Use of Bayesian Modeling to Assess Occurrence of Viral Pathogens in Multiple US Drinking Water Systems

SUPPORTING INFORMATION

Eunice A. Varughese[*†], Nichole E. Brinkman[†], Emily M. Anneken[†], Jennifer L. Cashdollar[†], G. Shay Fout[†], Edward T. Furlong[‡], Dana W. Kolpin[§], Susan T. Glassmeyer[†], Scott P. Keely[†]

†USEPA, Office of Research and Development, National Exposure Research Laboratory, 26 W. Martin Luther King Dr., Cincinnati, OH 45268, United States; varughese.eunice@epa.gov; brinkman.nichole@epa.gov; anneken.emily@epa.gov; cashdollar.jennifer@epa.gov; fout.shay@epa.gov; glassmeyer.susan@epa.gov; keely.scott@epa.gov

‡USGS, National Water Quality Laboratory, Denver Federal Center, Bldg 95 Denver, CO 80225, United States; efurlong@usgs.gov

§ USGS, 400 S. Clinton St, Rm 269 Federal Building Iowa City, IA 52240, United States; dwkolpin@usgs.gov

* Corresponding author:

Email: varughese.eunice@epa.gov; Phone: (513) 569-7628; Fax: (513) 569-7117

Supplementary Methods:

Comparison of nucleic acid extraction techniques

After organic flocculate concentration, EPA Method 1615 uses a tertiary concentration step, followed by small volume nucleic acid extraction (0.2 ml) to isolate viral nucleic acid. In lieu of this two-step method, a one-step method of large volume nucleic extraction (0.5 ml) has been used by our lab previously to further concentrate the sample and extract viral nucleic acid (Brinkman et al., 2013). In this study, the two-step and one-step method were compared to determine which method produced better viral recovery.

The two-step method involved Vivaspin concentration followed by mini-extraction, where only 1 ml of organic flocculation concentrate is used to produce approximately 200 µL of concentrated nucleic acid. Briefly, Vivaspin 20 centrifugal concentrators (Sartorius-Stedim) with a 30K MWCO (Molecular Weight Cut Off) were pre-soaked overnight with 1× Dulbecco's PBS, 0.2% BSA (USB) to bind non-specific sites inside the Vivaspin concentrator. After removal of the solution, 10 ml of prepared blank organic flocculate was seeded with primary effluent concentrate as a source of virus. Primary effluent is wastewater that has been settled to remove the majority of suspended and floating solids. We further concentrated the primary effluent in the laboratory as described in previous work (Brinkman et al., 2013), and measured the concentrate for levels of the multiple virus types used in this study. The primary effluent was found to contain all 5 virus types (assessed by qPCR, data not shown) that were analyzed in this study. For each method, at least 3 replicates of 9 ml of sterile organic flocculate concentrates were seeded with 0.675 ml of primary effluent concentrate. The seeded organic flocculate was added to the Vivaspin and centrifuged at 5000 × g until the retained sample volume was about 0.5 ml. Afterwards, each Vivaspin was washed twice by adding 1 ml of 1×Dulbecco's PBS, followed by centrifugation at 5000 × g until the retained volume was less than 0.4 ml. The retained sample volume

693 was brought up to 0.4 ml with 1× Dulbecco's PBS, pH 7.0, of which 200 µl of this concentrate was
694 extracted using the QIAamp DNA Blood Mini Kit (Qiagen). Nucleic acids were eluted twice with 50 µl of
695 buffer AE and extracts were stored at −70°C, until processed by qPCR, as detailed below.

696 The one step method involved extraction of 10 ml of the organic flocculate seeded with primary effluent
697 concentrate, using only the QIAamp DNA Blood Maxi Extraction Kit (Qiagen), with slight modifications to
698 manufacturer's instructions. The lysis buffer provided in the kit was replaced with Buffer AVL (Qiagen),
699 as it is a better buffer for viral lysis. Additionally, carrier RNA (Qiagen) was supplemented at 275 µg per
700 extraction event to buffer against loss of targeted nucleic acids and the protease digestion step was
701 omitted. Finally, during elution of nucleic acids from the maxi column, 1 ml Buffer AE was
702 supplemented with 400 units of Recombinant RNasin Ribonuclease Inhibitor (Promega) and the volume
703 of the eluate was reloaded onto the column for a second elution, for a total extract volume of 1 ml.
704 Extracts were stored at −70°C, until processed by qPCR, as detailed below. The products of the two-step
705 and one-step extraction samples were analyzed by (RT)-qPCR as described in the manuscript.

706 **Whole Genome Amplification**

707 Whole genome amplification (WGA) was performed to identify samples that may be positive, but were
708 below the limit of qPCR quantification. The samples that were determined to be negative for AdV
709 and/or PyV DNA by qPCR, were subjected to WGA, using the Illustra Genome Phi V2 Amplification Kit
710 (GE Healthcare). