīgā biomarķiera daudzumu var aprēķināt aptuveno iedzīvotāju skaitu, kas ikdienā lieto pilsētas notekūdeņu sistēmu.

Informācija par biomarķieru noteikšanas analītiskās metodes parametriem ir iekļauta zinātniskajā publikācijā "*Rapid determination of pharmaceuticals in wastewater by direct infusion HRMS using target and suspect screening analysis*" (doi.org/10.1016/j.scitotenv.2020.142688).

IETEICAMĀ METODOLOĢIJA NOTEKŪDEŅU MONITORINGAM

Paraugu atlase, ņemšana un loģistika

Notekūdeņu paraugus ievāc, īstenojot 24 stundu kombinēto notekūdeņu paraugu ņemšanu, ko iespējams īstenot ar atbilstošām paraugu ņemšanas iekārtām (3. attēls). Tādejādi ir iespējams iegūt vienu diennakts notekūdeņu kvalitāti reprezentējošu paraugu, šāda veida paraugu ņemšana ir izplatīta ikdienas notekūdeņu attīrīšanas procesu efektivitātes monitoringa veikšanā, līdz ar to plaši pārbaudīta pasaules mērogā. Kā arī notekūdeņu sistēmu inženieru un zinātnieku vidū ir pieņemts, ka šāda parauga faktiskā diennakts notekūdeņu kvalitātes dinamikas un absolūtā piesārņojuma reprezentācija ir ar augstu ticamību. Paraugu ņemšanas iekārtas pašvaldību uzņēmumiem tiek nodrošinātas monitoringa programmas ietvaros. Lai īstenotu visaptverošu monitoringu, notekūdeņu paraugus plānots ņemt 1 reizi nedēļā. Gadījumos, ja tiek konstatēts iespējams saslimstību uzliesmojums vai uzsākta intensīva iedzīvotāju vakcinācija attiecīgajā pašvaldībā, paraugu ņemšanas regularitāte var tikt palielināta. Notekūdeņu paraugu ņemšanu notekūdeņu attīrīšanas iekārtās (NAI) nodrošina pašvaldības ūdenssaimniecības darbinieki. Parauga ievākšana tiek uzsākta attiecīgās dienas 10:00 un parauga ievākšana notiek līdz nākamās

dienas (pēc 24 stundām) 10:00 (ne vēlāk), lai nodrošinātu parauga laicīgu nogādāšanu analīžu veikšanas vietā. Vienā paraugu ņemšanas reizē nepieciešams ievākt 2 litru kombinēto 24 stundu notekūdeņu paraugu. Pēc paraugu ņemšanas tā transportēšanas laikā nepieciešams nodrošināt līdz 6 °C temperatūru. Notekūdeņu paraugi, izmantojot BIOR rīcībā esošo loģistikas sistēmu, tiek nogādāti reģionālajos paraugu pieņemšanas centros un secīgi Rīgā analīžu veikšanai.

3. Attēls. 24 stundu kombinēto notekūdeņu paraugu ņemšanas iekārta notekūdeņu attīrīšanas iekārtas ieplūdē (fotofiskācijas iegūtas projekta īstenošanas laikā). Iekārta ievāc 7,2 litru kompozītparaugu, kas tiek ievākts ar paraugu ņemšanas intesitāti 300 ml 1 reizi stundā (parauga ievākšana notiek 24 stundu laikā, veidojoties 1 attiecīgo laika periodu raksturojošam notekūdeņu paraugam)

SARS-CoV-2 vīrusa klātbūtnes monitorings/ testēšana notekūdeņos

Pēc paraugu nonākšanas laboratorijā jāveic paraugu sagatavošana vīrusu RNS un mikrobioloģiskā DNS izdalīšanai. Tās ietvaros notekūdeņus saskalina un 400 µl no kopējā parauga ievāc priekš mikroorganismu DNS izdalīšanas, kā arī 180 ml no kopējā parauga ievāc priekš vīrusu RNS izdalīšanas.

DNS izdalīšana jāveic, izmantojot piemērotus augsnes mikrobiālā DNS izdalīšanas reaģentu un materiālu komplektus. Šo komplektu iegādes izvēlei ir jābūt balstītai uz diviem faktoriem: kopējā DNS ieguvuma, kur lielāks iegūtās DNS daudzums ir uzskatāms

par labāku, un paraugu dezintegrācija, kur metodes, kas vienlīdz labi sagrauj gan gram+, gan gram- baktērijas ir uzskatāmas par piemērotākām. Pēc DNS iegūšanas jāveic iegūtā materiāla kvalitātes un daudzuma novērtēšana, izmantojot agarozes gela elektroforēzes un interkolējošās fluorescentās krāsas kvantifikācijas metodes. Ja iegūtie rezultāti ir apmierinoši, tiek veikta visas DNS lielapjoma paralēlās sekvencēšanas bibliotēkas sagatavošana un tās sekvencēšana ar minimālo datu apjomu, kas līdzvērtīgs 23 miljoniem nolasījumu uz paraugu. Pēc jēldatu ieguves tie tiek analizēti, izmantojot uz attiecīgo brīdi modernākās datu analīzes programmas un datubāzes, ar mērķi noteikt mikroorganismu kopienas struktūru un novērtēt paraugā esošos patogenus.

Eksperimentālās WBE DNS metožu izstrādes ietvaros veicot darbības saskaņā ar ražotāja norādījumiem notika pielietoti MP Biomedicals (ASV) ražotie "FastDNA™ Spin Kit for Soil DNA Extraction", Qubit™ 1X dsDNA HS Assay Kit, MGIEasy Universal DNA Library Prep Set un DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE150).

Lai novērtētu vīrusu RNS izdalīšanas efektivitāti, pirms tālāko procesu uzsākšanas paraugam ir jāpievieno piemērots daudzums surogātvīrusa daļiņu. Izvēlētajam surogātvīrusam ir jābūt līdzīga izmēra un uzbūves RNS vīrusam, kas dabīgi nav sastopams vidē, no kura nākuši paraugi. Izdalīšanas procedūru uzsāk ar lielākā izmēra daļiņu frakcijas izgulsnēšanu, ko veic centrifugējot 30 min pie 8000 x g. Vīrusus no iegūtā supernantanta iegūst pievienojot 24 g polietilēna glikola MW8000 un 5.4 g nātrija hlorīda, inkubējot 2 h istabas temperatūrā un izgulsnējot ar 10 min centrifugēšanu pie 12000 x g. RNS no vīrusiem jāizdala, izmantojot piemērotus vīrusa RNS izdalīšanas reaģentu un materiālu komplektus. Pie šo komplektu iegādes izvēlei ir jābūt balstītai uz diviem faktoriem: kopējā RNS ieguvuma, kur lielāks iegūtās nukleīnskābes daudzums ir uzskatāms par labāku, un spējas izdalīt gan liela, gan neliela izmēra RNS molekulas, jo notekūdeņu paraugos vīrusa genoma RNS var būt fragmentēta. Pēc RNS iegūšanas jāveic iegūtā materiāla daudzuma novērtēšana, izmantojot fluorescentās krāsas kvantifikācijas metodes. Ja iegūtais daudzums ir pietiekams, tad materiāls ir uzskatāms par piemērotu tālākām darbībām.

Eksperimentālās WBE RNS izdalīšanas metožu izstrādes ietvaros, veicot darbības saskaņā ar ražotāja norādījumiem, tika pielietotas rekombinantas Semliki forest vīrusam līdzīgās daļiņas, Trizol un Qubit™ RNA HS Assay Kit.

Specifiskā vīrusa daudzuma noteikšanai vispiemērotākā ir jaunākās paaudzes kPĶR metode, jeb digitālais PĶR (dPĶR), jo atšķirībā no klasiskās kPĶR metodes, dPĶR sniedz iespēju noteikt izvēlēto vīrusa gēnu daudzumu līdz precīzam vienību skaitam vienā notekūdeņu litrā. Tomēr, tā kā vīrusa RNS var būt fragmentēta, tad ir ieteicams veikt mērījumus vairākiem vīrusa genoma rajoniem, kā arī surogātvīrusa genoma rajonam, lai novērtētu izdalīšanas procesa efektivitāti. Kvantifikācija jāveic izmantojot atbilstošas dPĶR iekārtas un tām piemērotus reaģentus un reaģentu un materiālu komplektus. Pie šo reaģentu un komplektu iegādes, izvēlei ir jābūt balstītai uz sekojošiem apsvērumiem: izmantojot ar dažādām fluorescentām krāsvielām iezīmētas TaqMan tehnoloģijas zondes, mērījumi ir specifiskāki, un pastāv iespēja dažādus mērījumus multipleksēt, kā arī vēlams, lai izmantotie reaģentu komplekti vienas reakcijas ietvaros veiktu gan reverso transkripciju, gan amplifikāciju. Pēc reakcijas maisījumā esošo vīrusa genoma daļu kopiju skaita noteikšanas ir jāveic atbilstoši aprēķini, lai izskaitļotu attiecīgā vīrusa daudzumu vienā notekūdeņu litrā un iegūtais rezultāts ir jāsalīdzina ar iepriekšējos periodos iegūto. Ja ir novērojams būtisks skaitliskā rādījuma pieaugums, tad uzskatāms, ka saslimušo skaits attiecīgajā teritorijā ir pieaudzis, ja kritums - tad samazinājies. Izdalīšanas efektivitāti aprēķina, kā procentus no atgūtā surogātvīrusa.

Eksperimentālās WBE vīrusa kvantifikāijas metožu izstrādes ietvaros, veicot darbības saskaņā ar ražotāja norādījumiem, tika pielietotas BioRad QX200 dPĶR sistēma, SARS-CoV-2 vīrusa N1, N2 un E genoma rajoniem specifiskas zondes, rekombinantās Semliki forest vīrusam līdzīgo daļi ņu GFP genoma rajonam specifiska zonde un One-Step RT-ddPCR Advanced Kit for Probes.

Lai novērtētu dažādu vīrusa variantu klātbūtni notekūdeņos, tiek veikta visa vīrusa genomiskās RNS sekvencēšana. Tās ietvaros lielapjoma paralēlās sekvencēšanas bibliotēkas sagatavo un tās sekvencē ar minimālo datu apjomu, kas līdzvērtīgs 23 miljoniem nolasījumu uz paraugu. Pēc jēldatu ieguves tie tiek analizēti, izmantojot uz attiecīgo brīdi modernākās datu analīzes programmas. Attiecīgā vīrusa subtipa klātbūtne ir uzskatāma par pierādītu, ja tā atrodama vismaz 1 % no vīrusa atiecīgām genoma rajonam piederīgo nolasījumu. Iegūtais rezultāts ir jāsalīdzina ar iepriekšējos periodos iegūto. Ja ir novērojams būtisks attiecīgā subtipa proporcijas rādījuma pieaugums, tad uzskatāms, ka ar šo subtipu inficēto skaits attiecīgajā teritorijā ir pieaudzis, ja kritums - tad samazinājies.

Eksperimentālās WBE vīrusa sekvencēšanas metožu izstrādes ietvaros, veicot darbības saskaņā ar ražotāja norādījumiem, tika pielietoti ATOPlex RNA Library Prep Set un DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE100).

SARS-CoV-2 vīrusa klātbūtnes monitorings notekūdeņos 2021.- 2022. gadā Latvijā

Paraugu ievākšanas vietas

Ņemot vērā 2021. gada epidemioloģisko situāciju Latvijā attiecībā uz Covid-19 izplatību, SARS-CoV-2 monitoringa notekūdeņu sistēmās uzdevumi būtu:

1) pēc SPKC rekomendācijām un epidemioloģiskām indikācijām izvērtēt prioritārās paraugu ņemšanas vietas Latvijā;
2) novērtēt un iespēju robežās prognozēt iespējamu jaunu SARS-CoV-2 saslimšanas gadījumu skaita pieaugumu attiecīgajās pašvaldībās vai pilsētās;
3) sniegt papildu informāciju par iespējamu jaunu SARS-CoV-2 vīrusa variantu izplatību konkrētās pašvaldībās;
4) izveidot ilgtspējīgu notekūdeņu monitoringa sistēmu Latvijā turpmāko epidēmiju efektīvākai uzraudzībai un kontrolei.

Balsoties uz līdz šim veiktajos pētījumos iegūtajiem rezultātiem un starptautisku pieredzi, notekūdeņu paraugus varētu paredzēt ievākt vismaz 10 Latvijas pilsētu notekūdeņu attīrīšanas iekārtās (kuras pēc epidemioloģiskām indikācijām varētu arī mainīt), kur nonāk notekūdeņi no lielākās daļas attiecīgās apdzīvotās vietas iedzīvotāju notekūdeņiem. Tādejādi iespējams monitorēt maksimāli liela iedzīvotāju skaita veselības stāvokli. Sākotnēji rekomendējamās pašvaldības, kur uzsākt notekūdeņu epidemioloģisko monitoringu un lielākoties inženiertehniskas piezīmes par to izvēles pamatojumu apkopotas 1.tabulā. Vēršam atkārtotu uzmanību, ka pašvaldību, kurās veikt monitoringu, identifikācija jāveic balsoties gan uz inženiertehniskiem ierobežojumiem vai iespējām, gan epidemioloģisko situāciju, līdz ar to kvalitatīvas monitoringa sistēmas ieviešanas un faktiskas pielietošanas pamatā ir cieša notekūdeņu sistēmu inženieru un pētnieku un epidemiologu sadarbība.

1.tabula

Notekūdeņu epidemioloģijas monitoringa programmā iespējami iekļaujamās pilsētas

Pilsēta/ pašvaldība	Piezīmes
Rīga	liela pilsēta, daudz saslimušo, neskaidra vakcinācijas aktivitāte, daudz nenoteiktības, notekūdeņi no Rīgas, Jūrmalas, Garkalnes, Babītes, Mārupes
Jelgava	tiek turpināts esošais monitorings, liela pilsēta, salīdzinoši vienkārša sistēma
Kuldīga	tiek turpināts esošais monitorings, salīdzinoši maza pilsēta, salīdzinoši vienkārša sistēma
Daugavpils	liela pilsēta, neliela vakcinācijas aktivitāte, daudz saslimšanas gadījumu
Rēzekne	liela pilsēta, neliela vakcinācijas aktivitāte, daudz saslimšanas gadījumu
Liepāja	liela pilsēta, aktīva vakcinācija, vidēji daudz saslimšanas gadījumu
Valmiera	liela pilsēta, aktīva vakcinācija, vidēji daudz saslimšanas gadījumu
Sigulda	vidēja pilsēta, daudz iedzīvotāju strādā Rīgā – mobilitāte, tūrisma gala mērķis (tuvojas pavasaris un aktīvs tūrisms)
Ogre	vidēja pilsēta, daudz iedzīvotāju strādā Rīgā – mobilitāte
Saulkrasti	maza pilsēta, milzīga mobilitāte un iedzīvotāju skaita pieaugums vasarā
Citas pilsētas	Jūrmala (liela pilsēta, sarežģīta sistēma, daļa no notekūdeņiem 30% tiek pārsūknēti uz Rīgu). Ventspils, Cēsis – kā papildus alternatīvas.

Rezultātu apstrāde un interpretācija

SARS-CoV-2 klātbūtnes noteikšanas testa rezultāts tiek izteikts kā RNS kopijas ml notekūdens parauga un pārrēķināts uz 100 000 iedzīvotāju. To nosaka, izmantojot papildu datus, kas iegūti no informācijas par notekūdens sistēmas uzbūvi (piem., vai ir kopēja ar lietus kanalizāciju, tīklu garums), hidroloģiskiem un klimata datiem (nokrišņu daudzums, temperatūra) kā arī biomarķieru mērījumiem. Mērījumu precizitāte pieaug ar mērījumu skaitu.

Notekūdeņu visa metagenoma analīzes rezultāts ir saraksts ar notekūdeņos identificētajiem mikroorganismiem un to proporcionālo sadalījumu. Veicot saraksta padziļinātu izpēti, iespējams identificēt patogeno mikroorganismu sugas, novērtēt to avotu un iedzīvotāju veselības stāvokli attiecīgās saslimšanas kontekstā.

SARS-CoV-2 genoma analīzes rezultāts ir saraksts ar identificētajām paraugā esošā vīrusa genoma mutācijām. veicot saraksta padziļinātu izpēti iespējams novērtēt jau zināmo “bīstamo” subtipu izplatību un savlaicīgi identificēt jaunu – potenciāli bīstamu subtipu rašanos.

Rekomendāciju kopsavilkums

Pirmie notekūdeņu monitoringa izmantošanas rezultāti Latvijā tika prezentēti Eiropas Savienības Pētniecības institūta (JRC) organizētā konferencē 2020.gada jūnijā. Sākoties VPP programmai, tika ievākti paraugi un nogādāti uz JRC metodikas pilnveidošanai. Regulāri notika kopējās sapulces, kurās par veiksmīgu WBE izmantošanu EU ziņoja daudz dalībvalstu zinātnieki. Ar notekūdeņu monitoringu ir iespējams savlaicīgi noteikt dažādu infekcijas slimību uzliesmojumu sākšanos un apjomu apdzīvotās vietās un identificēt apdzīvoto vietu reģionus ar paaugstinātu saslimušo skaitu. Izmantojot esošo vides paraugu ievākšanas un loģistikas tīklu, to iespējams samērā īsā laikā un ar salīdzinoši nelieliem papildus izdevumiem ieviest arī praksē jau no 2021.gada.

1) Notekūdeņu monitoringam (WBE) ir būtiska nozīme vīrusu epidemioloģijā tās izmantošana var veicināt savlaicīgāku un pamatotāku lēmumu pieņemšanu Covid-19 (un ne tikai) uzraudzībā un kontrolē.
2) WBE izmantošana nodrošina epidemioloģiska monitoringa veikšanu, nesaskaroties ar saslimušajiem, līdz ar to neapdraudot pašu monitoringa veikšanā iesaistīto kritiski svarīgu speciālistu veselību.
3) WBE pieejas izmantošanai Latvijā kritiski nozīmīga ir pašvaldību un ūdenssaimniecību (centralizēto notekūdeņu savākšanas un attīrīšanas sistēmu apsaimniekotāju), Slimību Profilakses un Kontroles centra (SPKC), Pārtikas un veterinārā dienesta (PVD) un pētniecības organizāciju sadarbības nodrošināšana paraugu ievākšanā, izmeklējumu veikšanā un rezultātu interpretācijā.
4) Ilgtermiņā WBE var tikt izmantota ne tikai Covid-19, bet arī citu slimību monitoringā vai sabiedrības veselības pētījumos.
5) Pielietojot WBE var būtiski samazināt inficēto cilvēku skaitu un monitoringa izmaksas, jo šādi iegūtā informācija sniedz iespēju savlaicīgi identificēt uzliesmojumus, operatīvi informēt attiecīgā reģiona/rajona sabiedrību un veselības aprūpes speciālistus un fokusēt (citādi visas valsts mēroga) izdevumus problemātisko teritoriju iedzīvotāju masveida diagnostikai.

REFERENCES

1) Barcelo D. 2020. Wastewater-based epidemiology to monitor Covid-19 outbreak: Present and future diagnostic methods to be in your radar. *Case Studies in Chemical and Environmental Engineering.* 2 (100042).
2) Daughton C.G. 2020. Wastewater surveillance for population-wide COVID-19: The present anf future. *Science of the Total Environment.* 736 (139631).
3) Davo L., Segui R. Et al. 2021. Early detection of SARS-CoV-2 infection cases or outbreaks at nursing homes by targeted wastewater tracking. *Clinical Microbiology and Infection.* https://doi.org/10.1016/j.cmi.2021.02.003.
4) Pērkons I., Ruško J., Začs D., Bartkevičs V. 2021. Rapid determination of pharmaceuticals in wastewater by direct infusion HRMS using target and suspect screening analysis. *Science of the Total Environment.* 755 (142688). - **VPP_COVID-19 projekta ietvarā izstrādātā publikācija**.
5) Polo D., Quintela-Baluja M., et al. 2020. Making waves: Wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. *Water Research.* 186 (116404).
6) Hart O.E., Halden R.U. 2020. Computational analysis of SARS-CoV-2/ COVID-19 surveillance by wastewater-based epidemiologylocally and globally: Feasibility, economy, opportunities and challenges. *Science of the Total Environment.* 730 (138875).

PIELIKUMI

Water Research

Detection of SARS-CoV-2 in wastewater and importance of population size assessment in smaller cities: an exploratory case study from Latvia

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	SARS-CoV-2; WBE; wastewater; ddPCR; 5-HIAA; mobile phone data.
Corresponding Author:	Davids Fridmanis Latvian Biomedical Research and Study Centre: Latvijas Biomedicinas petijumu un studiju centrs LATVIA
First Author:	Dita Gudra, M.Sc.
Order of Authors:	Dita Gudra, M.Sc.
	Sandis Dejus
	Vadims Bartkevics
	Ance Roga
	Martins Strods
	Anton Rayan
	Kristina Kokina
	Anna Zajakina
	Uga Dumpis
	Laura Elina Ikkere
	Irina Arhipova
	Gundars Berzins
	Aldis Erglis
	Juris Binde
	Evija Ansonska
	Aivars Berzins
	Talis Juhna
	Davids Fridmanis
Abstract:	Wastewater-based epidemiology (WBE) has regained global importance during the COVID-19 pandemic. Mobility of people and other factors such as precipitation are lowering the accuracy of the disease prevalence estimates, which is crucial for proper crisis management. These estimations are particularly challenging in urban areas with moderate or low numbers of inhabitants. Here, we demonstrate that by combining data of detected SARS-CoV-2 RNA copy number, 5-Hydroxyindoleacetic acid analyses in wastewater and mobile call detail record it was possible to provide accurate assessment of COVID-19 epidemiological situation in small and medium-sized towns. This is the first study demonstrating WBE for monitoring COVID-19 outbreaks in Latvia and showing that the application of such population size estimation measurements as total 5-HIAA and call detail record data improve the accuracy of the WBE approach.
Suggested Reviewers:	Helene Norder helene.norder@gu.se
	Gloria Sanchez gloriasancez@iata.csic.es

	Raul Gonzalez rgonzalez@hrsd.com
	Daniel Gerrity daniel.gerity@snwa.com

Detection of SARS-CoV-2 in wastewater and importance of population size assessment in smaller cities: an exploratory case study from Latvia

Authors:

Dita Gudra[1]*, Sandis Dejus [2]*, Vadims Bartkevics[3]*, Ance Roga[1], Martins Strods[2], Anton Rayan[2], Kristina Kokina[2], Anna Zajakina[1], Uga Dumpis[5], Laura Elina Ikkere[3], Irina Arhipova[4], Gundars Berzins[5], Aldis Erglis[5], Juris Binde[6], Evija Ansonska[5], Aivars Berzins[3]**, Talis Juhna[2]**, Davids Fridmanis[1]**

Affiliations:

1 - Latvian Biomedical Research and Study Centre, Ratsupites iela 1, Riga, LV-1067, Latvia

2 - Riga Technical University, Laboratory of Water Research and Environmental Biotechnology, Kipsalas iela 6a/6b, Riga, LV-1048, Latvia

3 - Institute of Food Safety, Animal Health and Environment BIOR, Lejupes iela 3, Riga, LV-1067, Latvia

4- Latvia University of Life Sciences and Technologies, Liela iela 2, Jelgava, LV-3001, Latvia

5- University of Latvia, Aspazijas bulvaris 5, Riga, LV-1050, Latvia

6- LLC "Latvian Mobile Telephone", Ropazu iela 6, Riga, LV-1039, Latvia

* These authors contributed equally

** These authors contributed equally

Corresponding authors:

Sample collection from WWTP: talis.juhna@rtu.lv

Physicochemical analysis of wastewater: vadims.bartkevics@bior.lv

Genetic analysis of wastewater: davids@biomed.lu.lv

ABSTRACT

Wastewater-based epidemiology (WBE) has regained global importance during the COVID-19 pandemic. Mobility of people and other factors such as precipitation are lowering the accuracy of the disease prevalence estimates, which is crucial for proper crisis management. These estimations are particularly challenging in urban areas with moderate or low numbers of inhabitants. Here, we demonstrate that by combining data of detected SARS-CoV-2 RNA copy number, 5-Hydroxyindoleacetic acid analyses in wastewater and mobile call detail record it was possible to provide accurate assessment of COVID-19 epidemiological situation in small and medium-sized towns. This is the first study demonstrating WBE for monitoring COVID-19 outbreaks in Latvia and showing that the application of such population size estimation measurements as total 5-HIAA and call detail record data improve the accuracy of the WBE approach.

Keywords: SARS-CoV-2, WBE, wastewater, ddPCR, 5-HIAA, mobile phone data.

Abbreviations

ddPCR - Droplet PCR

WW - Wastewater

WWTP - Wastewater treatment plant

PEG - Polyethylene glycol

WBE - Wastewater based epidemiology

CDR - call detail record

SD – standard deviation

1 Introduction

The COVID-19 pandemic caused by the SARS-CoV-2 virus reached Latvia in March 2020, with an average of 7.4 new cases daily and a mean 14-day cumulative incidence of 3.1±4.5 until late September 2020. During this period several outbreaks of SARS-CoV-2 were detected and successfully contained due to extensive testing and contact tracing led by the Disease Prevention and Control Centre. However, the incidence increased rapidly starting from October, reaching its maximum on January 10th, 2021, with a 14-day cumulative incidence of 693.9. Since then the number of COVID-19 cases has remained consistently high, with only periodical small-scale fluctuations [1], which was partially associated with the spread of emerging SARS-CoV-2 variants. At the moment of manuscript preparation, 133'098 confirmed infection cases and 2'370 deaths have been reported in Latvia [2].

The conventional approach of monitoring the spread of SARS-CoV-2 virus in the communities mainly relies on testing of symptomatic patients. This approach is time-consuming and is often limited by testing capacity in a particular region. In addition, people with mild or asymptomatic disease might not seek diagnostic tests thus contributing to the silent spread of the virus. In addition to conventional diagnostic tests, wastewater (WW) based SARS-CoV-2 tracking has recently gained wide popularity in the world. Although SARS-CoV-2 is primarily a respiratory virus, recent studies indicate that it might be excreted in feces and urine from presymptomatic, asymptomatic and recently recovered patients [3–7]. There is even a suggestion that SARS-CoV-2 might replicate within the gastrointestinal tract [8,9], however, not all confirmed individual stool samples have been tested positive for SARS-CoV-2 [5,10]. The duration of viral shedding in feces has been observed to vary among patients with an average of 25-28 days [10–13], and with a magnitude of 10^2 up to 10^8 RNA copies per gram [6,7,10]. Recently several studies have analyzed continuous data of SARS-CoV-2 presence in the WW from different countries - France [14], USA [15,16], Qatar [17], Finland [18], Spain [19], Germany [20], UK [21], Sweden [22] and other, - concluding that tracing viruses present in raw WW have proved to be a quick and timely method to track the dynamics of the infection within a community [23,24]. Nevertheless, estimation of the approximate number of infected individuals remains a worldwide challenge due to a number of reasons. Such reasons include, but are not limited to: diverse viral shedding rates in a population, high dilution of viral RNA and their concentration fluctuations in the WW by rain and industrial WWs, presence of compounds or particles that may either degrade or physically protect the virus, and, loss of viral RNA during the transit time through the WW network [21,25]. While some of the above-mentioned reasons might have a low impact on the large cities for the detection and quantification of the viral RNA in the WWs, smaller size populations are affected to a much greater extent.

To address this problem several population biomarkers such as 5-HIAA (5-Hydroxyindoleacetic Acid) [26], coffee or artificial sweeteners were used [27,28]. However, they are affected by degradation and dilution in WW streams and some of them also by particular product consumption habits within the local population. Thus the accuracy of these population size measurements should be verified at specific locations and for that the use of

mobile data as a real-time data source has become increasingly relevant [29], which allows, for example, to assess the trends of regional economic development or to determine the change in population mobility patterns.

The objectives of this study were (i) to demonstrate the applicability of WBE in small and medium-sized municipalities of Latvia and (ii) to test whether combination of data from biomarkers (5-HIAA) and mobile call detail record (CDR) is able to increase the accuracy of relationship assessment between the amount of detected SARS-CoV-2 RNA copies in WW and confirmed SARS-CoV-2 cases.

2 Methods

2.1 Sample collection

Two municipalities of Latvia were selected for continuous monitoring of SARS-CoV-2 presence in the WWs - Jelgava (~55'000 inhabitants) and Kuldiga (~11'000 inhabitants). Composite 23-hour raw WW samples (7.2 L) were collected at the WWTP (after grit removal) using a portable autosampler P6 MINI MAXX (MAXX Mess und Probenahmetechnik GmbH, Germany), which was operating at time dependant mode (300 mL/h). Collected samples were immediately transferred to the laboratory, stored at 4 ^{o}C and further processed within 24 hours. Samples at each municipality were collected once or twice per week. Samples from Jelgava WWTP were collected starting from 17th of August 2020 till 30th of November 2020, resulting in the acquisition of 16 samples; while samples from Kuldiga WWTP - starting from 7th of October 2020 till 29th of November 2020, resulting in a collection of 17 samples (Supplementary Table 1). Assumptions behind the selection of municipalities that were included in monitoring and collection schedules are explained in the discussion section.

2.2 WW treatment and RNA extraction

180mL (4x45ml) of WW sample was centrifuged at 8000xg for 30min at 4 °C to remove larger particles, such as bacterial cells and debris. The supernatant was transferred to new tubes and virus particles were precipitated using Polyethylene glycol (PEG) 8000 as described elsewhere [30]. Briefly, 24g of PEG 8000 (8% w/v, Sigma-Aldrich) and 5.4g of NaCl (PanReac AppliChem, Germany) were added to 180ml of supernatant. The mixtures were incubated for 2h at RT with gentle agitation. The precipitated virus particles were recovered by centrifugation at 12000xg for 10min. Total RNA was isolated with Tri reagent (Sigma-Aldrich, USA) according to manufacturer's instructions and eluted in 30µl molecular grade water. RNA samples were stored at - 80 °C and subjected to RT-qPCR analysis within 24h after the RNA extraction. RNA concentration was estimated using the Qubit RNA HS Assay kit and Qubit 2.0 fluorometer (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. To assess the efficiency of the extraction procedure and estimate the viral recovery rate, each sample was supplemented with 5µl of surrogate - recombinant, replication defective and GFP gene containing Semliki Forest Virus (SFV) particles. The concentration of the surrogate was $2{,}27 \times 10^{7}$ of infective particles per 1ml as estimated by determination of infectious particle titter while the number of GFP copies per 1µl of surrogate was 100'700 as determined by Droplet Digital PCR (ddPCR). The procedure of copy number determination is described in subsequent sections.

2.3 ddPCR

ddPCR approach was employed for ultra-precise quantification of SARS-CoV-2 RNA copies within WW samples and assessment of surrogate recovery rate. For that purpose, analyses were carried out on two regions of SARS-CoV-2 nucleocapsid (N) gene (N1 and N2), single region of SARS-CoV-2 Envelope small membrane protein (E)

gene and single region of surrogate - recombinant SFV GFP gene (Table 1). All reactions were carried out in a single plex employing One-Step RT-ddPCR Advanced Kit for Probes (BioRad, Germany). The reaction mixture contained 5µL of Supermix, 2µL of reverse transcriptase, 1µL of 300mM dithiothreitol, appropriate primers (forward and reverse) and probes (Metabion, Germany) to final concentration of 1.21µM each, 2µL of extracted WW RNA and RNAse free water to total volume of 22µL. The following steps included droplet generation with QX200 Droplet Generator (Bio-Rad, Germany), amplification in T100 Thermal Cycler (Bio-Rad, Germany) (under following conditions: Ramp Rate setting 1; 50ºC for 60min; 95ºC for 60min; 40 cycles of 94ºC for 30sec and 60ºC for 2min; 98ºC for 10min), 4h equilibration and droplet stabilization in room temperature and positive/negative droplet quantification in QX200 Droplet Reader (Bio-Rad, Germany). Acquired data were analysed employing QuantaSoft software (Bio-Rad, Germany) and calculated to the number of copies per mL of WW. The average from three SARS-CoV-2 analyses was used in all further viral concentration assessments. The recovery of surrogate SFV was calculated as the percentage of recovered GFP RNA coding copies.

Table 1. Names and sequences of employed oligonucleotides. All were synthesized by Metabion GmbH (Germany)

Oligonucleotide name	Oligonucleotide sequence	Source
nCoV_N1 Forward	GAC CCC AAA ATC AGC GAA AT	Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Center for Disease Control and Prevention, Atlanta, GA, USA
nCoV_N1 Probe	(6-Fam) ACC CCG CAT TAC GTT TGG TGG ACC (BHQ-1)	
nCoV_N1 Reverse	TCT GGT TAC TGC CAG TTG AAT CTG	
nCoV_N2 Forward	TTA CAA ACA TTG GCC GCA AA	
nCoV_N2 Probe	(6-Fam) ACA ATT TGC CCC CAG CGC TTC AG (BHQ-1)	
nCoV_N2 Reverse	GCG CGA CAT TCC GAA GAA	
nCoV_E Forward	ACA GGT ACG TTA ATA GTT AAT AGC GT	Corman *et al.* [31]
nCoV_E Probe	(6-Fam) ACA CTA GCC ATC CTT ACT GCG CTT CG (BHQ-1)	
nCoV_E Reverse	ATA TTG CAG CAG TAC GCA CAC A	
GFP Forward	CTG CTG CCC GAC AAC CAC	Wang *et al.* [32]
GFP Probe	(Hex) CCA GTC CGC CCT GAG CAA AGA CC (BHQ-1)	
GFP Reverse	TCA CGA ACT CCA GCA GGA C	

2.4 5-Hydroxyindoleacetic acid (5-HIAA) measurements within WW

Solid-phase extraction was performed on Strata-X cartridges (200 mg/3 mL). Cartridges were conditioned with 3 mL of methanol and 3 mL of deionized water. The samples were filtered through glass fiber filters using a vacuum. After acidification with 100 µL of formic acid, the samples were loaded on columns at the approximate flow rate of 5 mL min^{-1}. Columns were rinsed with 3 mL of 5% methanol solution in water and then dried for 60 min under vacuum, followed by elution with 2 x 3 mL of methanol. The eluates were then evaporated to dryness under a gentle nitrogen stream in a water bath at 40ºC temperature. The samples were reconstituted in 200 µL of water/methanol (80/20, v/v) containing the internal standard of salbutamol-d3 at a concentration of 0.05 ng μL^{-1}.

Extracts were filtered through PVDF 0.22 µm centrifugal filters. For the analysis of 5-HIAA final extracts were diluted ten-fold with water/methanol (80/20, v/v).

Instrumental analysis was performed on the UHPLC-Orbitrap-HRMS system. Chromatographic separation was achieved on the Hypersil Gold C18 analytical column (100 × 2.1 mm, 1.9 µm) column using mobile phases of 0.01 % acetic acid in water (A) and acetonitrile (B) at flow rate of 0.3 ml min^{-1}. A gradient program was used: 5% of mobile phase (B) was used from 0 to 1.0 min, 5% (B) to 95% (B) from 1.0 to 5.0 min, maintained at 95% (B) from 5.0 to 9.0 min, then decreased back to 5% (B) from 9.0 to 10.0 min and finally, the column was re-equilibrated with 5% (B) from 10.0 to 15 min. A 10 µL aliquot of the extract was injected. The column and autosampler were maintained at 30°C and 15°C, respectively.
Orbitrap-HRMS was equipped with a heated electrospray ionization probe (HESI II) operating in the positive ionization mode. 5-HIAA was detected in full-scan mode by measuring two characteristic ions – 192.0655 *m/z* (quantitative) and 146.0606 *m/z* (qualitative) using mass tolerance of 10 ppm. Quantification was performed by external calibration using a five-point calibration curve obtained by spiking pooled WW samples at the concentration range of 2.5 to 50 µg L^{-1}.

2.5 Estimation of population size through employment of mobile activity data (Call detail record)

The unique mobile phone user data set consisted of the aggregated Call Detail Records (CDR) collected during the time period from October 2020 to November 2020 from all base stations in Kuldiga municipality cellular network (outgoing or incoming call, SMS). The data was aggregated within the boundaries of 24 hours of a specific day and thus it can be overlapped with the number of mobile phone users in a previous day.

2.6 Data visualization and statistics

Population normalization for viral load in WW was performed by application of relative population markers as 5-HIAA. Normalization regarding 5-HIAA was done by dividing the total copies of detected viral RNA with the total 24h 5-HIAA amount. Dataset of confirmed SARS-CoV-2 cases per municipality was acquired from the Centre for Disease and Prevention and Control and can be downloaded from https://data.gov.lv/dati/lv/dataset/covid-19-pa-adm-terit. Cumulative incidence for 14, 21, 28 and 35 days per each municipality was calculated and correlation analysis of total viral copies in millions and millions of copies per mg of 5-HIAA versus cumulative incidence of respective periods was carried out with the following equation:

$$Correlation\ coeff.(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

Two-tailed Mann-Whitney test [33] was used to compare (i) 5-HIAA measurements between different sampling days of the week (ii) correlation coefficients between total viral copies, 5-HIAA normalized viral copies and cumulative incidence, (iii) normalized viral load measurements between different sampling days of the week within a municipality, and (iiii) to compare average number of phone calls between weekends and weekdays. Pearson correlation analysis was performed to evaluate the association between the total amount of 5-HIAA and average number of phone calls in Kuldiga.

Cluster analysis was used to group similar day observations into a number of clusters based on the observed values of daily aggregated mobile phone user's data and 5-HIAA data. The cluster membership was compared with the distribution of weekdays and weekends.

All figures were created using matplotlib [34] and seaborn [35] libraries within the python environment.

3 Results

3.1 Physicochemical analysis of WW samples

Following the sample collection, a number of physical and chemical measurements were carried out to account for factors that could exert a substantial impact on the stability of SARS-CoV-2 RNA and fluctuations in population size - pH, temperature, electrical conductivity and 5-HIAA concentration measurements. In addition, to estimate the total volume of WW that flowed through the site during sample collection time we also copied the appropriate water meter reading records. However, because of access restrictions, the meter reading times did not match those of sample collection start and end time, therefore for this study the total volume of municipality WW that flowed into the WWTP during sample collection was estimated through calculation of average WW volume per hour during the time period between two-meter records and multiplication by the amount of time the sample collection was carried out. If the sample collection spanned several meter record time periods, then calculation was carried out for each period individually and subsequently summed.

The analysis of parameters that were measured on-site revealed that to some extent they were municipality-specific, but otherwise rather stable over the whole study time (Supplementary Table 1). Thus, the average WW pH value in Jelgava was 7.77 (Standard deviation (SD)=0.14) and the average electrical conductivity was 2015.38 (SD=99.75) µS/cm, while in Kuldiga they were 7.67 (SD=0.15) and 1215.11 (SD=111.70) µS/cm, respectively. Greater variations were observed in WW temperature, which in Jelgava on average was 15.01 (SD=4.64) °C while in Kuldiga 11.78 (SD=1.65) °C, but these variations correlated with the temperature of the surrounding environment and were appropriate for Latvian seasonal changes.

Since it is a well-known fact that industrial WW, as well as rainwater, can significantly affect the concentration of specific analytes through dilution effect, we decided to work with total amount values and not the concentration values. In the case of 5-HIAA, this strategy is also supported by "ENETS Consensus Guidelines" where 24h total 5-HIAA measurements are recommended for diagnostics of Neuroendocrine Tumours [36] and the normal secretion range of this metabolite is 2 to 9 mg/24h per person [37]. Thus, the multiplication of results from 5-HIAA concentration measurements with the total amount of WW from specific collection events revealed that there is considerable day-to-day variability in the amount of this metabolite (Supplementary Table 1). We observed that on average inhabitants of Jelgava produced 103'080.70mg of 5-HIAA per day, the SD was 14'489.43mg, which corresponds to 14.06% of deviation from the average, while the relative range was 52.34% points of the average. Although the cause of these fluctuations has not been investigated in this study, we believe that it might be due to combinatorial effects of such factors as metabolite degradation rate in WW, metabolite secretion rate, and most prominently population movement. In comparison to Jelgava, the average of the total 5-HIAA amount in WW of Kuldiga was 6.02 times lower (17'136.14mg), which fits well with the differences in population size of both locations, but unlike in the former, here we also observed greater SD 3'686.6 mg (21.51%) and greater relative range of the measurements (82.54% points of the average). Due to the fluctuations of total 5-HIAA amount, we divided datasets of both populations according to the day of the week of sample collection start date (Mon, Tue and Thu for Jelgava and Mon, Tue, Thu and Sun for Kuldiga) and in the case of those that contained a sufficient number of data points (at least three) all were compared in pairs using two-tailed Mann-Whitney test. Acquired results revealed that the differences between Mon and Thu datasets from Jelgava were not significant (p=0.3710) while differences between Thu and Sun datasets from Kuldiga were significant

(*p*=0.0080). Considering that daytime activities of the general population during weekdays differ from those of weekends, we decided to perform a comparison of data that were grouped according to this division. Acquired results revealed that both data sets differed significantly (*p*=0.0237). The average of the total 5-HIAA amount in WW collected on weekdays was 18’751.69 mg (SD=2’987.32 mg or 15.93%), while in WW that was collected on weekends it was 14’174.30 mg (SD=3061.76 mg or 21.60%), but the relative range of measurements was 54.91% points and 48.31% points of the average respectively, which in both cases is closer to that of Jelgava. Since the employment of 5-HIAA as population size marker is considered experimental while the uncovered difference was indeed significant, we decided that an additional verification of this observation employing a WW unrelated approach is necessary.

3.2 Validation of observed differences between weekday and weekend total 5-HIAA measurements by CDR

Employment of mobile phones has become prevalent in the predominant majority of European countries and since these devices are always located in close proximity of their owners their territorial density reflects the density of population within a specific area. Therefore, we speculated that if observed differences in total amount of 5-HIAA in Kuldiga between weekdays and weekends were true then these should also be reflected in mobile phone usage data, and we performed the evaluation of mobile phone call activities in conjunction with 5-HIAA measurements.

As explained in materials and methods section the accrued mobile phone usage data were the 24h Call Detail Record (CDR) aggregates and the aggregation start and end time for particular day was midnight, which did not match the sample collection start and end time. Therefore, the assessment of mobile phone activity for the period of sample collection was calculated in a similar manner as previously the volume of WW that passed the collection site during sample collection. Thus, the average activity per hour was calculated for each day and multiplied by the amount of time the sample collection was carried out during that day. The total mobile phone usage activity during sample collection time was acquired by summing the respective results from two days (Supplementary table 1).

To evaluate if there is a correlation between total amount of 5-HIAA and mobile phone usage activity we carried out the Pearson correlation analysis and acquired results revealed that there is a rather weak (*r*=0.6194) yet significant (*p*<0.01) correlation between the two datasets. The analysis related scatter plot (Figure 1) also uncovered that all values were grouped in two clusters, which correspond to weekdays and weekends. Further comparison of mobile phone activities between weekdays and weekends revealed that two datasets were indeed significantly different (*p*=0.0011, two-tailed Mann-Whitney test).

Since our initial assumption that population size, which contributed to formation of WW, differed significantly between weekdays and weekends was confirmed, it was decided that all further Kuldiga related data analyses shall be carried out separately for the whole Kuldiga data set, Kuldiga weekday data set, and Kuldiga weekend dataset.

3.3 Validation SARS-CoV-2 RNA extraction procedure through employment of surrogate virus

We employed Recombinant Semliki Forest Virus (rSFV) as a surrogate to monitor the efficiency and success rate of extraction procedures. Acquired results revealed that the rSFV recovery rate in Jelgava on average was 1.12 (median - 1.00) and in Kuldiga - 0.64 (median - 0.58). However, it varied greatly from extraction to extraction

and displayed no correlation with any of the other measurements that were carried out. Plausible explanations and recommendations are negotiated within the discussion section.

3.4 SARS-CoV-2 RNA copy number correlation with cumulative incidence

To assess whether the changes in acquired WW viral copy number correlated with the theoretical number of infected patients, we carried out a correlation analysis. However, since several researchers have reported that feces remain SARS-CoV-2 positive for up to 30 days after the diagnosis [38], the correlation analysis between viral copy number and 21, 28, and 35-day cumulative incidence of COVID-19 cases were also carried out. In addition to assessing whether the normalization against population size marker would provide an improvement to acquired results we also calculated the number of viral RNA copies per 1mg of 5-HIAA for all samples (Supplementary Table 1) and performed the same correlation analyses. Acquired results revealed that in Jelgava the correlation between both total and 5-HIAA normalized SARS-CoV-2 copy number and the cumulative incidence of disease cases was high regardless of the number of days that was used for calculation, but the correlations which involved total copy number were slightly higher (r>0.985, p<0.01) than those which involved 5-HIAA normalized number (r>0.972, p<0.01) (Figure 2, Figure 3A). Unlike with the former, in the case of Kuldiga, the correlation between all RNA copy number values and cumulative incidences of cases was very weak, the best one being between total RNA copy number and 35-day cumulative incidence of COVID-19 cases (r=0.44, p>0.05). However, both total and 5-HIAA normalized copy number of Kuldiga weekday samples displayed a strong correlation with 35-day cumulative incidence of SARS-CoV-2 positive tests, and the r-value of 5-HIAA normalized was slightly higher (r=0.949, p<0.05) than that of total copy number (r=0.910, p<0.05). In contrast to previous, weekend copy number data displayed a stronger correlation with a 14-day cumulative incidence (r=0.706 for total and r=0.630 for 5-HIAA normalized), but the p-value of both results was above 0.05 indicating that they were not significant (Figure 2, Figure 3B).

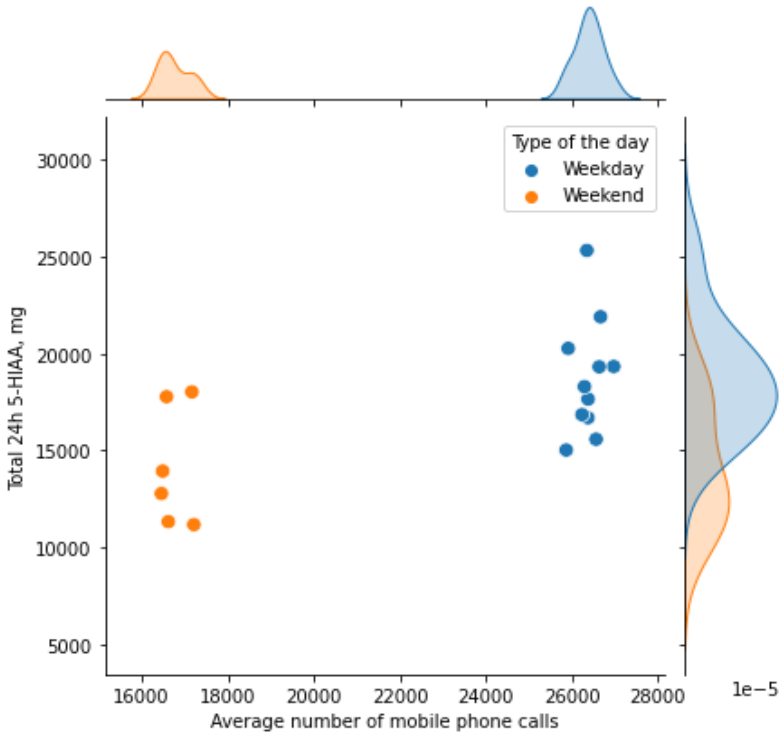


Figure 1. The scatter plot of 5-HIAA and CDR data in Kuldiga shows the effect of population activity on 5-HIAA measurement data depending on the type of the day (e.g., weekday or weekend).

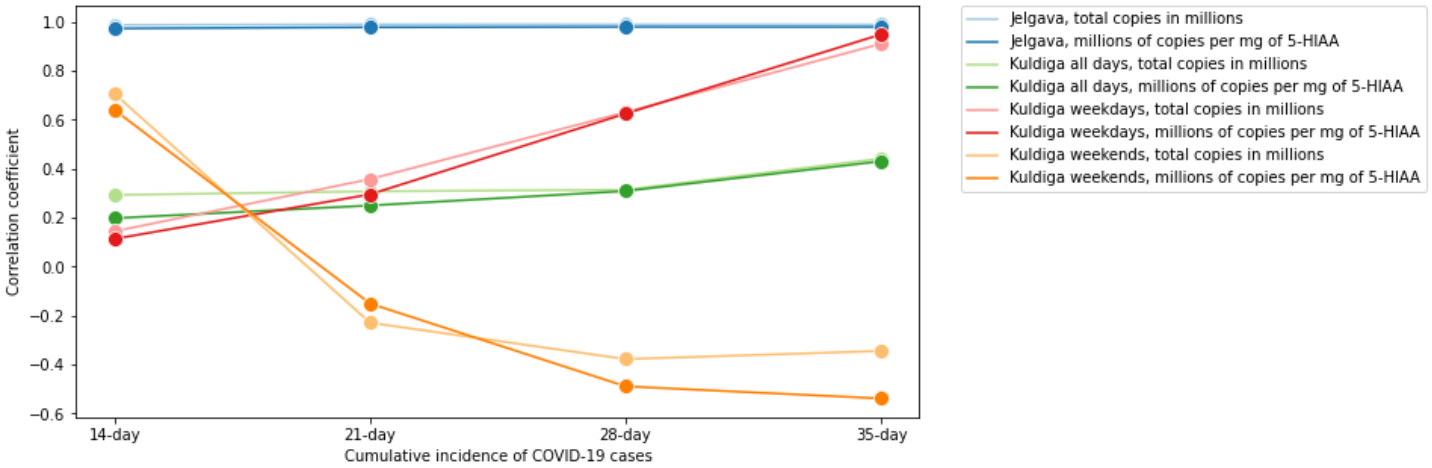


Figure 2. Changes in correlation coefficient between detected SARS-CoV-2 RNA copy number and a cumulative incidence of COVID-19 cases depending on the number of days that was used for the calculation of cumulative incidence.

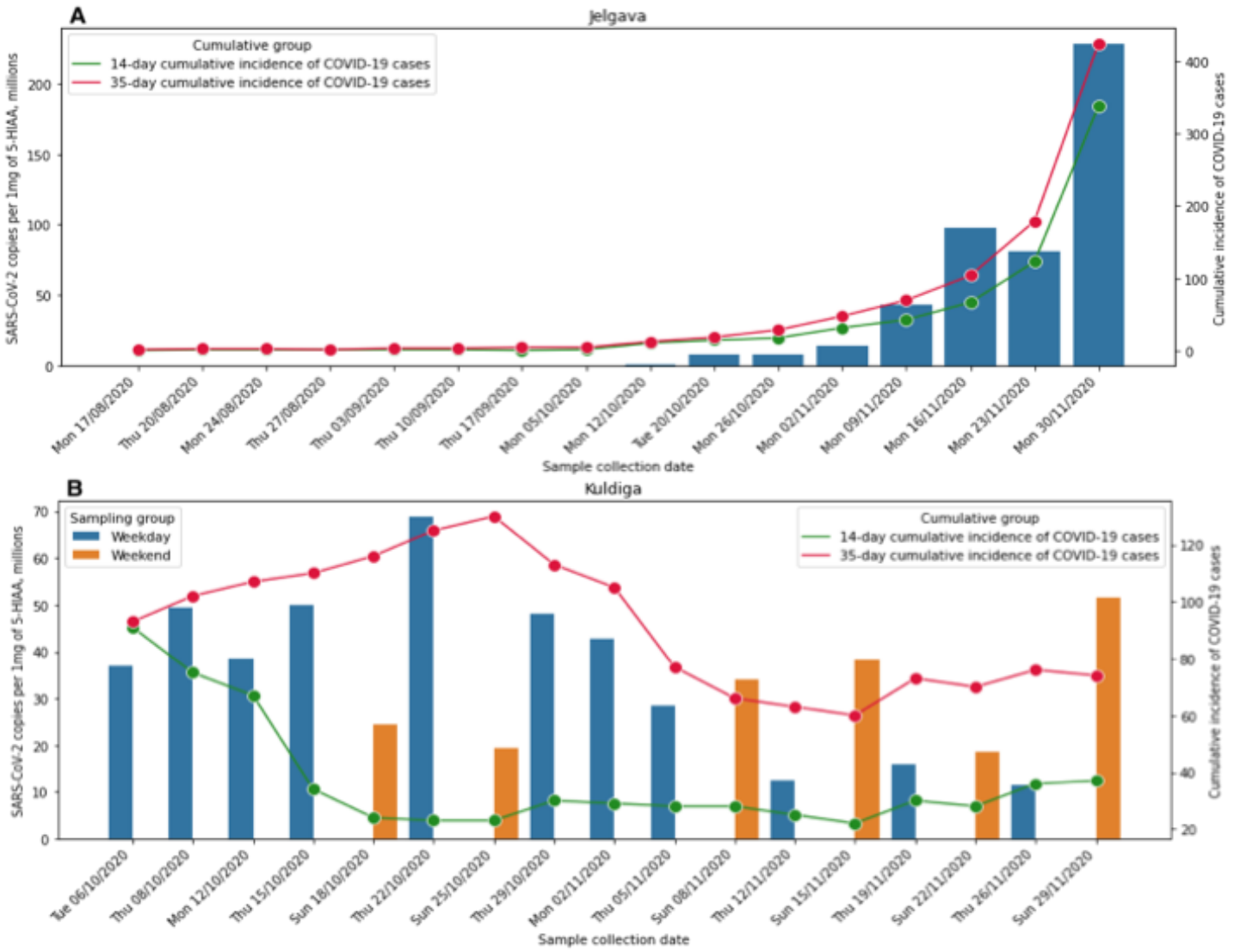


Figure 3. 5-HIAA normalized quantity of SARS-CoV-2 RNA within the wastewater of two Latvian locations A - Jelgava and B - Kuldiga.

4 Discussion

Although WW epidemiology has been used before to study various public health aspects, it gained worldwide interest during COVID-19 pandemic. This approach played an important role as an additional tool for observation of trends during disease surveillance, particularly in urban areas. Most of the studies have focussed on the large cities where weekly variation in population size is relatively small. The purpose of this study was to evaluate the applicability of SARS-CoV-2 monitoring in the WW for early detection of viral outbreaks and to assess the overall epidemiological situation during an outbreak in medium and small-sized municipalities. To achieve this goal the strategy of this study was based on longitudinal WW testing for the presence of SARS-CoV-2 RNA. Considering that this was a pilot study and thus limited in its scope and resources it was decided to perform the surveillance in only two municipalities.

Since Riga is the capital city of Latvia and the largest city in the Baltic region that accommodates ~30% of the Latvian population, it was not surprising that the outbreaks of COVID-19 at this location were detected regularly.

Therefore, we deemed that this city, although relevant from an epidemiological point of view, was not suitable for studies involving early detection of outbreaks. To reach this goal we selected a municipality that was relatively "free of SARS-CoV-2", was located in a no more than two-hour trip distance from Riga, maintains a strong economic connection to the capital - ensuring daily population movements between the locations, and has an independent economy, meaning that there is a strong necessity for communication between its inhabitants thus providing a good opportunity for the spread of the virus. Considering all these factors we selected Jelgava, which is the fourth largest city in the country, a major regional economic centre and residence of ~ 55 '000 inhabitants. Additionally, it was decided to closely monitor the epidemiological situation in all COVID-19 free Latvian towns with suitable sewage collection systems and initiate the collection of WW samples in one of them upon the first news of an outbreak. Such an opportunity presented itself at the beginning of October 2020, when an outbreak was detected in Kuldiga, which is the 15th largest town in Latvia and has a population of approximately 11'000 inhabitants.

The assessment of SARS-CoV-2 RNA presence in collected WW samples was carried out through ddPCR based quantification of three viral genome regions. The selection of primers and probes for Nucleocapsid (N) gene detection was based on "US CDC" recommendations while for envelope (E) gene detection on "WHO" recommendations. Acquired data yielded by all three probe/primer sets were comparable, but on average the numbers generated by the N1 set were higher than those generated by the other two sets and those generated by the N2 set were higher than those of the E set. However, this rule was not universal and since observed differences could be the result of the combinatorial effect of such factors as sample-specific RNA degradation, differences in probe/primer set detection efficiencies and other, we decided that rather than focussing on results of one assay we shall employ the average of three. Also, as explained earlier, to avoid the dilution effect, the total number of SARS-CoV-2 RNA copies within WW from each collection event were calculated.

To explain the observed variance in the extraction of surrogate, we performed additional literature data analysis on the stability of both viruses and acquired data revealed that SARS-CoV-2 remains viable in a wider range of conditions than rSFV. Thus the former can tolerate pH values in a range from 4 to 11 [39], while the latter can be inactivated at pH below 6.0 [40]. Also, it should be considered that WW is an aggressive environment that contains detergents and other compounds that may affect both viruses differentially and as the rule, viral particles of SARS-CoV-2 are spending a considerably longer period of time in this environment than those of rSFV, thus it is possible that at the moment of extraction they have reached some stability equilibrium and SARS-CoV-2 RNA is deteriorating at a slower rate than that of surrogate. Therefore, because of named reasons, we believe that it is essential to employ some kind of surrogate for the plus-minus test to assess whether the extraction as such was successful but acquired numbers should not be employed for correction of target virus quantification results unless it has been determined that stability of both target and surrogate under various conditions is highly similar.

In this study, the 5-HIAA measurements were performed to account for fluctuations in the size of the population that contributed to the formation of WW and the choice of this metabolite as a population size marker was based on studies published earlier [26,41–43]. To further validate the size of the population we added a call detail record of the study region. To our best knowledge, in this study we were able to show for the first time that CDR data along with the 5-HIAA measurements can support the disease prevalence studies with the estimation of population size that contributes to the WW. This might be especially important for cities with a population size below 100'000, as the WWs here are more heavily influenced by rain, industrial WWs and population movement, thus introducing the fluctuations of the detected SARS-CoV-2 genetic material and human excreted biomarkers.

However, further investigations towards the estimation of the population size by CDR and 5-HIAA in small-size cities are necessary.

In this study we observed a high level of fluctuation within detected SARS-CoV-2 RNA between consecutive sample collections. We speculated that this might be due to weekly changes in population size, which are characteristic for such small-sized municipalities as Kuldiga, when people leave the town during the weekend and return during the weekdays. This hypothesis was supported by 5-HIAA measurements and subsequently also confirmed by CDR data analysis. This, however led to the conclusion that although persons with confirmed infection cases have been put in strict isolation, a considerable proportion of detected SARS-CoV-2 RNA in the WW might arise from asymptomatic individuals, who unknowingly continue to spread the virus. At the same time, recovered individuals might still excrete the virus for a prolonged time. An additional aspect of the small-sized municipalities is a proportion of people that are travelling during the weekdays from the countryside to their workplace in the town, thereby, if they are COVID-19 positive, their virus genetic material would be detectable only during the workdays, while some proportion of people, who reside in the town, during the weekend travel abroad, thus introducing additional bias towards the detection of SARS-CoV-2 RNA in the WW. An additional explanation to observed fluctuations might also be the rural nature of the town, where WW for significant number households is collected in designated collection tanks and transportation of accumulated WW from both infected and non-infected households to WWTP is carried out exclusively on weekdays which might increase the concentration of viral RNA. Hence, these results denote that the collection of WW samples from small-sized municipalities, such as Kuldiga, should be performed either on a specific day of the week, preferably during the workdays or daily to either control the effect of population migration or provide sufficient data to recognize the pattern.

The 14-day cumulative incidence is used for the assessment of the COVID-19 incidence in different countries. The rationale behind the employment of this indicator is based on the observation that symptomatic disease lasts approximately 14 days and after this period the viral load of the patient is not sufficient to infect others. However, our data showed that a 14-day cumulative incidence displayed a good correlation with detected SARS-CoV-2 RNA copies in the WWs only in Jelgava, where the incidence of COVID-19 cases during the study period were either non-existent or on the steep rise, which might have a significant impact on the outcome. Due to the above-mentioned reasons, we evaluated cumulative incidence of different durations and concluded that the 35-day cumulative incidence was superior in explaining the variation of SARS-CoV-2 RNA copies in the WW. This correlation was not surprising as studies worldwide have reported that the virus can be excreted via human body fluids up to 40 days from the onset of symptoms [3–5,38]. Introducing the 35-day cumulative incidence of confirmed COVID-19 cases might provide better resolution in the prediction of the outbreak and possibly aid in developing a model to predict the number of infected individuals in a community based on WW testing.

An additional crucial aspect that should be considered during the conduction of WBE studies is the sample collection. Collection and viral RNA extraction procedures that are presented in this study are efficient yet labour intensive and difficult to apply in massive surveillance programmes. Also, the number of virus particles in the WW is relatively low, which sets limits to the accuracy of detection methods. Thus, there is a need to improve the sample collection and pre-treatment procedures through concentration of the samples without a significant increase in labour and costs. Several approaches are proposed including the addition of sorbent materials in the WW [44]. However, it would be simpler to employ the sample concentration through the multi-step membrane ultrafiltration that has been used in surface water analyses [45,46]. Moreover, sample collection coupled with

WW flow rate measurements and adjusted by this data would improve the overall representativity of collected sample and increase the accuracy of viral RNA and population size estimation even further.

The use of mobile data as a real-time data source has become increasingly relevant, which allows, for example, to assess the trends of regional economic development or to determine the change in population mobility patterns. In this study, the analysis of aggregate data from mobile network operator was successfully integrated to confirm 5-HIAA measurement observations and to evaluate whether the highest activity of the population is related to changes in SARS-CoV-2 RNA abundance in WW. However, in our study the correlation between these two datasets was weak and further studies are required to account for various WW collection system and mobile phone usage related parameters that could influence the correlation.

Many countries have recently adopted WBE as a cost-effective approach for wide-scale screening of the SARS-CoV-2 RNA in the WW, that along with the clinical data could provide information of potential virus transmission within the community, thus aiding the decision-making process for authorities of the public health sector. However, several issues have to be taken into account when using WBE as an epidemiological tool. At first, there is a gap of knowledge which prohibits precise estimation of the number of COVID-19 affected individuals using the WBE. Secondly, for medium-sized and smaller municipalities, the population movement should be considered while evaluating the presence of the SARS-CoV-2 RNA in the WW. Thirdly, there is a necessity to develop and implement time-effective, cost-effective and reliable methods to detect emerging SARS-CoV-2 variants within the population through the WBE surveillance.

In conclusion, this study validates the application of WBE as a reliable tool to investigate the prevalence of the infection within a community. We confirmed that WBE based SARS-CoV-2 RNA monitoring could be performed in middle and small size municipalities and acquired viral RNA data correlate with a 35-day cumulative incidence of SARS-CoV-2 infection cases. In addition, in small size municipalities due to the high influence of population migration on the SARS-CoV-2 RNA in the WW, the collection day of the WW samples should be carefully selected and following the continuation of the monitoring should be performed only on that particular weekday or carried out on a daily basis to facilitate pattern recognition.

5 Conclusion

- WBE approach is applicable in both larger cities and smaller municipalities with population size as small as 10'000.
- For smaller municipalities it is essential to perform measurements that enable population size control.
- 5-HIAA measurements are applicable for assessment of changes in population size.
- Weekly population movement and/or differences of population daytime activities on different days of the week are significantly affecting the WBE results in smaller municipalities, thus samples should be collected either daily or weekly on a specific day of the week.
- Collection of information on the total volume of WW that passed the collection site during sample collection provides control over dilution effect and enables assessment of total Viral RNA copy number and total 5-HIAA amount that was excreted by population.
- Observed differences in Kuldiga between weekdays and weekends are the result of either population movement or sanitation activities that are carried out during weekdays.
- There is room for improvements of sample collection procedures with the introduction of multistep raw WW ultrafiltration and coupling of sample collection rate to WW flow rate.

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.

Acknowledgments

We thank Latvian Water and Wastewater Works Association for supporting the sampling and access to wastewater plants.

Funding

The work was supported by project No. VPP-COVID-2020/1-0008: "Multidisciplinary approach to monitor, control and confine the COVID-19 and other future epidemics in Latvia."

Supplementary materials

Supplementary material associated with this study can be found in the online version.

REFERENCES

1 Center for Disease Prevention and Control. Covid-19 Statistics. 2021.https://www.spkc.gov.lv/lv/covid-19-statistika

2 Center for Disease Prevention and Control. COVID-19 tests, confirmed cases and outcomes. Published Online First: 2021.https://data.gov.lv/dati/lv/dataset/covid-19/resource/d499d2f0-b1ea-4ba2-9600-2c701b03bd4a

3 Lo IL, Lio CF, Cheong HH, *et al.* Evaluation of SARS-CoV-2 RNA shedding in clinical specimens and clinical characteristics of 10 patients with COVID-19 in Macau. *Int J Biol Sci* 2020;**16**:1698–707. doi:10.7150/ijbs.45357

4 Xing Y-H, Ni W, Wu Q, *et al.* Prolonged viral shedding in feces of pediatric patients with coronavirus disease 2019. *J Microbiol Immunol Infect* 2020;**53**:473–80. doi:10.1016/j.jmii.2020.03.021

5 Chen Y, Chen L, Deng Q, *et al.* The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *J Med Virol* 2020;**92**:833–40. doi:10.1002/jmv.25825

6 Lescure F-X, Bouadma L, Nguyen D, *et al.* Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *Lancet Infect Dis* 2020;**20**:697–706. doi:10.1016/S1473-3099(20)30200-0

7 Pan Y, Zhang D, Yang P, *et al.* Viral load of SARS-CoV-2 in clinical samples. *Lancet Infect Dis* 2020;**20**:411–2. doi:10.1016/S1473-3099(20)30113-4

8 Zhang H, Kang Z, Gong H, *et al.* The digestive system is a potential route of 2019-nCov infection: a bioinformatics analysis based on single-cell transcriptomes. *bioRxiv* 2020;:2020.01.30.927806. doi:10.1101/2020.01.30.927806

9 Bogler A, Packman A, Furman A, *et al.* Rethinking wastewater risks and monitoring in light of the COVID-19 pandemic. *Nat Sustain* 2020;**3**:981–90. doi:10.1038/s41893-020-00605-2

10 Wölfel R, Corman VM, Guggemos W, *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* 2020;**581**:465–9. doi:10.1038/s41586-020-2196-x

11 Wu Y, Guo C, Tang L, *et al.* Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *Lancet Gastroenterol Hepatol* 2020;**5**:434–5. doi:10.1016/S2468-1253(20)30083-2

12 Xu Y, Li X, Zhu B, *et al.* Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nat Med* 2020;**26**:502–5. doi:10.1038/s41591-020-0817-4

13 Wang X, Zheng J, Guo L, *et al.* Fecal viral shedding in COVID-19 patients: Clinical significance, viral load dynamics and survival analysis. *Virus Res* 2020;**289**:198147. doi:10.1016/j.virusres.2020.198147

14 Trottier J, Darques R, Ait Mouheb N, *et al.* Post-lockdown detection of SARS-CoV-2 RNA in the wastewater of Montpellier, France. *One Heal* 2020;**10**:100157. doi:https://doi.org/10.1016/j.onehlt.2020.100157

15 Gerrity D, Papp K, Stoker M, *et al.* Early-pandemic wastewater surveillance of SARS-CoV-2 in Southern Nevada: Methodology, occurrence, and incidence/prevalence considerations. *Water Res X* 2021;**10**:100086. doi:https://doi.org/10.1016/j.wroa.2020.100086

16 Gonzalez R, Curtis K, Bivins A, *et al.* COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. *Water Res* 2020;**186**:116296. doi:https://doi.org/10.1016/j.watres.2020.116296

17 Saththasivam J, El-Malah SS, Gomez TA, *et al.* COVID-19 (SARS-CoV-2) outbreak monitoring using wastewater-based epidemiology in Qatar. *Sci Total Environ* 2021;**774**:145608. doi:https://doi.org/10.1016/j.scitotenv.2021.145608

18 Hokajärvi A-M, Rytkönen A, Tiwari A, *et al.* The detection and stability of the SARS-CoV-2 RNA biomarkers in wastewater influent in Helsinki, Finland. *Sci Total Environ* 2021;**770**:145274. doi:10.1016/j.scitotenv.2021.145274

19 Randazzo W, Truchado P, Cuevas-Ferrando E, *et al.* SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Res* 2020;**181**:115942. doi:https://doi.org/10.1016/j.watres.2020.115942

20 Westhaus S, Weber F-A, Schiwy S, *et al.* Detection of SARS-CoV-2 in raw and treated wastewater in Germany – Suitability for COVID-19 surveillance and potential transmission risks. *Sci Total Environ* 2021;**751**:141750. doi:https://doi.org/10.1016/j.scitotenv.2020.141750

21 Hillary LS, Farkas K, Maher KH, *et al.* Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. *Water Res* 2021;:117214. doi:https://doi.org/10.1016/j.watres.2021.117214

22 Saguti F, Magnil E, Enache L, *et al.* Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Res* 2021;**189**:116620. doi:10.1016/j.watres.2020.116620

23 Ahmed W, Angel N, Edson J, *et al.* First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci Total Environ* 2020;**728**:138764. doi:10.1016/j.scitotenv.2020.138764

24 Peccia J, Zulli A, Brackney DE, *et al.* SARS-CoV-2 RNA concentrations in primary municipal sewage sludge as a leading indicator of COVID-19 outbreak dynamics. *medRxiv* 2020;:2020.05.19.20105999. doi:10.1101/2020.05.19.20105999

25 Polo D, Quintela-Baluja M, Corbishley A, *et al.* Making waves: Wastewater-based epidemiology for COVID-19 - approaches and challenges for surveillance and prediction. *Water Res* 2020;**186**:116404. doi:10.1016/j.watres.2020.116404

26 Thai PK, O'Brien JW, Banks APW, *et al.* Evaluating the in-sewer stability of three potential population biomarkers for application in wastewater-based epidemiology. *Sci Total Environ* 2019;**671**:248–53.

doi:10.1016/j.scitotenv.2019.03.231

27 O'Brien JW, Thai PK, Eaglesham G, *et al.* A Model to Estimate the Population Contributing to the Wastewater Using Samples Collected on Census Day. *Environ Sci Technol* 2014;**48**:517–25. doi:10.1021/es403251g

28 Rico M, Andrés-Costa MJ, Picó Y. Estimating population size in wastewater-based epidemiology. Valencia metropolitan area as a case study. *J Hazard Mater* 2017;**323**:156–65. doi:https://doi.org/10.1016/j.jhazmat.2016.05.079

29 Arhipova I, Berzins G, Brekis E, *et al.* Mobile phone data statistics as a dynamic proxy indicator in assessing regional economic activity and human commuting patterns. *Expert Syst* 2020;**37**:e12530. doi:https://doi.org/10.1111/exsy.12530

30 Fuqing W, Jianbo Z, Amy X, *et al.* SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *mSystems* 2021;**5**:e00614-20. doi:10.1128/mSystems.00614-20

31 Corman VM, Landt O, Kaiser M, *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 2020;**25**. doi:https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045

32 Wang G, Gentry TJ, Grass G, *et al.* Real-time PCR quantification of a green fluorescent protein-labeled, genetically engineered Pseudomonas putida strain during 2-chlorobenzoate degradation in soil. *FEMS Microbiol Lett* 2004;**233**:307–14. doi:10.1111/j.1574-6968.2004.tb09497.x

33 Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann Math Stat* 1947;**18**:50–60. doi:10.1214/aoms/1177730491

34 Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* 2007;**9**:90–5. doi:10.1109/MCSE.2007.55

35 Waskom ML. seaborn: statistical data visualization. *J Open Source Softw* 2021;**6**:3021. doi:10.21105/joss.03021

36 Oberg K, Couvelard A, Delle Fave G, *et al.* ENETS Consensus Guidelines for Standard of Care in Neuroendocrine Tumours: Biochemical Markers. *Neuroendocrinology* 2017;**105**:201–11. doi:10.1159/000472254

37 Chernecky C, Berger BH. *Laboratory Tests and Diagnostic Procedures*. 6th ed. Philadelphia, PA: 2013.

38 Gupta S, Parker J, Smits S, *et al.* Persistent viral shedding of SARS-CoV-2 in faeces – a rapid review. *Color Dis* 2020;**22**:611–20. doi:https://doi.org/10.1111/codi.15138

39 Chan K-H, Sridhar S, Zhang RR, *et al.* Factors affecting stability and infectivity of SARS-CoV-2. *J Hosp Infect* 2020;**106**:226–31. doi:https://doi.org/10.1016/j.jhin.2020.07.009

40 S.E. S, L. L, J.R. G. Sensitivity of bluetongue virus to lipid solvents, trypsin and pH changes and its serological relationship to arboviruses. *JHyg* 1966;**64**:339–45.

41 Pandopulos AJ, Gerber C, Tscharke BJ, *et al.* A sensitive analytical method for the measurement of neurotransmitter metabolites as potential population biomarkers in wastewater. *J Chromatogr A* 2020;**1612**:460623. doi:https://doi.org/10.1016/j.chroma.2019.460623

42 Chen C, Kostakis C, Gerber JP, *et al.* Towards finding a population biomarker for wastewater epidemiology studies. *Sci Total Environ* 2014;**487**:621–8. doi:https://doi.org/10.1016/j.scitotenv.2013.11.075

43 Choi PM, Tscharke BJ, Donner E, *et al.* Wastewater-based epidemiology biomarkers: Past, present and

future. *TrAC Trends Anal Chem* 2018;**105**:453–69. doi:https://doi.org/10.1016/j.trac.2018.06.004

44 Freda K. The myriad ways sewage surveillance is helping fight COVID around the world. 10 May. 2021. doi:https://doi.org/10.1038/d41586-021-01234-1

45 Ferguson C, Kaucner C, Krogh M, *et al.* Comparison of methods for the concentration of Cryptosporidium oocysts and Giardia cysts from raw waters. *Can J Microbiol* 2004;**50**:675–82. doi:10.1139/w04-059

46 Rusiñol M, Martínez-Puchol S, Forés E, *et al.* Concentration methods for the quantification of coronavirus and other potentially pandemic enveloped virus from wastewater. *Curr Opin Environ Sci Heal* 2020;**17**:21–8. doi:10.1016/j.coesh.2020.08.002

Manuscript "Detection of SARS-CoV-2 in wastewater and importance of population size assessment in smaller cities: an exploratory case study from Latvia"

HIGHLIGHTS

- Traces of SARS-CoV-2 RNA were detected in Latvian municipal WW.
- Population size control measurements improved the accuracy of COVID-19 prevalence estimation.
- Habits of citizen movement must be considered using WBE in small/medium-sized towns.

Dear editors of Water Research,

We are now submitting a manuscript entitled "**Detection of SARS-CoV-2 in wastewater and importance of population size assessment in smaller cities: an exploratory case study from Latvia**".

We confirm that this work is original and has not been published elsewhere, nor is currently under consideration for publication by another journal.

Although wastewater-based epidemiology (WBE) has gained the global importance during the COVID-19 pandemic, a number of aspects remain unassessed. Recently, the EU Commission have encouraged to apply WBE as a tool to monitor SARS-CoV-2 epidemiological situation in cities with population rate above 100'000. While wastewater of large cities is less affected by such factors as dilution of viral RNA due to the inflow of rain or industrial wastewater, population migration, presence of compounds that might degrade the virus and other, the wastewater of smaller municipalities is affected by these factors to a much greater extent. To mitigate the effects of above-mentioned factors on the estimation of viral RNA load in the wastewater, we conclude that quantification of SARS-CoV-2 RNA should be accompanied by additional cost- and time-effective population size measurements, which can aid the interpretation of the acquired results. Even more, we were able to show, that 35-day cumulative incidence of confirmed COVID-19 cases displayed strong positive correlation with the detected viral RNA quantities within the wastewater.

The study presented in article reflects on detection of SARS-CoV-2 traces in wastewater of small municipalities and surveillance of epidemiological situations through such analysis. It also reflects on the importance of population size control and explores possibilities that are provided by 5-HIAA – an experimental population size biomarker and digital records of mobile phone usage activity.

This is the first report on WBE based study in Latvia and one of few from Baltic region. We believe that this article gives novel knowledge on interpretation of SARS-CoV-2 RNA load in the wastewater and may serve as a basis for development of improved recommendations for epidemiological surveillance. Therefore, we hope that this manuscript will be acceptable for publication in Water Research.

Sincerely yours,

Davids Fridmanis, PhD on behalf of authors of this multidisciplinary research.

31.05.2021.

Manuscript "Detection of SARS-CoV-2 in wastewater and importance of population size assessment in smaller cities: an exploratory case study from Latvia"

DECLARATION OF INTEREST STATEMENT

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.

Electronic Supplementary Material (for online publication only)

Supplementary table 1.xlsx

Figure-1

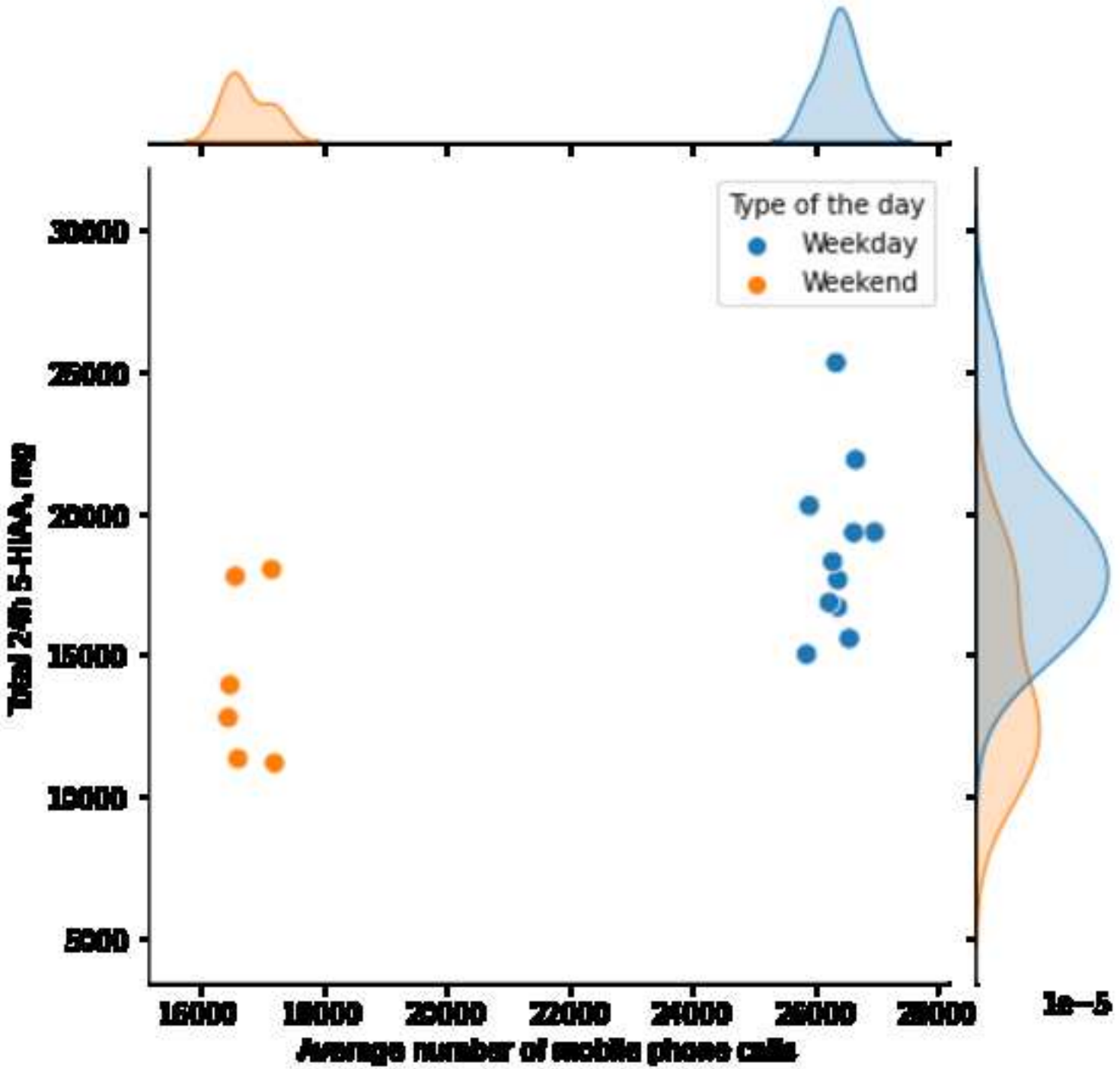

Type of the day
Weekday
Weekend
30000
25000
20000
15000
10000
5000
16000
18000
20000
22000
24000
26000
28000
1e−5
Average number of mobile phone calls

Figure-2

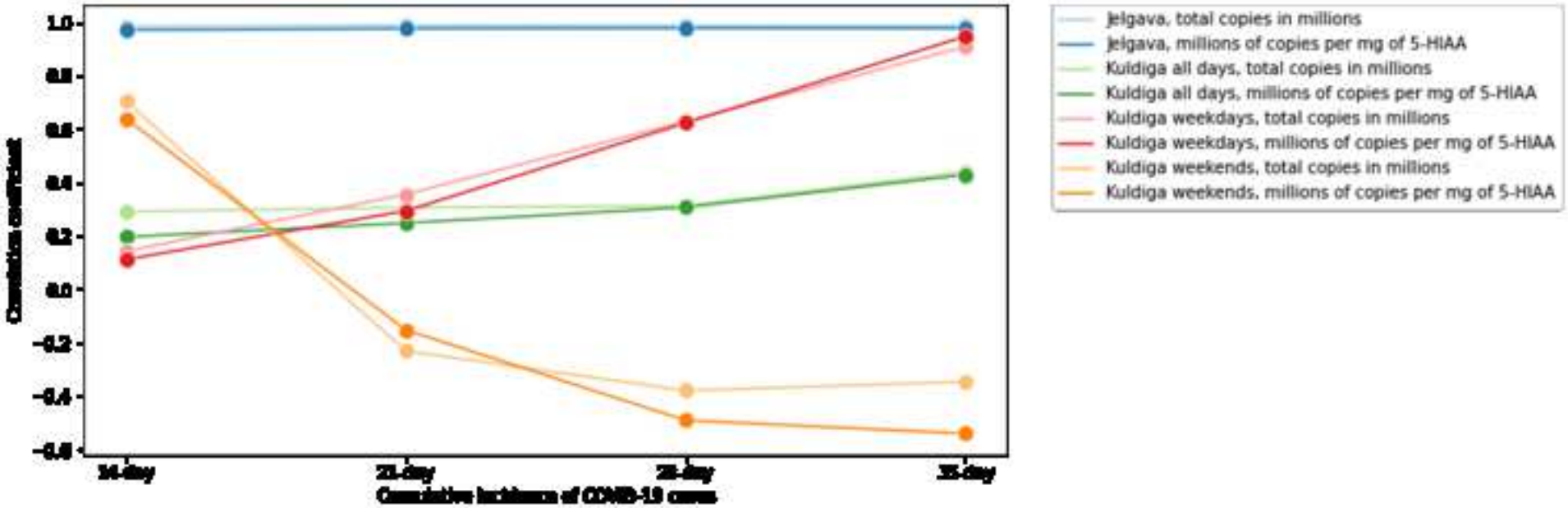
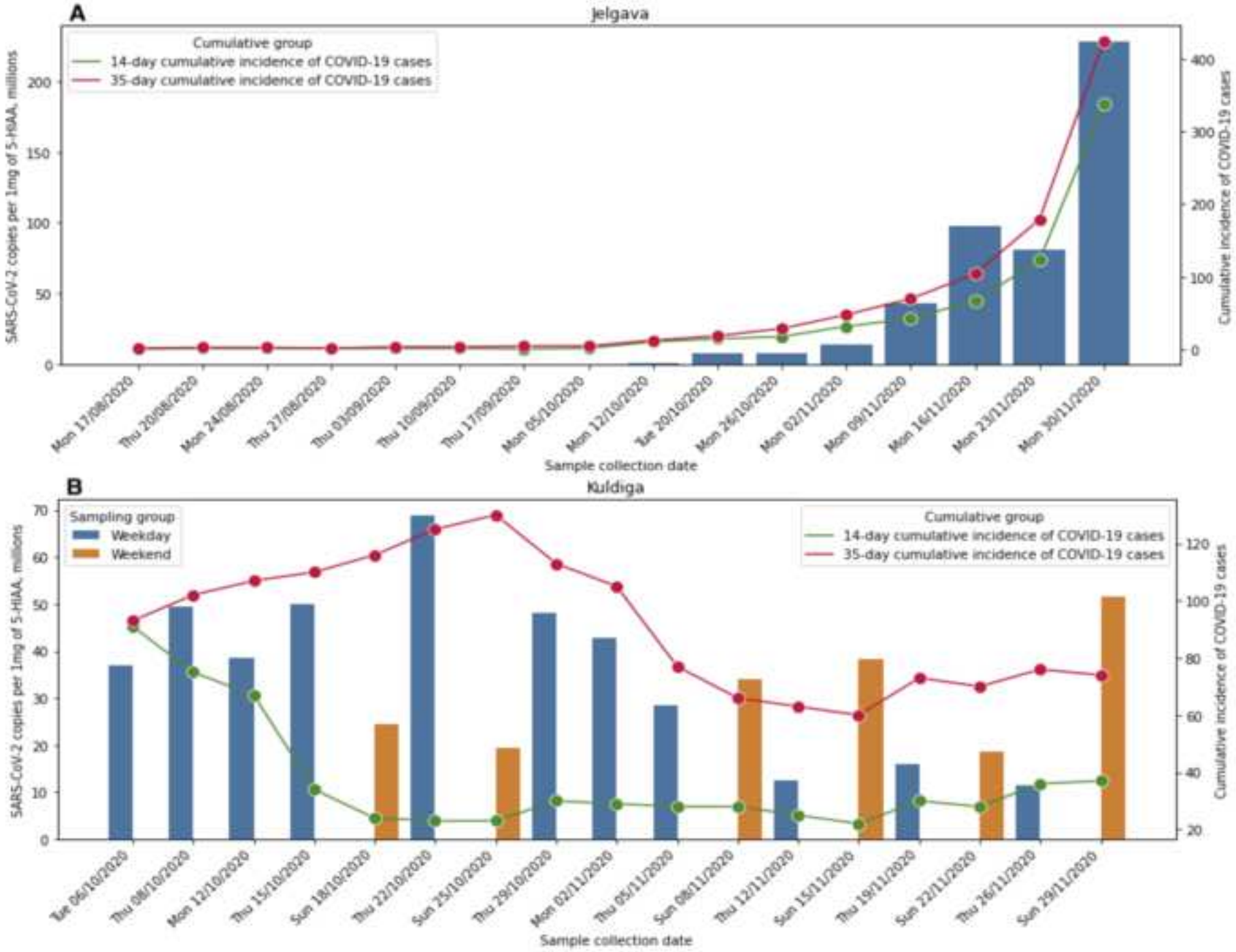

Jelgava, total copies in millions
Jelgava, millions of copies per mg of 5-HIAA
Kuldiga all days, total copies in millions
Kuldiga all days, millions of copies per mg of 5-HIAA
Kuldiga weekdays, total copies in millions
Kuldiga weekdays, millions of copies per mg of 5-HIAA
Kuldiga weekends, total copies in millions
Kuldiga weekends, millions of copies per mg of 5-HIAA

Figure-3


A
Jelgava
Cumulative group
14-day cumulative incidence of COVID-19 cases
35-day cumulative incidence of COVID-19 cases
SARS-CoV-2 copies per 1mg of 5-HIAA, millions
Cumulative incidence of COVID-19 cases
Sample collection date
B
Kuldiga
Sampling group
Weekday
Weekend
Cumulative group
14-day cumulative incidence of COVID-19 cases
35-day cumulative incidence of COVID-19 cases
SARS-CoV-2 copies per 1mg of 5-HIAA, millions
Cumulative incidence of COVID-19 cases
Sample collection date

Oligonucleotide name	Oligonucleotide sequence
nCoV_N1 Forward	GAC CCC AAA ATC AGC GAA AT
nCoV_N1 Probe	(6-Fam) ACC CCG CAT TAC GTT TGG TGG ACC (BHQ-1)
nCoV_N1 Reverse	TCT GGT TAC TGC CAG TTG AAT CTG
nCoV_N2 Forward	TTA CAA ACA TTG GCC GCA AA
nCoV_N2 Probe	(6-Fam) ACA ATT TGC CCC CAG CGC TTC AG (BHQ-1)
nCoV_N2 Reverse	GCG CGA CAT TCC GAA GAA
nCoV_E Forward	ACA GGT ACG TTA ATA GTT AAT AGC GT
nCoV_E Probe	(6-Fam) ACA CTA GCC ATC CTT ACT GCG CTT CG (BHQ-1)
nCoV_E Reverse	ATA TTG CAG CAG TAC GCA CAC A
GFP Forward	CTG CTG CCC GAC AAC CAC
GFP Probe	(Hex) CCA GTC CGC CCT GAG CAA AGA CC (BHQ-1)
GFP Reverse	TCA CGA ACT CCA GCA GGA C

Source

Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Center for Disease Control and Prevention, Atlanta, GA, USA

Corman *et al.* [31]

Wang *et al.* [32]

Science of the Total Environment 755 (2021) 142688

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www. elsevier. com/ locate/scitotenv

Rapid determination of pharmaceuticals in wastewater by direct infusion HRMS using target and suspect screening analysis

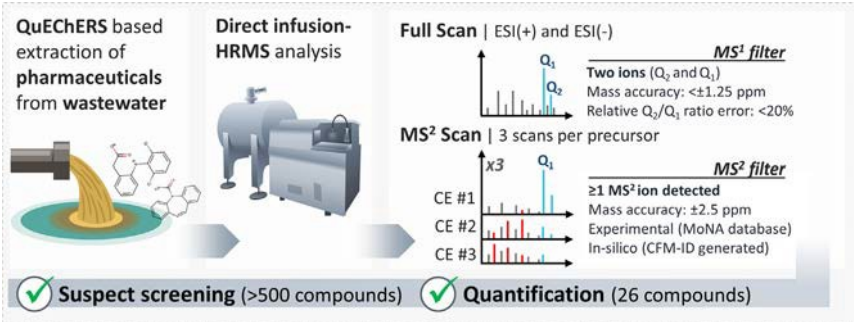
Ingus Perkons [a,b,*], Janis Rusko [a,b], Dzintars Zacs [a], Vadims Bartkevics [a,b]

[a] Institute of Food Safety, Animal Health and Environment "BIOR", Lejupes iela 3, Riga LV-1076, Latvia
[b] University of Latvia, Faculty of Chemistry, Jelgavas iela 1, Riga LV-1004, Latvia

HIGHLIGHTS

- Direct infusion-HRMS was used for suspect and target screening of pharmaceuticals.
- A modified QuEChERS approach was used for sample extraction and clean-up.
- 79 suspects and 24 target compounds were detected in 72 wastewater samples.
- Diclofenac, metoprolol, and telmisartan were the most frequently detected compounds.
- Diclofenac and carbamazepine metabolites showed high prevalence in wastewater samples.

GRAPHICAL ABSTRACT



article info

Article history:
Received 29 July 2020
Received in revised form 11 September 2020
Accepted 26 September 2020
Available online 3 October 2020

Editor: Yolanda Picó

Keywords:
Suspect screening
High resolution mass spectrometry
Pharmaceuticals
Wastewater
QuEChERS
FT-ICR-MS

abstract

A wide-scope screening of active pharmaceutical ingredients (APIs) and their transformation products (TPs) in wastewater can yield valuable insights and pinpoint emerging contaminants that have not been previously reported. Such information is relevant to investigate their occurrence and fate in various environmental compartments. In this study, we explored the applicability of direct infusion high resolution mass spectrometry (DI-HRMS) for comprehensive and rapid detection of APIs and their TPs in wastewater samples. The method was developed using a Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) system and incorporated both wide-scope suspect screening and semi-quantitative determination of selected analytes. The identification strategy was based on the following criteria: narrow accurate mass window (±1.25 ppm) for two most abundant full-MS signals, isotopic pattern fit and additional confirmation on the basis of MS^2 spectra at three fragmentation levels. The tentative identification of suspects and target compounds relied on an in- house database containing more than 500 different APIs and TPs. The measured fragment spectra were matched against experimental MS^2 patterns obtained from a publicly available spectral library (MassBank of North America) and *in-silico* generated fragmentation features (from the CFM-ID algorithm). In total, 79 suspects were identified and 24 target compounds were semi-quantified in 72 wastewater samples. The highest detection frequencies in treated wastewater effluents were observed for diclofenac, metoprolol and telmisartan, while hydroxydiclofenac, dextrorphan, and carbamazepine metabolites were the most frequently detected TPs. The obtained API profiles were in accordance with the national consumption statistics and the origin of wastewater samples. The developed method is suitable for rapid screening of APIs in wastewater and can be used as a

* Corresponding author at: Institute of Food Safety, Animal Health and Environment "BIOR", Lejupes iela 3, Riga LV-1076, Latvia.
E-mail address: ingus.perkons@bior.lv (I. Perkons).

https://doi.org/10.1016/j.scitotenv.2020.142688

complementary tool to characterize API emissions from wastewater treatment facilities and to identify problematic compounds that require more rigorous monitoring.

1. Introduction

Active pharmaceutical ingredients (APIs), their metabolites and transformation products (TPs) are increasingly being recognized as environmental contaminants of growing concern. Although these substances are classified as emerging contaminants, their unwanted presence in the environment, especially water bodies, has been already identified for several decades (Daughton and Ternes, 1999). Much of this pollution is released from urban wastewater treatment plants (WWTPs) due to two reasons: many pharmaceuticals are excreted from the human body either unchanged or as metabolites that can display biological activity, and conventional biological treatment processes are not sufficiently effective to completely remove these substances from the wastewater (WW) (Gao et al., 2012). Besides, there are several other pathways through which pharmaceuticals can enter the environment, e.g., direct discharge from API manufacturing sites (Phillips et al., 2010) and improper disposal of medicines (Bound and Voulvoulis, 2005). While more than a thousand monographs for APIs are included in the European Pharmacopoeia (Council of Europe, 2020), only a minor fraction of them are subject to routine environmental monitoring in the EU, for example, the ones listed in the surface water "Watch List" under the EU Water Framework Directive (Deloitte, 2018). Nevertheless, a large body of literature has provided strong evidence that environmentally relevant levels of APIs can pose a threat to the ecosystem (Desbiolles et al., 2018; Ebele et al., 2017). For example, adverse effects have been observed due to antidepressants such as fluoxetine and bupropion, which have shown the capability to induce behavioral abnormalities in fish (Schultz et al., 2010; Weinberger and Klaper, 2014). This is just the tip of an iceberg, because in contrast to many other chemical pollutants, APIs are designed to alter biochemical processes at low concentrations. Therefore, it should not come as a surprise that many of them can cause unintentional harm to other species. In addition, previously undocumented risks can also emerge due to physicochemical interactions between certain APIs and other substances that are present in the environment, causing a so-called "cocktail effect" (Escher et al., 2020; Styrishave et al., 2011; Vasquez et al., 2014).

Comprehensive investigation of pharmaceutical residues in environmental samples requires highly sophisticated methods due to their diverse chemical properties and relatively low concentration levels. Historically, the multi-residue methods applied to resolve analytical challenges presented by polar organic contaminants were almost exclusively based on liquid chromatography (LC) and tandem mass spectrometry (MS^2) (Gros et al., 2006). However, recent advancements in high-resolution mass spectrometry (HRMS) have allowed researchers to implement novel analytical strategies that involve non-target and suspect screening techniques. Although targeted methods can provide highly accurate information about the occurrence of these substances, non-target methods can extend the current scope and acquire more detailed data that helps to identify additional compounds of interest. The most suitable MS technologies applied in this field are based on quadrupole time-of-flight-MS (QTOF-MS) and hybrid Orbitrap-MS systems that can achieve both high resolving power and scanning frequency. In addition, most of these applications involve simultaneous acquisition of MS^2 and even MS^3 spectra, significantly boosting the reliability of the acquired data (Bletsou et al., 2015). For instance, Singer et al. (2016) applied LC-Orbitrap-MS with incorporated data-dependent MS^2 acquisition to investigate more than 800 APIs in WW effluents (Singer et al., 2016), whereas a more recent study reported an LC-QTOF-MS application that can perform suspect screening of more than 1000 APIs and 250 metabolites in hospital WW (Wielens Becker et al., 2020). Alternatively, other types of HRMS instruments can be used, such as Fourier transform ion cyclotron resonance MS (FT-ICR-MS). The latter is frequently applied in several fields that require the analysis of unknowns, e.g. petroleomics, metabolomics, proteomics, and also for the characterization of dissolved organic matter. Apart from the cost, it has a critical disadvantage that is more pronounced compared to its counterparts - FT-ICR-MS technology can sustain ultra-high resolving power only at the expense of scanning speed. Hence, it is rarely used in combination with high-performance LC systems. Nevertheless, sample introduction techniques such as direct infusion (DI) and flow injection (FI) can help to unlock its full potential. DI-FT-ICR-MS has also been applied by one participant in EPA's (*US* Environmental Protection Agency) non-targeted analysis collaborative trial (Ulrich et al., 2019). In addition, this technique has been successfully used to study disinfection by-products in drinking water (Andersson et al., 2019) and for non-target screening of brominated compounds (Zacs et al., 2019; Ziegler et al., 2019).

Although this field is rapidly developing, many critical challenges remain. According to the conclusions derived from the first collaborative non-target screening trial run by the NORMAN Association, the analytical methods applied in this field are already reasonably well harmonized. Meanwhile, there is a great disparity in the data processing strategies (Schymanski et al., 2015). A significant discrepancy emerges between the workflows, because the inventory of applicable tools varies from method to method. For instance, a recent study by Hohrenk et al. (2020) investigated the consistency of data processing in terms of feature extraction from LC-MS data by applying different software tools. Their study found that the non-targeted mode yielded only around 10% overlap of features between the four programs, while between 40% and 55% of the features did not match with any other program. Not surprisingly, more repeatable results were achieved using the suspect screening strategy where the average rate of successful identifications between the programs varied from 64% to 88% (Hohrenk et al., 2020). There is no question that non-target screening strategies are particularly useful in situations where no or limited information is available regarding the compounds of interest. However, even in an ideal scenario when all the features are accurately extracted, correct identification of unknowns still remains a difficult puzzle to solve, thus suspect screening can be considered as a more reliable strategy when compound-specific information for suspects is available (Krauss et al., 2010). Meanwhile, the DI approach might in some cases serve as an attractive alternative for LC-based applications, allowing to reduce difficulties during feature extraction.

In this study, we applied the FT-ICR-MS technology for target and suspect screening of APIs and their TPs in WW samples from Latvia. The sample preparation method was derived from a modified QuEChERS protocol and was able to provide a wide analytical scope. Furthermore, combining experimental MS^2 data from MassBank with *in-silico* generated fragment spectra allowed rapid MS^2 fingerprinting of the analytes of interest. The main objectives of this study were as follows: (i) to evaluate the applicability of FT-ICR-MS technology for target and suspect screening of APIs in WW samples, (ii) to gain new insights regarding the occurrence of selected APIs and their TPs, especially those characterized by poor removal efficiency during biological WWTP treatment processes and (iii) to compare our findings with literature reports and pharmaceutical consumption data.

2. Materials and methods

2.1. Samples

A total of 72 samples (36 influents, 36 effluents) were collected from different WWTPs from Latvia during March and April of 2019. The effluents and influents were collected in pairs on the same day from each sampling location. The samples were collected in amber glass bottles, kept at +4 °C during transportation and were filtered on the day of the delivery to the laboratory through 1.2 μm glass microfiber filters (GF/C, Whatman, UK). After that, a 20 mL aliquot (3 replicates per sample) was transferred to a 50 mL polypropylene tube and stored at −16 °C in darkness until the time of the analysis. The storage period ranged from one to four weeks. More detailed information about each sample, including the main characteristics (biochemical oxygen demand, suspended solids, and chemical oxygen demand) and the WWTPs of origin are presented in the Supplementary material (Table S1).

2.2. Standards and sample preparation

All the solvents were of HPLC grade and were purchased from Fluka (Buchs, Switzerland) or Sigma-Aldrich (St. Louis, MO, USA). Deionized water was generated with a Milli-Q water purification system (Millipore, Billerica, MA, USA). All the analytical standards used in this study were obtained either from Sigma-Aldrich (St. Louis, MO, USA) or Fluka (Buchs, Switzerland). Sepra™ C_{18}-E (Bulk Packing, 50 μm, 65 Å), Sepra™ (Bulk Packing, 50 μm, 70 Å), and Strata™-X-A 33 μm Polymeric Strong Anion sorbent were obtained from Phenomenex (Torrance, CA, USA), while anhydrous magnesium sulfate (>98%) was purchased from Chempur (Poland).

Each polypropylene tube containing a frozen 20 mL aliquot of WW was uncapped, covered with aluminum foil and freeze-dried using Benchtop "K" Series freeze dryer (VirTis, Gardiner, NY, USA). The freeze-drying process was paused when the volume reached about 5 mL. The samples were then thawed, vortexed for 30 s, frozen and again subjected to the freeze-drying procedure. The aforementioned step was included to ensure that all dry matter from the sample remained at the bottom of the container. After the freeze-drying proce- dure, 1 mL of acetonitrile/water (9:1, *v*/v) was added to the dry residue, vortexed for 1 min, and sonicated for 10 min. The extract was then cen- trifuged for 10 min at 3500 rpm and carefully transferred to a 15 mL polypropylene tube, followed by the addition of 1 mL of acetonitrile/ water (1:9, v/v). Phase separation was induced by adding 500 mg of an- hydrous magnesium sulfate to the extract. In order to minimize the pos- sible decomposition of target analytes due to the exothermic dissolution of $MgSO_4$, the extract was cooled to +4 °C before the addition. After that, the sample was vortexed for 1 min, centrifuged for 5 min at 3500 rpm, and the acetonitrile layer (600 μL) was transferred to a 2 mL V-shaped glass vial that contained 4 mg of dSPE sorbent (C_{18}-E/ Strata-X-A, 3:1, w/w). The mixture was immediately vortexed for 30 s and centrifuged for 5 min at 3500 rpm. Finally, a 100 μL aliquot was transferred to a 2 mL amber glass vial, diluted with 1100 μL of acetonitrile/water (4:1, v/v), and analyzed within 24 h by DI-FT-ICR-MS.

2.3. Direct infusion high-resolution mass spectrometry

The analysis was performed on a 7 T SolariX FT-ICR-MS system (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization source (ESI). Samples were introduced into the system by direct infusion via syringe pump at a flow rate of 5 μL/min. The mass spectra were acquired in both positive and negative ionization modes using a broadband mode (*m/z* range 50–1000). The full MS data were ac- quired by accumulating 32 scans, whereas MS^2 spectra were acquired without multiple spectra accumulation. The time domain data size was set at 4 M for all modes. The observed resolving power (FWHM) using these settings was around 490,000 at *m/z* 250 and 275,000 at *m/ z* 500. Three MS^2 spectra were acquired per precursor at different colli- sion energy (CE) values (5 V, 15 V, and 25 V) using an isolation window of ±2.5 *m/z*. In order to increase the fragment ion intensities, the ion accumulation time was set to 500 ms, whereas for the full-MS experiments this parameter was set to 100 ms and 20 ms for the positive and negative ionization modes, respectively. The main source parame- ters were as follows: nebulizer pressure 0.5 bar, dry gas flow rate 6 L/ min, dry temperature 200 °C, and capillary voltage 3.5 kV. The instru- ment was externally calibrated with sodium formate clusters, while the lock mass was applied to maintain high mass accuracy (see the Supplementary material, Table S2). Data acquisition was performed using ftmsControl 2.2.0 software (Bruker) and raw data processing was done with the DataAnalysis 5.0 software (Bruker).

2.4. Suspect list database

The initial list of suspects was derived from a publicly available database developed by the German Federal Environmental Agency. The database resulted from a systematic review of 1519 publications and comprises 178,708 data entries that summarize the occurrence data of 771 APIs and their TPs (aus der Beek et al., 2016). In order to select the compounds that can be investigated by this HRMS-based screening method, the following filters were applied: (i) monoisotopic mass range from 100 to 1000 Da, (ii) literature source credibility "good", and (iii) at least one MS^2 spectrum available in the MassBank of North America (MoNA) repository that has been obtained by ESI ionization. In addition, entries that can be classified as salts or mixtures were converted back to their parent molecules. After the removal of unsuitable entries, ~600 unique suspects remained on the list. As the next step, each entry was supplemented with molecular structure information (e.g., the molecular formula, SMILES and InChIKey) and full-MS features for both ionization modes, such as the accurate mass and isotopic pattern. The latter two were calculated for the $[M + H]^+$ and $[M-H]^-$ species by using the enviPat R package (Loos et al., 2015). LogD calculations were carried out by ChemAxon's web calculator "LogD Predictor" at pH 7.0 (https://disco.chemaxon.com/calculators/demo/plugins/logd/).

In order to obtain an MS^2 fingerprint, each compound was matched against the MoNA repository. In total, more than 8000 experimental records were found for compounds on the list. Each spectrum was subsequently reprocessed with MetFrag (Ruttkies et al., 2016), an *in-silico* fragmentation algorithm (mass error threshold ±5 ppm, relative abundance >10%). This step was implemented to annotate the fragment ions and to assign the correct *m/z* values, because the experimental data, for the most part, lacked fragment formula annotations and could contain erroneous peaks due to poor mass accuracy or interferences. The obtained dataset with experimental MS^2 traces was then aggregated on the basis of the respective precursors and polarity, and the five most prevalent fragments were selected as the MS^2 fingerprint. However, it was not always possible to obtain five features due to limited data. Hence, a complementary approach was applied, which involved the prediction of MS^2 spectra via Competitive Fragmentation Modeling-ID (CFM-ID) (Blaženović et al., 2017). Finally, the fragments obtained in this manner were again processed analogously to the experimental data and added to the database in order to improve the accuracy of the method, especially for suspects that gave fingerprints containing less than five *m/z* values (Chao et al., 2020). The steps performed to prepare the suspect list have been summarized in a flowchart (Fig. 1, left side). The suspect database is available as a spreadsheet (dataset S1 in the Supplementary material) and contains the following information for each entry: structural information (formula, SMILES, and InChIKey), full-MS traces (*m/z* values, isotopic patterns), and MS^2 traces (*m/z* values and the corresponding fragment formulas for both experimental and predicted features).

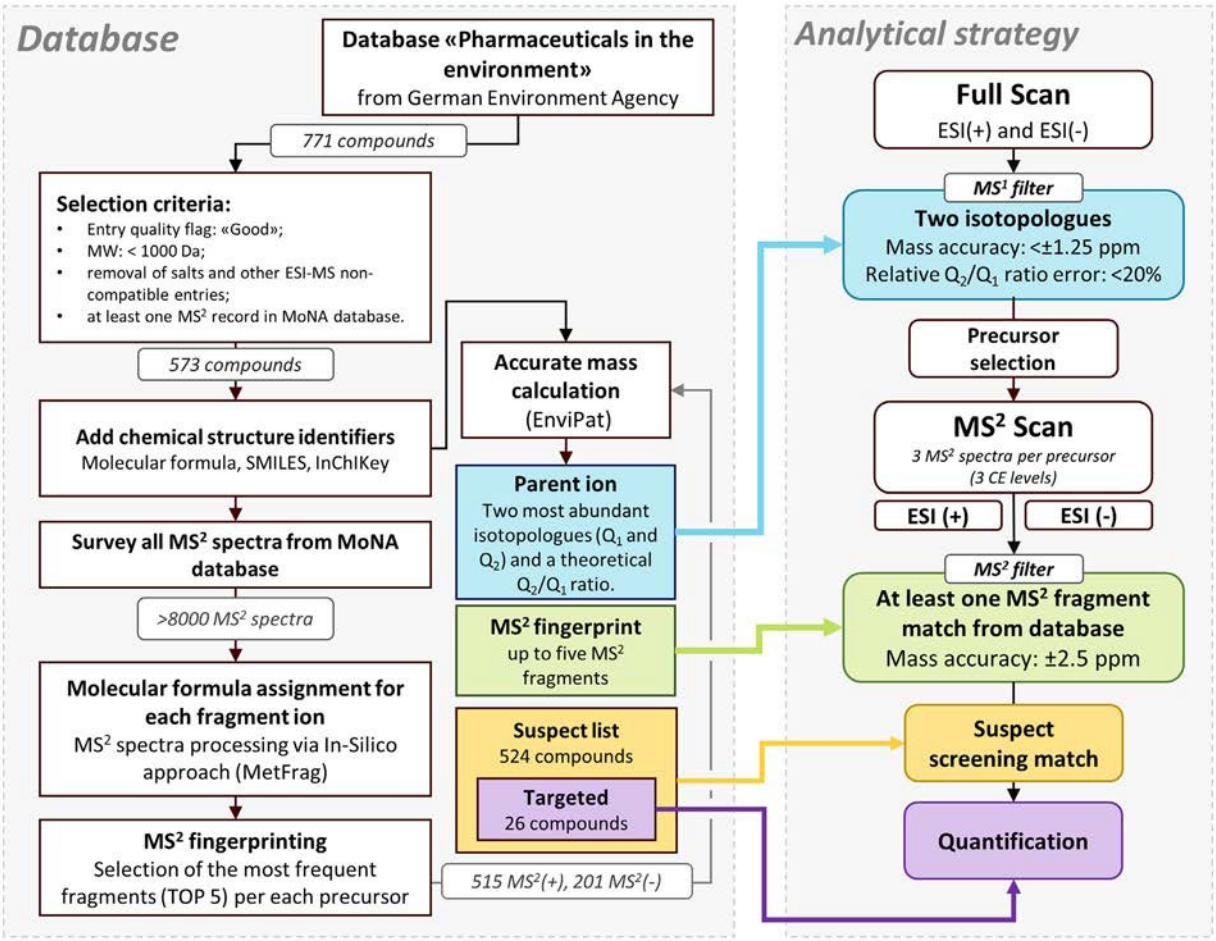


Fig. 1. A schematic overview of the screening workflow and the in-house database.

2.5. Target and suspect screening workflow

Initially, one full-MS spectrum (accumulation of 32 scans) was acquired for each ionization mode. Lock mass calibration and blank subtraction (from matrix-free reagent blank) was performed using Data Analysis 5.0 software (Bruker) and automated with VBScript (a subset of Visual Basic from Microsoft). The list of all peaks including *m/z* values and the corresponding peak intensities within the selected *m/z* range (from 100 to 1000) was exported as comma-separated values (CSV) file and further processed in the RStudio environment (v. 1.1.463) with R (v. 3.5.3). The full-MS features were extracted and matched against the suspect list using the following criteria: accurate mass error for the two most abundant ions (Q_1 and Q_2) ±1.25 ppm and the relative error between the theoretical and experimental ion ratios (±20% for suspect screening; from ±20% to ±50% for targeted screening). After this, suitable matches were further advanced to the next step – the MS^2 acquisition. Three MS^2 spectra were obtained per precursor at different CE values (5, 15, and 25 V). In order to save time, minimize human error, and to assist in data handling, the MS^2 acquisition method files were created automatically in the Rstudio environment by modifying a template method. In brief, the method consisted of stitched segments where each segment acquired one scan per precursor using a defined CE value. For instance, if the full-MS data yielded 30 compliant matches from the suspect list in positive mode, an MS^2 acquisition method was automatically generated with 90 scan segments and the measurements were immediately performed by the HRMS system. The set of measurements produced only one raw file for each ionization mode, which was automatically processed by VBScript in the Data Analysis software, sliced into 90 individual scans, exported as separate CSV files, and, finally, evaluated using R. The same procedure was applied for the negative mode. The features obtained from the MS^2 data were subsequently matched against the suspect list database that contained accurate mass information of both experimental and predicted fragmentation fingerprints. If at least one fragment showed a match (mass accuracy ±2.5 ppm) then the suspect was considered to be tentatively identified.

The target analysis was carried out simultaneously with the suspect screening. Quantification was based on Q_1 ion intensities from the full-MS data, using a matrix-matched calibration. The total time of the workflow depended on the number of matching suspects. In particular, the data processing (~5 min) and acquisition of two full-MS spectra (2 × 68 s) required approximately 7 min, while each MS^2 measurement extended the total time of the analysis by 4.5 s per precursor. Therefore, the average length of the workflow ranged from 10 to 15 min per sample. The outline of the workflow is depicted in the flowchart on the right side of Fig. 1 and a more comprehensive step-by-step guide is illustrated in Fig. S1 (Supplementary material). The R scripts applied in this study and an example dataset can be accessed online (https://github.com/ingusperkons/HRMS-screening-of-pharmaceuticals-in-wastewater).

2.6. Validation and quality control

After optimization, the target analysis section of the workflow was validated. Validation was carried out on a pooled WWTP influent and effluent matrix at three fortification levels (six replicates per level) over a period of two days. More detailed information about the fortification

levels is presented in Table S3 in the Supplementary material. The method was validated in terms of linearity, detection limit, decision capability, intra-day repeatability, between-day repeatability, and recovery. A matrix-matched calibration curve was applied to compensate for the matrix effect and to facilitate more accurate quantification. In addition, quality control (QC) samples and matrix-free procedure blanks were prepared once every 10 samples to monitor the performance of the method. The validation results along with a more detailed discussion are presented in Section 3.3.

At the same time, quality control for non-target methods is much more difficult to implement. Besides, additional challenges arise in the absence of chromatographic separation, preventing the application of additional evaluation criteria, such as retention time deviation. Nevertheless, the availability of ultra-high resolving power can at least partially outweigh the aforementioned drawback. In our case, a critical issue is the possibility of a false positive match that can occur according to two scenarios. First, the FT-ICR-MS instrument itself produces artifact peaks (e.g. Gibbs, harmonic, and magnetron peaks). Fortunately, these peaks can be excluded via the peak-picking algorithm integrated in the software. On the contrary, non-artificial peaks caused by isobaric compounds (i.e., substances that have the ability to produce MS signals that cannot be resolved from the target signal) and constitutional isomers cannot be removed. Therefore, additional measures were taken to minimize the rate of false positives, as further discussed in Section 3.3.

2.7. Optimization experiments

All optimization experiments described in Sections 3.1 and 3.2 were performed in triplicate using a pooled WWTP influent sample (unless stated otherwise). Selected fortification level for all optimization experiments corresponded to the concentration that was used for the second validation level. All sorbents (C18, PSA, and Strata-X-A) were washed with acetonitrile prior application to eliminate procedural interferences. Absolute recovery of APIs was evaluated by comparing the signal intensities of target analytes between procedural blanks (ultrapure water) that were spiked before and after the full sample preparation procedure. The matrix effect ($ME_{ionization}$) was calculated by dividing the analyte response in matrix matched extract by the response in procedure blank and multiplying this value by 100%. A value of 100% indicates no effect, whereas values below and above 100% indicate ionization suppression or ionization enhancement, respectively. The extracts that were used to calculate matrix effect were fortified using the post-extraction addition approach.

3. Results and discussion

3.1. Optimization of the sample preparation protocol

The initial sample preparation tests were conducted with a conventional solid phase extraction (SPE) technique that had been successfully applied in our previous study (Pugajeva et al., 2017). Even though this is a standard technique in numerous LC-MS based multi-analyte methods, high matrix suppression was observed in the case of DI-HRMS. This effect was especially pronounced in the negative ESI mode and required considerable dilution (>20-fold). Therefore, alternative sample preparation protocols were evaluated. In this context, QuEChERS (quick, easy, cheap, effective, rugged, and safe) was recognized as a potential substitute for SPE-based extraction. While this methodology is mainly used for pesticide residue analysis, it has recently gained increased recognition in the area of pharmaceutical residue detection (Peña-Herrera et al., 2019; Pérez-Lemus et al., 2019). For instance, a QuEChERS protocol has been applied for the determination of 25 APIs in sediment samples, producing satisfactory results (analyte recovery between 64% and 101%) (Nannou et al., 2019). Nevertheless, QuEChERS is rarely considered as a practical option when it comes to water samples, because only a limited sample volume can be extracted at once. A way to overcome this limitation is freeze-drying, since water removal may allow to minimize the co-extraction of matrix components, because only a limited fraction of dry matter dissolves in the extraction medium (acetonitrile–water). Obviously, precautions must be taken to avoid significant analyte losses. However, methodologies exist where the freeze-drying technique has been successfully applied to WW matrixes for the determination of APIs (Boulard et al., 2018; Montes et al., 2017). Recently, Mechelke et al. (2019) reported an elegant way of WW sample preconcentration prior to suspect screening of polar emerging pollutants (including APIs). This study relied on vacuum-assisted evaporative concentration and further confirmed that evaporation techniques can be applied to target and non-target analysis of APIs. Based on these considerations, freeze-drying was used as a pretreatment step before the extraction and the QuEChERS approach was further modified to meet the needs of the DI-HRMS method.

A preliminary optimization was carried out without matrix to evaluate the compatibility of QuEChERS salts and buffering agents with the DI-HRMS method. Two troublesome issues were observed. First, sodium chloride (NaCl) is partially transferred to acetonitrile during the phase separation and consequently causes interfering MS signals that correspond to NaCl clusters: $[Na_nCl_{n+1}]^-$ and $[Na_{n+1}Cl_n]^+$ (Fig. S2 in supplementary material). Second, citric acid produced an extremely intense peak at *m/z* 191.01973 in the negative ionization mode. Neither NaCl nor citrate buffers impaired the analytical performance of LC-MS based applications, yet they could pose serious problems in DI-HRMS analysis (e.g., suppress the ionization of analytes, oversaturate the ICR cell and promote the formation of additional adducts). Attempts were made to remove NaCl and citric acid residues from the final extract by evaporation and subsequent solvent exchange. For this purpose, we investigated the applicability of dichloromethane, ethyl acetate, acetone, and methyl tert-butyl ether. Unfortunately, none of these solvents provided satisfactory results (Fig. S2 in the supplementary material). Therefore, the use of NaCl and citrate salts was rejected and phase separation/extraction relied solely on $MgSO_4$. Next, the amount of anhydrous $MgSO_4$ was optimized. The most favorable conditions in terms of analyte recovery were found when the ratio (mg/μL) between $MgSO_4$ and acetonitrile-water (1:1, v/v) was from 0.225 to 0.275 (Fig. S3). The draft method was tested on WWTP influent samples and strong matrix suppression still prevailed.

In order to minimize the co-extraction of matrix components from freeze-dried material, different percentages of acetonitrile in water were investigated (50–100%). The results revealed that signal suppression decreased with higher organic phase content, but, at the same time, several APIs displayed poor extraction efficiency likely due to their relatively polar nature (e.g., caffeine, NSAIDs, and the macrolide antibiotics). Hence, acetonitrile-water (9:1, v/v) was selected as the most suitable option and an additional sonication step was incorporated to assist more efficient analyte transfer to the extraction medium (Fig. S4 in supplementary material).

However, these efforts were not sufficient to counter the ion suppression, which was particularly pronounced in the negative ionization mode. Although it is nearly impossible to pinpoint the substances responsible for this adverse effect, we noticed four ubiquitous signals at *m/z* 297.153, *m/z* 311.169, *m/z* 325.185, and *m/z* 339.200 that displayed around 10-fold higher intensity than all the other MS peaks. A putative identification was carried out, suggesting that these interferences are anionic surfactants (see Fig. S5B and C in the Supplementary material), in particular, linear alkylbenzene sulfonates ($C_nH_{2n-1}C_6H_5O_3S$, n = 10–13). Yet, surfactant residues should not be considered the primary underlying cause of suppression, since dissolved organic matter in wastewater comprises a complex mixture of organic substances with molecular mass up to 100 kDa. Therefore, further clean-up using dispersive SPE (dSPE) was evaluated. In particular, the traditional dSPE sorbents (C_{18} and PSA) and an anion exchange sorbent (Strata-X-A) were tested for the removal of surfactant residues and other co-extractable

interferences. Extracts from matrix-matched samples were treated with various amounts of dSPE sorbents (1, 3, and 5 mg) and evaluated with regard to matrix effects. In addition, analogously prepared blank samples were also analyzed to assess how the dSPE sorbents affected the analytical response under conditions where the matrix was absent. In matrix-matched experiments, the C_{18} sorbent was able to reduce ion suppression for almost all target analytes, whereas PSA and Strata-X-A produced contradictory results. Namely, PSA significantly impaired the analytical signal response of NSAIDs and other acidic APIs. This effect was even more pronounced in spiked procedural blanks, because the concentration of matrix substances was negligible, thus more binding sites were unoccupied, leading to an even greater affinity towards acidic APIs. Since WWTP effluents contain much less organic matter than untreated influents, unintentional loss of analytes may occur in WWTP effluents and lead to underestimated results. Meanwhile, the same amount of Strata-X-A sorbent impaired the analytical response of acidic analytes to a lesser extent but still provided a sound capability for the reduction of anionic surfactants (see Fig. S5A) in the Supplementary material). These results are likely to be related to the structural characteristics of each sorbent. PSA contains two types of binding sites (primary and secondary amine), while Strata-X-A has only one type of quaternary amine moiety, therefore, the first exhibits higher overall ion-exchange capacity. Based on these findings, PSA was excluded, while the amounts of C_{18} and Strata-X-A sorbents used in the final method were optimized to 3 mg and 1 mg, respectively.

Finally, dilution experiments were conducted to find the most appropriate dilution factor (1, 2, 4, 8, 12, 16, and 20). As seen in Fig. S6 (in the supplementary material), the final extract had to be diluted with injection phase at least 12 times before stable $ME_{ionization}$ was obtained. At these conditions, the mean $ME_{ionization}$ was 80.7 ± 24.5%, and the lowest $ME_{ionization}$ values was observed for paracetamol (19%), naproxen (38%), and spiramycin (58%).

The developed method was further verified by determining the absolute recoveries of APIs (Fig. 2A). The results revealed that dSPE treatment still caused a significant loss of analytes, especially acidic APIs. For instance, valsartan recovery decreased from 76% to 24% as a result of the clean-up procedure. Even though the absolute recovery was lower compared to the extracts obtained by $MgSO_4$-assisted phase separation alone, the dSPE clean-up step was able to sufficiently reduce the matrix effect (Fig. 2B) and improved the mass accuracy (Fig. 2C and D), allowing the extracts to be analyzed by DI-HRMS. Furthermore, the application of dSPE allowed to gain more information about the samples and the total amount of full-MS signals increased by 11% and 6% for negative and positive ionization modes, respectively. Nevertheless, the matrix-matched standard calibration curve (MMSCC) approach had to be applied to compensate for analyte losses and to improve the method precision.

3.2. Optimization of DI-HRMS method

HRMS parameters were optimized to achieve adequate sensitivity, signal stability, and mass accuracy. Apart from the source parameters, the following instrumental factors were recognized as the most influential: flow rate, ion accumulation time, resolving power, and injection phase. Special attention was paid to the injection phase, because its composition plays a crucial role in mitigating the matrix effect and can drastically affect the ionization efficiency. First, experiments were conducted on procedural blanks to find the most suitable acetonitrile percentage (100%, 95%, 87%, 75%, 50%, 25%, and 0%) in the injection phase (Fig. S7). After that, the most promising compositions were tested on matrix-matched samples to observe whether the ionization behavior remained unchanged in the presence of matrix components. The results indicated that the ionization efficiency increased along with the concentration of acetonitrile in the injection phase. Meanwhile, the experiments with the highest acetonitrile content (100% and 95%) produced a sudden drop of analytical response (with the exception of xylazine). This observation was in accordance with the hypothesis that hydrogen bonds play a crucial role in the formation of solvent shells that are necessary for the solvation in ESI process. Therefore, pure aprotic polar solvents produced poor results when the amount of water was insufficient (Kruve, 2016). Interestingly, the results from spiked procedure blanks and matrix-matched samples showed an equivalent trend and in both cases acetonitrile concentrations at 87% and 75% displayed the maximum sensitivity for almost all compounds, hence 80% acetonitrile in water (v/v) was selected as the injection phase.

The flow rate was optimized over the range from 1 to 20 μL/min. As expected, lower flow rates produced higher sensitivity. At the same time, poor repeatability was observed when the flow rate was below 4 μL/min. The observed instability can be attributed to the instrumental configuration, because the applied ESI source was not designed for infusion experiments at ultra-low flow rates, causing irregularities during ionization. According to the specification, the equipped sprayer needle is not recommended for experiments at flow rates below 2 μL/min. These considerations were taken into account and the flow rate for the main method was set at 5 μL/min. The ion accumulation time was investigated by analyzing a fortified WWTP influent sample. For FT-ICR-MS system, prolonged accumulation of ions can cause ICR cell oversaturation and negatively impact both sensitivity and mass accuracy. The final accumulation times were set at 100 ms and 20 ms for positive and negative ionization modes, respectively. The latter was maintained at 20 ms to reduce the impact of anionic surfactants on the full-MS spectra acquisition. Meanwhile, for MS^2 experiments, the accumulation time was increased to 500 ms per scan, which was sufficient for obtaining fragment ion traces at the first calibration level. The CE values were optimized during MS^2 CID experiments by obtaining fragment spectra for the target compounds at 5, 15, 25, 35, and 45 V. The obtained MS^2 spectra were matched against the developed in-house database to find the most suitable CE values that provided the most comprehensive fragmentation patterns for all target compounds. The results revealed that a single CE value failed to provide sufficient MS^2 information. Thus, the acquisition of fragment spectra in the final method was carried out using three CE values (5 V, 15 V, and 25 V) per precursor.

Finally, the transient size, which is directly linked to the resolving power, was deliberately kept at 4 M. The authors are well aware that this value is considerably lower than would be expected from a 7 T FT-ICR-MS system. The rationale behind this decision was to show that the developed method was not necessarily limited to FT-ICR-MS and could be transferred to other HRMS systems (e.g., Orbitrap-MS analyzers).

3.3. Performance of the target and suspect screening methods

In order to investigate the capability of the method and to assign proper concentration levels for the validation, an initial assessment of sensitivity was carried out by analyzing the target analytes at various fortification levels via the MMSCC approach. Decision limit (CCα) values were calculated based on the obtained calibration curve and corresponding signal-to-noise ratios (S/N). During this step, we increased the typical S/N threshold, since background noise produced by the FT-ICR-MS instrument was not negligible, especially for direct infusion analysis. According to Decision 2002/657/EC, for substances that have not had a permitted limit established, CCα values can be calculated at S/N of 3 (The European Parlament and the Council of the European Union, 2012). Yet, for FT-ICR-MS instrumentation, excessively low S/N threshold can be a major cause of poor reproducibility (Sleighter et al., 2012), hence the S/N limit for the peak-picking algorithm was increased to 5. Furthermore, all full-MS data were recorded in profile mode to enable the estimation of S/N during the post-processing stage. Taking into account that MS^2 spectra were acquired only for those compounds for which both ions were detected within the acceptable mass accuracy range (±1.25 ppm), CCα was therefore defined as the lowest concentration at which the least abundant analyte signal (Q_2) could be

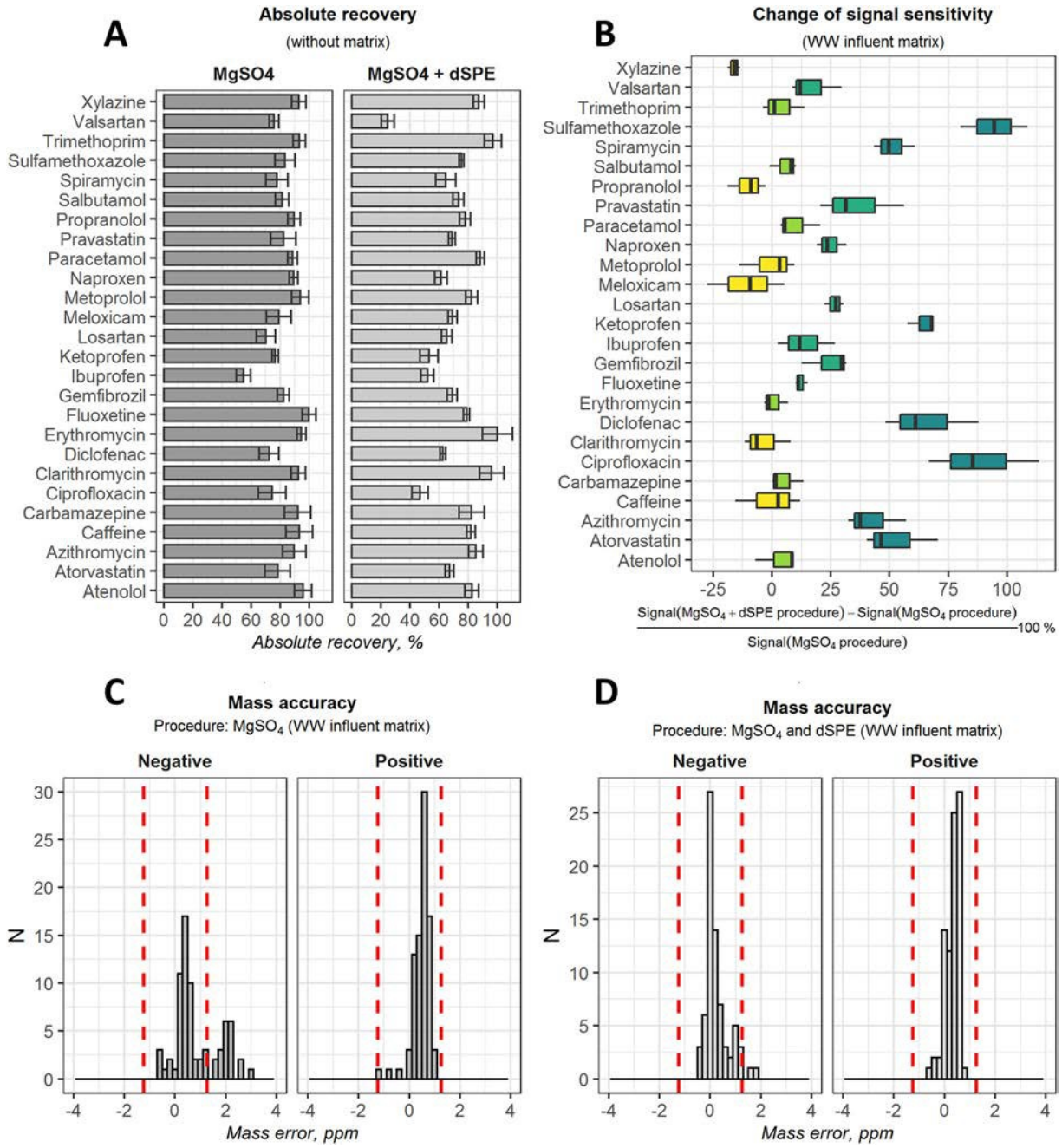


Fig. 2. Performance of the sample preparation protocol in terms of absolute recoveries (A), change of signal intensity (B), and mass accuracy (C and D) for targeted APIs.

measured at S/N of 5 instead of 3. CCβ was calculated as follows: CCα value plus 1.64 times the standard deviation of the within-laboratory reproducibility that was obtained from the measured content at the lowest validation level. Two ranges were used for the maximum permitted tolerance with regard to the ion abundance ratios (Q_2/Q_1). To reduce the number of false positives, suspect screening relied on a "strict" range that was set at 20% for all compounds. Meanwhile, a "wide" range was used only in target screening and its limits were directly adapted from Decision 2002/657/EC (≥0.5–20%, 0.2 to 0.5–25%, 0.1 to 0.2–30%, and ≤0.1–50%).

As mentioned in the experimental part, MS^2 matching was carried out using experimental library spectra and *in-silico* generated spectra. The latter option was justified by numerous studies indicating that *in-silico* fragmentation can be a complementary technique for compounds lacking experimental data. For instance, Blaženović et al. (2017) compared four different *in-silico* fragmentation algorithms (MetFragCL, CFM-ID, MAGMa+, and MS-FINDER) on a dataset from the 2016 CASMI challenge. According to this study, a workflow that relied solely on experimental spectra yielded 60% correct hits. Meanwhile, a workflow that was supplemented with *in-silico* generated spectra was able to achieve an overall success rate of 87% (Blaženović et al., 2017). The aforementioned findings indicate that a suspect screening workflow that combines both experimental and *in-silico* generated MS^2 spectra can achieve significantly higher accuracy. The workflow in our method relied on three MS^2 spectra that were acquired at three CE levels. The measured MS^2 features were matched against the experimental and predicted MS^2 fingerprints. If at least one fragment trace within the mass accuracy limits (±2.5 ppm) was detected in experimental library spectra, the suspect was reported as tentatively identified. Meanwhile, the generated fragment features were used to increase confidence,

because they were able to provide product ions for compounds that lacked comprehensive experimental data.

The target approach was validated using an in-house validation approach. A summary of the validation criteria is presented in Table 1. A detailed list of all investigated performance criteria is available in the Supplementary material (Tables S3–S8). The screening detection limit (SDL) was established as the lowest level for which the most abundant ion (Q_1) was detected in all parallel measurements. The limit of identification (LOI) was set as the lowest level for which a compound fulfilled all identification criteria at the "wide" range witha success rate of ≥90% (Diaz et al., 2013). From the perspective of quantitative analysis, the method can be considered sensitive and reliable only for a limited number of analytes, because the obtained performance criteria were explic- itly compound-specific. For example, atenolol, metoprolol, and propranolol (beta-blockers) were detected with high sensitivity and satisfactory performance was achieved even at concentrations below 50 ng/L. On the contrary, low molecular weight APIs such as paraceta- mol and naproxen required 10 times higher concentrations to meet the same criteria.

The ion abundance ratio was recognized as the most critical factor for successful identification of APIs. For example, at the lowest fortification level (level A) only 17 out of 26 target compounds met this criterion (with ≥90% success rate). Moreover, a slight bias towards negative residual error was observed for the experimentally determined ratios. This observation indicated that low abundance signals were occasionally rejected by the peak-picking algorithm because they failed to achieve the required S/N ratio. Thus, the accumulated full-MS spectra displayed systematically distorted isotopic patterns at low concentration levels. At the same time, the measured fragmentation features were in a good agreement with both experimental (MoNA) and predicted (CFM-ID algorithm) MS^2 fingerprints. The total number of correct hits was slightly higher for experimental fragment features (94%) than for the predicted fragmentation patterns (83%). On average, 2.3 and 1.6 fragments were found for each compound in the experimental and *in-silico* generated libraries, respectively. Only two substances could not be identified based on the predicted spectra (clarithromycin and spiramycin). As expected, the experimental database showed higher success rate, yet the *in-silico* generated fragmentation patterns were shown to be moderately accurate and thus could be used as a complementary tool for identification.

Altogether, 73% of compounds were successfully identified at a concentration close to CCβ while 89% and 94% success rates were observed for the two upper levels (Table 2). Only for 4 APIs (azithromycin, spiramycin, sulfamethoxazole, and valsartan) the LOI could not be established due to insufficient success rate. In order to verify the target screening method, a quality control (QC) sample was analyzed for each sample batch (10 samples). The QC sample was obtained by fortifying WWTP influent and effluent aliquots at a concentration that corresponded to the second validation level. Regrettably, the QC results were only partially consistent with the validation data and indicated that the method suffers from interferences to a greater extent than previously thought. In particular, the maximum recovery during the validation study was almost invariably below 150%. Meanwhile, the QC data displayed elevated recovery values on multiple occasions (14 from the total of 208 QC data points, Fig. S8). This finding suggested that the robustness was not properly assessed during the validation study, because all experiments (including MMSCC) were carried out on pooled WWTP influent and effluent matrixes, thus the inconsistency of matrix effects among diverse samples was not truly taken into account. Since the QC samples were obtained from different WWTP influents and effluents, MMSCC calibration curves could not completely mimic the matrix effect, causing potential errors in the quantification. However, the overall detection frequency of APIs during real WW sample analysis did not show statistically significant ($p > 0.05$) relationship with any of the main WW indicator parameters - biochemical oxygen demand, chemi- cal oxygen demand, and suspended solids (see Fig. S9 in the Supplementary material). Therefore, it may be presumed that the matrix effect did not significantly reduce the qualitative screening potential of the method.

Even highly selective HRMS applications, which involve chromatographic separation prior to MS detection, may often suffer from false positives. Therefore, careful attention must be paid to eliminate these risks, especially when the interpretation of results depends solely on MS data. First, procedure blanks were used for blank subtraction to remove background peaks. Next, we analyzed extracts of ten different matrices (e.g., fruits, vegetables, and non-contaminated commodities of animal origin), which presumably should not contain detectable amounts of APIs or their TPs. The samples were processed according to the workflow. A total of 39 suspects, including one target compound – gemfibrozil ($C_{15}H_{22}O_3$), displayed a false positive match by this approach and were excluded from the data analysis. Among the false positive hits, 74% belonged to suspects containing only C, H, and O atoms, while additional 23% of the false positives consisted of C, H, O, and N. These findings were in accordance with the general observation that the absence of heteroatoms other than nitrogen and oxygen gives far more false positive hits in HRMS-based screening methods (Kunzelmann et al., 2018).

Furthermore, five WW samples that were found to contain the highest number of suspects, were additionally investigated by HPLC-HRMS using the same sample preparation protocol (without the final dilution). The HPLC method was adapted from our previous work (Pugajeva et al., 2017), while simultaneous acquisition of full-MS and fragment spectra was achieved by broadband collision-induced dissociation (bbCID) technique on the same FT-ICR-MS instrument in both polarities at the lower transient size of 1 M. Ion chromatograms were constructed using the *m/z* values registered for Q_1 ions and suspects that produced multiple chromatographic peaks, indicating the presence of an interference. It must be noted that the HPLC analysis was not able

Table 1
The main validation criteria and instrumental capabilities in terms of compliance rates that correspond to mass accuracy, ion ratio, and MS^2 fingerprints.

Parameter	Conc. Level	Mean value	Median value	Range
Main validation criteria				
CCα, ng/L	–	128.4	65.5	18–693
CCβ, ng/L	–	234.2	98.0	27–1234
Coefficient of determination	–	0.983	0.98	0.95–0.99
RSD_{Sr}, %	A	16	16	5–38
	B	13	12	3–30
	C	12	10	5–29
RSD_{SwR}, %	A	23	21	8–61
	B	24	23	6–53
	C	20	17	7–51
Recovery, %	A	105.3	108	73–138
	B	101.1	102	75–121
	C	95.0	96	75–116
Mass accuracy and ion ratios				
Q_1 compliance rate (mass error ± 1.25 ppm), %	A	97%	100%	75%–100%
	B	98%	100%	75%–100%
	C	100%	100%	92%–100%
Q_2 compliance rate (mass error ± 1.25 ppm), %	A	85%	100%	25%–100%
	B	96%	100%	67%–100%
	C	99%	100%	83%–100%
Q_2/Q_1 ratio compliance rate (Q_2/Q_1 error ± 20%)	A	77%	96%	17%–100%
	B	90%	100%	33%–100%
	C	92%	100%	50%–100%
MS^2 fingerprinting				
Experimental MS^2 fingerprint match (≥1 matched fragment, MoNA database)	A	87%	92%	33%–100%
	B	96%	100%	75%–100%
	C	100%	100%	100%
Predicted MS^2 fingerprint match (≥1 matched fragment, CFM-ID)	A	73%	92%	0%–100%
	B	83%	100%	0%–100%
	C	92%	100%	0%–100%

Table 2
Successful identification rates of target APIs using "strict" and "wide" identification thresholds.

Validation level	Q_1 detection rate (±1.25 ppm)	Overall ID rate (strict)	Overall ID rate (wide)	≥75% ID rate (wide)	≥90% ID rate (wide)	100% ID rate (wide)	Compounds that failed to meet the specified identification criteria (success rate ≤ 90%, wide range)
Level A	97.1%	70.8%	73.4%	15/26	14/26	11/26	Azithromycin, caffeine, fluoxetine, gemfibrozil, ibuprofen, ketoprofen, meloxicam, naproxen, paracetamol, pravastatin, spiramycin, sulfamethoxazole, and valsartan
Level B	98.4%	87.2%	88.5%	22/26	17/26	16/26	Azithromycin, gemfibrozil, meloxicam, paracetamol, pravastatin, spiramycin, sulfamethoxazole, and valsartan
Level C	99.7%	91.7%	94.2%	25/26	22/26	18/26	Azithromycin, spiramycin, sulfamethoxazole, and valsartan

to detect all the expected suspect features. On average, 27% of the features that were found by direct infusion analysis remained undetected by the HPLC approach, indicating that this approach can only yield partial estimation of interferences.

Finally, a possible source of interfering ion species was more closely examined for those suspects detected at least twice in the WW samples. Two public databases (Human Metabolome Database (HMDB) and Chemical Entities of Biological Interest (ChEBI)) were surveyed to extract information about molecular formula isomers, whereas isobaric species (mass accuracy range: ±2.5 ppm; elemental composition: C, H, N, and O; electron configuration: even; H/C ratio range: 0 to 3) were calculated using the MolWeightToFormula (v. 3.1., Bruker) software. The obtained data were evaluated and the compounds with a high likelihood of an interference were dismissed from further analysis or classified as moderately susceptible to false-positive. If the database search yielded a potential interference with the same molecular formula, that could occur in WW samples due to anthropogenic or natural causes (e.g., simple peptides, human excretion products/metabolites, food and cosmetic ingredients, etc.), then moderate confidence level was assigned to the suspect. Meanwhile, compounds that yielded only one chromatographic peak in LC-MS analysis, exhibiting MS^2 features that could not be matched against the reference fragmentation pattern, were discarded. If multiple peaks were detected and at least one of them was compatible in terms of MS^2 spectra, the suspect was not excluded. A moderate confidence level was also assigned to compounds that yielded ≥2 isobaric substances when using the MolWeightToFormula software within the set criteria. From 72 suspects that were detected at least twice in the samples, 24 were excluded due to concerns regarding high risk of false positive results, whereas 28 compounds were classified as moderately susceptible to false positive results (moderate confidence level). The latter classification was introduced to highlight the method limitations, but did not change the interpretation of the final results.

A similar examination was carried out for all target analytes. Moderate confidence level was given to 15 target APIs, while one compound was discarded from the targeted method. Specifically, a suspiciously large peak was found in the extracted ion chromatograms of gemfibrozil in 4 out of 5 samples. The corresponding MS^2 spectra had an uncharacteristic pattern, even though one fragment feature matched the database. Therefore, a fortified QC sample was analyzed by HPLC-FT-ICR-MS to obtain the actual retention time of this compound. A mismatch between the retention times was observed, indicating that a high detection frequency of gemfibrozil is possibly due to an isomeric interference. Moreover, the results of database query showed that gemfibrozil shares the same molecular formula of $C_{15}H_{22}O_3$ with octyl salicylate, a frequent ingredient of sunscreens and cosmetics. The measured MS/MS spectrum was compared with several fragmentation patterns of octyl salicylate from MoNA database and significantly higher resemblance was found for octyl salicylate than for the target compound. These results were in accordance with the conclusions derived from the analysis of non-wastewater samples, where a false positive match was also found for gemfibrozil.

Overall, a coherent relationship was observed that a higher number of heteroatoms in the parent molecule, especially halogens, renders it less prone to interferences and at the same time results in a more distinguishable isotopic pattern. Thus, a more reliable detection via DI-HRMS can be achieved.

3.4. Target results

In our study, we analyzed 72 WW samples from 36 WWTPs in Latvia. Among the 25 target APIs, 20 and 24 different compounds were detected at least once in the WWTP effluents and influents, respectively. The most frequently detected compounds in untreated WWTP influents were diclofenac (NSAID, 86%), metoprolol (beta-blocker, 78%), clarithromycin (macrolide antibiotic, 53%), and ibuprofen (NSAID, 47%). A similar pattern was observed for WWTP effluents. However, compounds that show higher removal efficiency during municipal WW treatment processes were detected less frequently in the effluent samples. For instance, considerably high levels (446–2183 ng/L) of paracetamol were detected in six influent samples, whereas it remained undetected in effluents. Similarly, ibuprofen, salbutamol, and losartan were less frequently detected in effluents. Meanwhile, some APIs (e.g. diclofenac, erythromycin, clarithromycin, trimethoprim, and carbamazepine) did not follow this trend. Erythromycin was the only compound that was detected more frequently in treated WW samples. This observation is not particularly surprising, because poor removal of erythromycin has been repeatedly reported in the literature (Radjenović et al., 2009). Another key observation was the high prevalence of diclofenac in all sample types. This finding must be interpreted with caution as it may be associated with our detection method, because diclofenac is a particularly suitable analyte for HRMS-based applications as it bears two chlorine atoms. Thus, a definite and abundant isotopic pattern can be obtained, enabling higher sensitivity towards low abundance isotopologue signals. Nevertheless, the frequent occurrence of diclofenac is not unusual, because it exhibits highly variable removal efficiency in conventional WW treatment processes and is considered as a contaminant of emerging concern (Zhang et al., 2008). A different explanation could be related to the sample origin. Almost all WWTPs in our study receive sewage from rural villages and small towns. The average age in these locations is relatively high, with about 38% of the population being over 55 years old (Central Statistical Bureau of Latvia, 2019). Taking into account that diclofenac is administered not only for pain and inflammatory symptoms but also used for the treatment of arthrosis, which is a common condition among older populations, higher detection rates of this NSAID can be anticipated (Lindholm-Lehto et al., 2016). This explanation also supports the high prevalence of beta-blockers in WW samples.

As shown in Table 3, the concentration range of the detected APIs varied significantly. Some of this variability can be attributed to the instrumental capability, because sensitivity towards some analytes was insufficient (e.g. caffeine, valsartan, paracetamol, and ciprofloxacin), hence lower detection rates along with elevated mean concentrations were obtained. For example, caffeine is nearly ubiquitously present in domestic wastewaters due to the high consumption of beverages such as coffee, tea, and soluble dietary supplements. It can even be used as an anthropogenic marker for environmental pollution (Buerge et al.,

Table 3
The occurrence and detection frequencies of target APIs in WWTP influents and effluents.

Name	WWTP Effluents				WWTP Influents			
	Detection frequency, %	Mean, ng/L	Median, ng/L	Range (min - max), ng/L	Detection frequency, %	Mean, ng/L	Median, ng/L	Range (min - max), ng/L
Atenolol	36%	224	154	23–812	47%	250	59	16–1480
Atorvastatin	0%	ND[a]	ND	ND	19%	146	159	39–284
Azithromycin	17%	57	50	26–108	25%	117	127	24–231
Caffeine	11%	355	368	165–520	28%	6840	2030	152–43,900
Carbamazepine	22%	694	193	49–2170	28%	1600	262	111–8300
Ciprofloxacin	3%	108	108	108–108	3%	424	424	424–424
Clarithromycin	44%	379	269	64–1350	53%	453	152	17–2830
Diclofenac	78%	217	198	23–570	86%	570	253	27–3160
Erythromycin	19%	85	97	31–124	14%	98	68	51–196
Fluoxetine	6%	14	14	12–16	22%	12	12	9–16
Ibuprofen	17%	371	244	152–1070	47%	3320	658	108–28,500
Ketoprofen	8%	1590	521	511–3730	22%	1730	733	377–9090
Losartan	6%	56	56	34–79	22%	242	89	46–1170
Meloxicam	3%	153	153	153–153	25%	107	24	13–658
Metoprolol	61%	168	96	5–735	78%	470	231	10–5520
Naproxen	14%	611	570	306–961	19%	1980	775	350–4280
Paracetamol	0%	ND	ND	ND	17%	1190	1280	446–2180
Pravastatin	19%	188	176	119–274	33%	335	360	41–535
Propranolol	31%	10	8	5–23	44%	13	8	4–46
Salbutamol	3%	31	31	31–31	22%	39	35	17–84
Spiramycin	0%	ND	ND	ND	0%	ND	ND	ND
Sulfamethoxazole	19%	214	119	60–676	25%	1500	321	88–11,200
Trimethoprim	25%	52	49	12–95	28%	248	59	21–1650
Valsartan	0%	ND	ND	ND	22%	299	233	133–660
Xylazine	0%	ND	ND	ND	11%	13	13	7–20

[a] ND – not detected;

2003). However, in our study, we were able to quantitatively determine caffeine in only 28% and 11% of the WWTP influent and effluent samples, respectively. In general, our results did not show many atypical features and were in accordance with those reported from other European countries (Bodík et al., 2016; Kot-Wasik et al., 2016; Kötke et al., 2019; López-Serna et al., 2012). To further verify the findings of this study, we assessed the relationship between the obtained detection frequencies and national consumption of the individual pharmaceutical compounds. Consumption data were obtained from the annual report (data from 2018) by the State Agency of Medicines of Latvia and expressed as defined daily dose (DDD) per 1000 inhabitants per day (State Agency of Medicines of the Republic of Latvia, 2019). For six target analytes, information was not presented in the annual report. Fig. 3A reveals that, in most cases, APIs that are consumed more frequently show higher detection rates in WWTP influents. Nevertheless, a disproportionately high occurrence can be seen for two beta-blockers (atenolol and propranolol), while atorvastatin, which had the highest DDD per 1000 inhabitants per day, was detected in only 19% of WWTP influent samples. The latter can be attributed to the fact that atorvastatin undergoes rapid metabolism and thus only a minor fraction of the administered dose is excreted in unchanged form (Lennernäs, 2003).

Furthermore, an assessment of sulfamethoxazole/trimethoprim ratio was performed for samples that contained both of these APIs. There is some evidence that the concentration ratio between them can be used as a marker to determine the main origin of WW samples, because there is a difference between human and veterinary dosage forms. In the case of humans, the ratio between them in the formulation is 5:1, while the typical ratio in WW samples ranges from 1.1 to 3.3 (Thiebault, 2020). In our study, we were able to observe a consistent re- lationship only in WWTP influents where the mean ratio was 3.92 (Fig. 3B). Meanwhile, the WWTP effluents showed highly scattered ra- tios possibly due to different removal efficiencies of both antibiotics. Considering that all of the analyzed samples were from municipal WWTPs, our findings are in line with those of previous studies. Therefore, despite the semi-quantitative nature of the method, this study offers evidence that DI-HRMS can be applied for a rapid screening of APIs in WW samples (an illustrative example of carbamazepine with the measured full-MS and MS^2 features is presented in Fig. S10).

3.5. Suspect screening results

In total, 79 compounds from the suspect list were detected in the analyzed WW samples. Among these analytes, 54 APIs and 18 transformation products (TPs) were found in the WWTP influent samples, while 38 APIs and 16 TPs were detected in WWTP effluents. Calculated logD values (at pH 7.0) for detected APIs ranged from −5.69 (metformin) to 5.17 (telmisartan), indicating that the method can detect both lipophilic and hydrophilic compounds. The total recall rate from the suspect list was 15.1% (79/524). Out of all compliant signals that were obtained from full-MS data, 26% were discarded due to a mismatch between the measured and theoretical MS^2 fingerprint. This observation indicated that full-MS data can be rather reliable even when used alone. The matching of fragment spectra provided an additional degree of safety, especially for compounds that are more susceptible to interferences. Experimental MS^2 features that were obtained from the MoNA database again provided more comprehensive information and higher number of matched fragments, compared to the *in-silico* generated MS^2 fingerprints. Nevertheless, it is also likely that some suspects were falsely rejected/identified due to inaccurate or insufficiently characteristic MS^2 fingerprints. The relatively low prevalence of TPs in WW samples can be attributed to the content of the suspect list, because it contained only 71 different TPs, thus it is not surprising that these substances were detected less often.

Overall, the number of detected compounds varied to great extent among the samples and was largely influenced by the load of organic components and the origin of WW. As shown in Fig. 4B, the highest number of suspects was observed in samples from a WWTP that receives sewage from urbanized areas and health care institutions. On the contrary, lower prevalence of APIs was found in WW samples from sparsely populated areas such as villages (see Table S1 for more detailed information). Furthermore, the analysis of individual contributions from APIs belonging to the most common therapeutic groups in

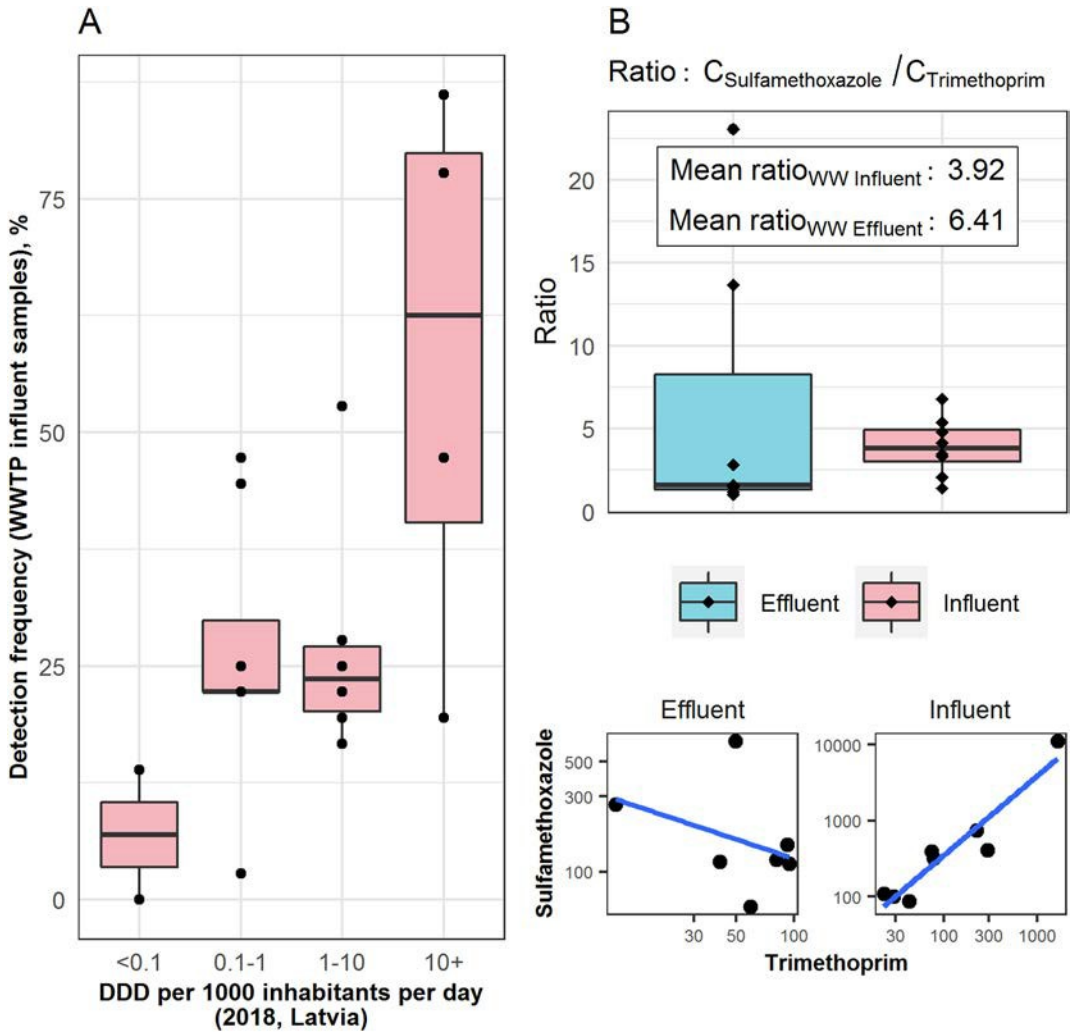


Fig. 3. Detection frequency of selected APIs in WWTP influent samples (A) and the ratio between sulfamethoxazole and trimethoprim concentrations that were found in the analyzed WWTP influents and effluents.

WWTP influent samples confirmed that the method can yield reliable results that are in accordance with the expected WW composition (Fig. 4A). For instance, WW discharged from a mental health clinic (sample "N") showed a considerably higher prevalence of antidepressants, anticonvulsants, and psychiatric medications. Meanwhile, samples of non-specific origin yielded a higher incidence of analgesics, antibiotics, and beta-blockers.

Among all of the suspect APIs, telmisartan had the highest detection rate. It was found in 83% and 92% of the WWTP influent and effluent samples, respectively. This result may be explained by three reasons: (i) it is the most prescribed member among angiotensin II receptor blockers in Latvia, (ii) telmisartan is excreted almost entirely as an unchanged drug and (iii) it has a relatively long elimination half-life (Stangier et al., 2000). No interfering peaks were found for telmisartan signal during the LC-HRMS analysis and the obtained spectra from bbCID fragmentation experiments were also in a good agreement with the experimental library spectra. Besides, other studies have also identified telmisartan as a problematic compound that exhibits poor removal efficiencies and a high prevalence in both sewage and surface water samples (Giebułtowicz et al., 2016; Mijangos et al., 2018). Other APIs that were repeatedly detected in WW samples are listed in Table 4. Among them, several APIs with high consumption volumes (bisoprolol, metformin, and rosuvastatin) and substances that are predominantly excreted as parent compound (e.g., amisulpride, sulpiride, and citalopram) were repeatedly detected. In fact, amisulpride and sulpiride (antipsychotics) have been identified as pseudo-persistent pollutants. In addition, lamotrigine (anticonvulsant) has shown to have higher concentrations in treated WW samples compared to WWTP influents, because of the deconjugation of the main human metabolite - lamotrigine-N2-glucuronide (Bollmann et al., 2016). Our data supported this interpretation and higher Q_1 signal intensities were indeed found in WWTP effluents (Fig. S11A). Meanwhile, the high detection rates of tramadol and oxcarbazepine should be interpreted with caution. Although both APIs have been frequently reported by other studies, they are not well suited for our methodology. The oxcarbazepine signal can overlap with two major carbamazepine metabolites (carbamazepine-10,11-epoxide and 2-hydroxycarbamazepine) that share the same molecular formula and, due to the high consumption of the parent compound, are co-occurring in WW samples. The same is true for tramadol, the detection of which can suffer from interfering signals that are caused by venlafaxine metabolites (*N*- and *O*-desmethylvenlafaxine). Nevertheless, we did not discard those analytes, because LC-HRMS analysis revealed that both parent APIs and interfering TPs are present in WW samples.

Another remarkable finding was that dextrorphan, an active metabolite of dextromethorphan, was detected suspiciously often (in 42% of the WWTP influent samples). Even though the parent compound is marketed as a cough suppressant, it is also abused as a recreational drug due to its side-effects. To avoid the misuse of dextromethorphan, it has been classified as a prescription drug in Latvia since 2009 and,

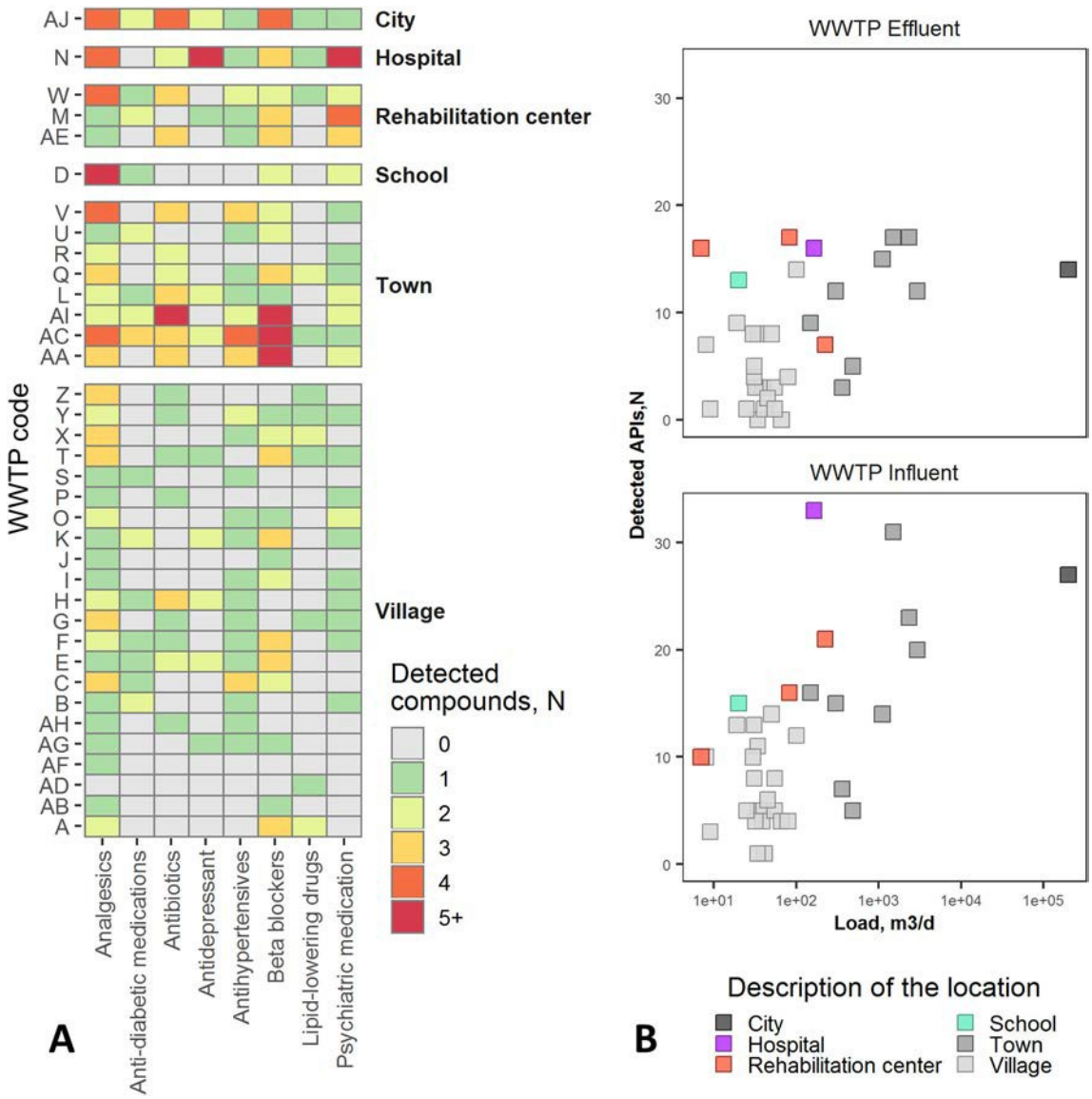


Fig. 4. The API profiles measured in WWTP influent samples (A) and the overall prevalence of APIs among WWTP influents and effluents (sum of all compounds obtained from the suspect screening and targeted approach).

as a consequence, its annual consumption has declined substantially and thus lower prevalence is expected. These results may indicate that the actual use of dextromethorphan is higher than reflected by the official consumption statistics and, despite legislative measures, it is still available for illicit use. Previous studies have also documented the presence of dextromethorphan metabolites in the environment (Campos-Mañas et al., 2019; Thurman and Ferrer, 2012). Other notable TPs that were detected during our study were hydroxydiclofenac, carboxyibuprofen, didesmethylvenlafaxine, desmethyltramadol, and carbamazepine metabolites. Among these compounds, "high" identification confidence was assigned only for hydroxydiclofenac. When compared to the parent molecule, traces of hydroxydiclofenac were found less often and in most cases with much lower Q_1 signal intensities. It has been shown that the ratio between the parent drug and hydroxyl- ated metabolites is dependent on the route of administration. Topical applications yield significantly higher levels of the parent substance than hydroxylated metabolites (Stülten et al., 2008). Our results seem to agree with this hypothesis, because topical diclofenac formulations (e.g., patches, creams, and gels) are available over the counter in Latvia, while almost all oral dosage forms can be purchased only by prescription.

As mentioned in the previous section, atorvastatin is the most widely used API among all prescription drugs. Yet, atorvastatin is extensively metabolised and only a small fraction is excreted unchanged (it was detected only in 19% of the WWTP influent samples via the targeted approach). Therefore, a retrospective analysis was conducted by investigating the presence of main atorvastatin metabolites (atorvastatin lactone, hydroxyatorvastatin, hydroxyatorvastatin lactone, and hydroxyatorvastatin glucuronide), which were not a part of the suspect database. Full-MS data were treated using isolation criteria that were identical to the main method and the most abundant Q_1 ion of deprotonated and protonated species was used for the comparison of signal intensities in the negative and positive ionization modes, respectively. Hydroxyatorvastatin was the only metabolite detected in WW samples. As shown in Fig. S11B, the analytical response was higher for the metabolite compared to the parent drug. These data were in agreement with the findings by Langford and Thomas (2011), who showed that *para*- and *ortho*-hydroxyatorvastatin were detected at higher levels than atorvastatin in Norwegian aquatic environment (Langford and Thomas, 2011). However, the comparison of signal intensities should be interpreted with caution, since the ionization of each compound can vary, giving misleading results about the actual occurrence of

Table 4
A summary of APIs and TPs from the suspect list that were detected most frequently in WWTP influents and effluents.

APIs

Name	Molecular formula	Detection frequency[a], %		MS^2 features[b]	Conf.[c]	Possible interferences	DID[d]
		Influent	Effluent				
Telmisartan	$C_{33}H_{30}N_4O_2$	83 (47)	92 (56)	1.7 (1.5)	H		5.5
Bisoprolol	$C_{18}H_{31}NO_4$	53 (36)	28 (22)	1.1 (1.0)	M	Nitrooctadecadienoic acids	19.2
Tramadol	$C_{16}H_{25}NO_2$	36 (36)	39 (36)	1.4 (1.4)	M	N/*O*-Desmethylvenlafaxine	1.5
Amisulpride	$C_{17}H_{27}N_3O_4S$	28 (28)	17 (17)	3.2 (3.2)	H		0.4
Oxcarbazepine	$C_{15}H_{12}N_2O_2$	28 (19)	14 (8)	2.3 (2.9)	M	Carbamazepine metabolites	0.4
Citalopram / Escitalopram	$C_{20}H_{21}FN_2O$	25 (3)	11 (0)	1.2 (1.0)	H		0.8/4.6
Sulpiride	$C_{15}H_{23}N_3O_4S$	25 (25)	25 (25)	1.9 (2.9)	H		0.2
Metformin	$C_4H_{11}N_5$	25 (25)	6 (6)	2.5 (1.5)	H		16.5
Rosuvastatin	$C_{22}H_{28}FN_3O_6S$	19 (14)	3 (3)	1.4 (1.6)	M	Isobaric substance $C_{33}H_{23}NO_3$	32.4
Sitagliptin	$C_{16}H_{15}F_6N_5O$	19 (22)	17 (19)	2.2 (1.5)	H		0.7
Lamotrigine	$C_9H_7Cl_2N_5$	17 (33)	19 (33)	2.1 (1.5)	H		0.8
Propafenone	$C_{21}H_{27}NO_3$	17 (17)	11 (8)	2.1 (1.1)	H		1.1

Transformation products of APIs

Name	Molecular formula	Detection frequency[a], %		MS^2 features[b]	Conf.[c]	Possible interferences	Parent compound (corresponding DID[d])
		Influent	Effluent				
Dextrorphan	$C_{17}H_{23}NO$	42 (0)	28 (0)	3.3 (0)	H		Dextromethorphan (0.1)
Hydroxydiclofenac	$C_{14}H_{11}Cl_2NO_3$	42 (39)	33 (28)	1.7 (1.7)	H		Diclofenac (18.1)
Carboxyibuprofen	$C_{13}H_{16}O_4$	36 (25)	28 (28)	1.1 (1.2)	M	Ethyl vanillin isobutyrate	Ibuprofen (23.7)
N,*N*-Didesmethylvenlafaxine	$C_{15}H_{23}NO_2$	36 (31)	39 (33)	1.1 (1.3)	M	*O*-Desmethyltramadol	Venlafaxine (1.0)
N,*O*-Didesmethylvenlafaxine	$C_{15}H_{23}NO_2$	36 (31)	42 (36)	1.3 (1.1)	M	*O*-Desmethyltramadol	Venlafaxine (1.0)
O-Desmethyltramadol	$C_{15}H_{23}NO_2$	36 (31)	39 (36)	1.2 (1)	M	Didesmethylvenlafaxine	Tramadol (1.5)
O-Desmethylvenlafaxine	$C_{16}H_{25}NO_2$	36 (39)	36 (33)	2 (1.3)	M	Tramadol	Venlafaxine (1.0)
N-Desmethylvenlafaxine	$C_{16}H_{25}NO_2$	33 (33)	39 (36)	1.2 (1.3)	M	Tramadol	Venlafaxine (1.0)
Acetoaminoantipyrine	$C_{13}H_{15}N_3O_2$	31 (31)	8 (3)	2.4 (1.5)	H		Metamizole (0.6)
Carbamazepine-2OH	$C_{15}H_{12}N_2O_2$	25 (22)	14 (8)	3.4 (3.3)	M	Oxcarbazepine	Carbamazepine (2.1)
Carbamazepine-10,11-epoxide	$C_{15}H_{12}N_2O_2$	22 (19)	8 (8)	2.8 (2.3)	M	Oxcarbazepine	Carbamazepine (2.1)

[a] Detection frequency obtained from the identification using experimental and predicted (in parentheses) MS^2 features.
[b] Mean number of fragment signals matched per precursor using experimental and predicted (in parentheses) MS^2 features.
[c] Identification confidence level ("H"- high and "M" - moderate).
[d] DDD per 1000 inhabitants (2018, Latvia).

these analytes. A similar retrospective analysis was carried out for carbamazepine and its metabolites (see Fig. S11C). The observed metabolite patterns showed a coherent trend. The analytical response for those metabolites increased along with increasing carbamazepine concentration. Besides, the most intense signals were found for samples AE, M, and N, that received sewage from healthcare facilities (AE and M - rehabilitation centre, N - hospital).

Generally, lower detection rates were observed for APIs on the suspect list, because the majority of suspect signals were dismissed due to non-compliant Q_2/Q_1 ratios. As mentioned before, target analytes were qualified using a "wide" ratio tolerance that ranged from 20% to 50%, whereas a strict tolerance limit (20%) was maintained for the qualification of suspects. In this way we were able to increase the reliability of the method, but only at the expense of sensitivity, resulting in lower detection frequencies that presumably narrowed the scope of the method. Nevertheless, the acquired results are comparable to screening methodologies based on LC-HRMS, thus showing that high-throughput screening of WW samples via DI-HRMS is possible and can yield valuable information regarding the presence of APIs and their TPs.

4. Conclusions

In this study, a novel suspect and target screening methodology using DI-FT-ICR-MS was developed. It was designed for semi-quantitative analysis of 26 compounds and qualitative screening of more than 500 APIs and their TPs. A total of 79 suspects and 24 target compounds were detected in 72 WW samples. The results indicate that the applied resolving power (490,000 at *m/z* 250 and 275,000 at *m/z* 500) was sufficient for screening purposes and thus the method was not limited to FT-ICR-MS systems, but could also be adapted to new generation Orbitrap mass analyzers. The main disadvantages of the method were: (i) higher susceptibility to false positive hits, especially for compounds that have numerous constitutional isomers (e.g., *N*- and *O*-desmethylvenlafaxine), (ii) poorer sensitivity compared to the conventional LC-HRMS applications, and (iii) limited capability for accurate quantification. However, it also provided unique advantages. First of all, it was rapid and required lesser amount of laborious post-processing steps that are an essential part of most LC-HRMS screening methods (e.g., peak picking and integration, retention time alignment, linking full-MS data with MS^2 traces). The interpretation of results relied solely on MS data; thus, the method could be more easily automated and data processing could be performed almost instantaneously, without much assistance from the operator. Second, it was possible to cover a potentially wider scope of APIs, as both hydrophobic and hydrophilic APIs could be analyzed simultaneously, as the limitations set by conventional reversed-phase chromatography were avoided. Thus, DI-HRMS can be considered a tool to observe previously overlooked compounds and possibly extend the scope of analytical procedures that are aimed to study environmental pollutants of emerging concern. Overall, the results of this study indicated that the method was very suitable for screening substances that contained at least two different heteroatoms. Even higher selectivity could be achieved towards halogenated APIs that can produce distinct isotopic patterns, such as diclofenac and losartan.

CRediT authorship contribution statement

Ingus Perkons: Conceptualization, Methodology, Software, Validation, Investigation, Visualization, Writing - original draft. Janis Rusko: Methodology, Validation, Formal analysis, Writing - review & editing,

Visualization. Dzintars Zacs: Formal analysis, Resources, Writing - review & editing, Supervision. Vadims Bartkevics: Conceptualization, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

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.

Acknowledgments

This work was supported by the Latvian National Research Program under project No. VPP-COVID-2020/1-0008.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.142688.

References

Andersson, A., Harir, M., Gonsior, M., Hertkorn, N., Schmitt-Kopplin, P., Kylin, H., Karlsson, S., Ashiq, M.J., Lavonen, E., Nilsson, K., Pettersson, Ä., Stavklint, H., Bastviken, D., 2019. Waterworks-specific composition of drinking water disinfection by-products. Environ. Sci. Water Res. Technol. 5, 861–872. https://doi.org/10.1039/c9ew00034h.

aus der Beek, T., Weber, F.A., Bergmann, A., Hickmann, S., Ebert, I., Hein, A., Küster, A., 2016. Pharmaceuticals in the environment—Global occurrences and perspectives. Environ. Toxicol. Chem. 35, 823–835. https://doi.org/10.1002/etc.3339.

Blaženović, I., Kind, T., Torbašinović, H., Obrenović, S., Mehta, S.S., Tsugawa, H., Wermuth, T., Schauer, N., Jahn, M., Biedendieck, R., Jahn, D., Fiehn, O., 2017. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: database boosting is needed to achieve 93% accuracy. J. Cheminform. 9, 1–12. https://doi.org/10.1186/s13321-017-0219-x.

Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. TrAC - Trends Anal. Chem. 66, 32–44. https://doi.org/10.1016/j.trac.2014.11.009.

Bodík, I., Mackuľak, T., Fáberová, M., Ivanová, L., 2016. Occurrence of illicit drugs and selected pharmaceuticals in Slovak municipal wastewater. Environ. Sci. Pollut. Res. 23, 21098–21105. https://doi.org/10.1007/s11356-016-7415-5.

Bollmann, A.F., Seitz, W., Prasse, C., Lucke, T., Schulz, W., Ternes, T., 2016. Occurrence and fate of amisulpride, sulpiride, and lamotrigine in municipal wastewater treatment plants with biological treatment and ozonation. J. Hazard. Mater. 320, 204–215. https://doi.org/10.1016/j.jhazmat.2016.08.022.

Boulard, L., Dierkes, G., Ternes, T., 2018. Utilization of large volume zwitterionic hydrophilic interaction liquid chromatography for the analysis of polar pharmaceuticals in aqueous environmental samples: benefits and limitations. J. Chromatogr. A 1535, 27–43. https://doi.org/10.1016/j.chroma.2017.12.023.

Bound, J.P., Voulvoulis, N., 2005. Household disposal of pharmaceuticals as a pathway for aquatic contamination in the United Kingdom. Environ. Health Perspect. 113, 1705–1711. https://doi.org/10.1289/ehp.8315.

Buerge, I.J., Poiger, T., Müller, M.D., Buser, H.R., 2003. Caffeine, an anthropogenic marker for wastewater contamination of surface waters. Environ. Sci. Technol. 37, 691–700. https://doi.org/10.1021/es020125z.

Campos-Mañas, M.C., Ferrer, I., Thurman, E.M., Sánchez Pérez, J.A., Agüera, A., 2019. Identification of opioids in surface and wastewaters by LC/QTOF-MS using retrospective data analysis. Sci. Total Environ. 664, 874–884. https://doi.org/10.1016/j.scitotenv.2019.01.389.

Central Statistical Bureau of Latvia, 2019. RIG010. The permanent resident population by sex and age groups in the statistical regions, cities under state jurisdiction, counties, towns, parishes, villages and Riga neighbourhoods (based on the boundaries in force at the beginning of 2020). [WWW Document]. URL. https://data1.csb.gov.lv/pxweb/en/iedz/iedz riga/RIG010.px/. (Accessed 2 March 2020).

Chao, A., Al-Ghoul, H., McEachran, A.D., Balabin, I., Transue, T., Cathey, T., Grossman, J.N., Singh, R.R., Ulrich, E.M., Williams, A.J., Sobus, J.R., 2020. In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples. Anal. Bioanal. Chem. 412, 1303–1315. https://doi.org/10.1007/s00216-019-02351-7.

Council of Europe, 2020. European Pharmacopoeia. (Ph. Eur.). 10th Edition: The 10th Edition at a Glance. [WWW Document]. URL. https://www.edqm.eu/en/european-pharmacopoeia-ph-eur-10th-edition. (Accessed 11 May 2020).

Daughton, C.G., Ternes, T.A., 1999. Pharmaceuticals and personal care products in the environment: agents of subtle change? Environ. Health Perspect. 107, 907–938. https://doi.org/10.1289/ehp.99107s6907.

Deloitte, 2018. Options for a Strategic Approach to Pharmaceuticals in the Environment - Final Report. https://doi.org/10.2779/87838.

Desbiolles, F., Malleret, L., Tiliacos, C., Wong-Wah-Chung, P., Laffont-Schwob, I., 2018. Occurrence and ecotoxicological assessment of pharmaceuticals: is there a risk for the Mediterranean aquatic environment? Sci. Total Environ. 639, 1334–1348. https://doi.org/10.1016/j.scitotenv.2018.04.351.

Diaz, R., Ibáñez, M., Sancho, J.V., Hernández, F., 2013. Qualitative validation of a liquid chromatography–quadrupole-time of flight mass spectrometry screening method for organic pollutants in waters. J. Chromatogr. A 1276, 47–57. https://doi.org/10.1016/j.chroma.2012.12.030.

Ebele, A.J., Abou-Elwafa Abdallah, M., Harrad, S., 2017. Pharmaceuticals and personal care products (PPCPs) in the freshwater aquatic environment. Emerg. Contam. 3, 1–16. https://doi.org/10.1016/j.emcon.2016.12.004.

Escher, B.I., Stapleton, H.M., Schymanski, E.L., 2020. Tracking complex mixtures of chemicals in our changing environment. Science 367 (6476), 388–392. https://doi.org/10.1126/science.aay6636.

Gao, P., Ding, Y., Li, H., Xagoraraki, I., 2012. Occurrence of pharmaceuticals in a municipal wastewater treatment plant: mass balance and removal processes. Chemosphere 88, 17–24. https://doi.org/10.1016/j.chemosphere.2012.02.017.

Giebułtowicz, J., Stankiewicz, A., Wroczyński, P., Nałęcz-Jawecki, G., 2016. Occurrence of cardiovascular drugs in the sewage-impacted Vistula River and in tap water in the Warsaw region (Poland). Environ. Sci. Pollut. Res. 23, 24337–24349. https://doi.org/10.1007/s11356-016-7668-z.

Gros, M., Petrović, M., Barceló, D., 2006. Development of a multi-residue analytical methodology based on liquid chromatography-tandem mass spectrometry (LC-MS/MS) for screening and trace level determination of pharmaceuticals in surface and wastewaters. Talanta 70, 678–690. https://doi.org/10.1016/j.talanta.2006.05.024.

Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., Schmidt, T.C., 2020. Comparison of software tools for liquid chromatography-high-resolution mass spectrometry data processing in nontarget screening of environmental samples. Anal. Chem. 92, 1898–1907. https://doi.org/10.1021/acs.analchem.9b04095.

Kötke, D., Gandrass, J., Xie, Z., Ebinghaus, R., 2019. Prioritised pharmaceuticals in German estuaries and coastal waters: occurrence and environmental risk assessment. Environ. Pollut. 255. https://doi.org/10.1016/j.envpol.2019.113161.

Kot-Wasik, A., Jakimska, A., Śliwka-Kaszyńska, M., 2016. Occurrence and seasonal variations of 25 pharmaceutical residues in wastewater and drinking water treatment plants. Environ. Monit. Assess. 188. https://doi.org/10.1007/s10661-016-5637-0.

Krauss, M., Singer, H., Hollender, J., 2010. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. Anal. Bioanal. Chem. 397, 943–951. https://doi.org/10.1007/s00216-010-3608-9.

Kruve, A., 2016. Influence of mobile phase, source parameters and source type on electrospray ionization efficiency in negative ion mode. J. Mass Spectrom. 51, 596–601. https://doi.org/10.1002/jms.3790.

Kunzelmann, M., Winter, M., Åberg, M., Hellenäs, K.E., Rosén, J., 2018. Non-targeted analysis of unexpected food contaminants using LC-HRMS. Anal. Bioanal. Chem. 410, 5593–5602. https://doi.org/10.1007/s00216-018-1028-4.

Langford, K., Thomas, K.V., 2011. Input of selected human pharmaceutical metabolites into the Norwegian aquatic environment. J. Environ. Monit. 13, 416–421. https://doi.org/10.1039/c0em00342e.

Lennernäs, H., 2003. Clinical pharmacokinetics of atorvastatin. Clin. Pharmacokinet. 42, 1141–1160. https://doi.org/10.2165/00003088-200342130-00005.

Lindholm-Lehto, P.C., Ahkola, H.S.J., Knuutinen, J.S., Herve, S.H., 2016. Widespread occurrence and seasonal variation of pharmaceuticals in surface waters and municipal wastewater treatment plants in Central Finland. Environ. Sci. Pollut. Res. 23, 7985–7997. https://doi.org/10.1007/s11356-015-5997-y.

Loos, M., Gerber, C., Corona, F., Hollender, J., Singer, H., 2015. Accelerated isotope fine structure calculation using pruned transition trees. Anal. Chem. 87, 5738–5744. https://doi.org/10.1021/acs.analchem.5b00941.

López-Serna, R., Petrović, M., Barceló, D., 2012. Occurrence and distribution of multi-class pharmaceuticals and their active metabolites and transformation products in the Ebro River basin (NE Spain). Sci. Total Environ. 440, 280–289. https://doi.org/10.1016/j.scitotenv.2012.06.027.

Mechelke, J., Longrée, P., Singer, H., Hollender, J., 2019. Vacuum-assisted evaporative concentration combined with LC-HRMS/MS for ultra-trace-level screening of organic micropollutants in environmental water samples. Anal. Bioanal. Chem. 411, 2555–2567. https://doi.org/10.1007/s00216-019-01696-3.

Mijangos, L., Ziarrusta, H., Ros, O., Kortazar, L., Fernández, L.A., Olivares, M., Zuloaga, O., Prieto, A., Etxebarria, N., 2018. Occurrence of emerging pollutants in estuaries of the Basque Country: analysis of sources and distribution, and assessment of the environmental risk. Water Res. 147, 152–163. https://doi.org/10.1016/j.watres.2018.09.033.

Montes, R., Aguirre, J., Vidal, X., Rodil, R., Cela, R., Quintana, J.B., 2017. Screening for polar chemicals in water by trifunctional mixed-mode liquid chromatography-high resolution mass spectrometry. Environ. Sci. Technol. 51, 6250–6259. https://doi.org/10.1021/acs.est.6b05135.

Nannou, C.I., Boti, V.I., Albanis, T.A., 2019. A modified QuEChERS approach for the analysis of pharmaceuticals in sediments by LC-Orbitrap HRMS. Anal. Bioanal. Chem. 411, 1383–1396. https://doi.org/10.1007/s00216-018-01570-8.

Peña-Herrera, J.M., Montemurro, N., Barceló, D., Pérez, S., 2019. Development and validation of an analytical method for determination of pharmaceuticals in fish muscle based on QuEChERS extraction and SWATH acquisition using LC-QTOF-MS/MS system. Talanta 199, 370–379. https://doi.org/10.1016/j.talanta.2019.01.119.

Pérez-Lemus, N., López-Serna, R., Pérez-Elvira, S.I., Barrado, E., 2019. Analytical methodologies for the determination of pharmaceuticals and personal care products (PPCPs) in sewage sludge: a critical review. Anal. Chim. Acta 1083, 19–40. https://doi.org/10.1016/j.aca.2019.06.044.

Phillips, P.J., Smith, S.G., Kolpin, D.W., Zaugg, S.D., Buxton, H.T., Furlong, E.T., Esposito, K., Stinson, B., 2010. Pharmaceutical formulation facilities as sources of opioids and

other pharmaceuticals to wastewater treatment plant effluents. Environ. Sci. Technol. 44, 4910–4916. https://doi.org/10.1021/es100356f.

Pugajeva, I., Rusko, J., Perkons, I., Lundanes, E., Bartkevics, V., 2017. Determination of pharmaceutical residues in wastewater using high performance liquid chromatography coupled to quadrupole-Orbitrap mass spectrometry. J. Pharm. Biomed. Anal. 133. https://doi.org/10.1016/j.jpba.2016.11.008.

Radjenović, J., Petrović, M., Barceló, D., 2009. Fate and distribution of pharmaceuticals in wastewater and sewage sludge of the conventional activated sludge (CAS) and advanced membrane bioreactor (MBR) treatment. Water Res. 43, 831–841. https://doi.org/10.1016/j.watres.2008.11.043.

Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J. Cheminform. 8, 1–16. https://doi.org/10.1186/s13321-016-0115-9.

Schultz, M.M., Furlong, E.T., Kolpin, D.W., Werner, S.L., Schoenfuss, H.L., Barber, L.B., Blazer, V.S., Norris, D.O., Vajda, A.M., 2010. Antidepressant pharmaceuticals in two U.S. effluent-impacted streams: occurrence and fate in water and sediment and selective uptake in fish neural tissue. Environ. Sci. Technol. 44, 1918–1925. https://doi.org/10.1021/es9022706.

Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I.M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., Thomaidis, N.S., Bletsou, A., Zwiener, C., Ibáñez, M., Portolés, T., De Boer, R., Reid, M.J., Onghena, M., Kunkel, U., Schulz, W., Guillon, A., Noyon, N., Leroy, G., Bados, P., Bogialli, S., Stipaničev, D., Rostkowski, P., Hollender, J., 2015. Non-target screening with high-resolution mass spectrometry: critical re- view using a collaborative trial on water analysis. Anal. Bioanal. Chem. 407, 6237–6255. https://doi.org/10.1007/s00216-015-8681-7.

Singer, H.P., Wössner, A.E., McArdell, C.S., Fenner, K., 2016. Rapid screening for exposure to "non-target" pharmaceuticals from wastewater effluents by combining HRMS-based suspect screening and exposure modeling. Environ. Sci. Technol. 50, 6698–6707. https://doi.org/10.1021/acs.est.5b03332.

Sleighter, R.L., Chen, H., Wozniak, A.S., Willoughby, A.S., Caricasole, P., Hatcher, P.G., 2012. Establishing a measure of reproducibility of ultrahigh-resolution mass spectra for complex mixtures of natural organic matter. Anal. Chem. 84, 9184–9191. https://doi.org/10.1021/ac3018026.

Stangier, J., Schmid, J., Türck, D., Switek, H., Verhagen, A., Peeters, P.A.M., van Marle, S.P., Tamminga, W.J., Sollie, F.A.E., Jonkman, J.H.G., 2000. Absorption, metabolism, and excretion of intravenously and orally administered [14C]telmisartan in healthy volunteers. J. Clin. Pharmacol. 40, 1312–1322. https://doi.org/10.1177/009127000004001202.

State Agency of Medicines of the Republic of Latvia, 2019. Statistics on Medicines Consumption 2018. [WWW Document]. URL. https://www.zva.gov.lv/sites/default/files/2019-06/2018.pdf.

Stülten, D., Zühlke, S., Lamshöft, M., Spiteller, M., 2008. Occurrence of diclofenac and selected metabolites in sewage effluents. Sci. Total Environ. 405, 310–316. https://doi.org/10.1016/j.scitotenv.2008.05.036.

Styrishave, B., Halling-Sørensen, B., Ingerslev, F., 2011. Environmental risk assessment of three selective serotonin reuptake inhibitors in the aquatic environment: a case study including a cocktail scenario. Environ. Toxicol. Chem. 30, 254–261. https://doi.org/10.1002/etc.372.

The European Parlament and the Council of the European Union, 2012. Commission decision of 14 august 2002 implementing council directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (notified under document number C(2002) 3044) (text with EEA relevance) (2002/657/EC). Off. J. Eur. Union 91, 8–36.

Thiebault, T., 2020. Sulfamethoxazole/trimethoprim ratio as a new marker in raw wastewaters: a critical review. Sci. Total Environ. 715, 136916. https://doi.org/10.1016/j.scitotenv.2020.136916.

Thurman, E.M., Ferrer, I., 2012. Liquid chromatography/quadrupole-time-of-flight mass spectrometry with metabolic profiling of human urine as a tool for environmental analysis of dextromethorphan. J. Chromatogr. A 1259, 158–166. https://doi.org/10.1016/j.chroma.2012.03.008.

Ulrich, E.M., Sobus, J.R., Grulke, C.M., Richard, A.M., Newton, S.R., Strynar, M.J., Mansouri, K., Williams, A.J., 2019. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Anal. Bioanal. Chem. 411, 853–866. https://doi.org/10.1007/s00216-018-1435-6.

Vasquez, M.I., Lambrianides, A., Schneider, M., Kümmerer, K., Fatta-Kassinos, D., 2014. Environmental side effects of pharmaceutical cocktails: what we know and what we should know. J. Hazard. Mater. 279, 169–189. https://doi.org/10.1016/j.jhazmat.2014.06.069.

Weinberger, J., Klaper, R., 2014. Environmental concentrations of the selective serotonin reuptake inhibitor fluoxetine impact specific behaviors involved in reproduction, feeding and predator avoidance in the fish Pimephales promelas (fathead minnow). Aquat. Toxicol. 151, 77–83. https://doi.org/10.1016/j.aquatox.2013.10.012.

Wielens Becker, R., Ibáñez, M., Cuervo Lumbaque, E., Wilde, M.L., Flores da Rosa, T., Hernández, F., Sirtori, C., 2020. Investigation of pharmaceuticals and their metabolites in Brazilian hospital wastewater by LC-QTOF MS screening combined with a preliminary exposure and in silico risk assessment. Sci. Total Environ. 699, 134218. https://doi.org/10.1016/j.scitotenv.2019.134218.

Zacs, D., Perkons, I., Bartkevics, V., 2019. Evaluation of analytical performance of gas chromatography coupled with atmospheric pressure chemical ionization Fourier transform ion cyclotron resonance mass spectrometry (GC-APCI-FT-ICR-MS) in the target and non-targeted analysis of brominated and chlorinated flame retardants in food. Chemosphere, 368–377 https://doi.org/10.1016/j.chemosphere.2019.03.047.

Zhang, Y., Geißen, S.U., Gal, C., 2008. Carbamazepine and diclofenac: removal in wastewater treatment plants and occurrence in water bodies. Chemosphere 73, 1151–1161. https://doi.org/10.1016/j.chemosphere.2008.07.086.

Ziegler, G., Gonsior, M., Fisher, D.J., Schmitt-Kopplin, P., Tamburri, M.N., 2019. Formation of brominated organic compounds and molecular transformations in dissolved organic matter (dom) after ballast water treatment with sodium dichloroisocyanurate dihydrate (DICD). Environ. Sci. Technol. 53, 8006–8016. https://doi.org/10.1021/acs.est.9b01064.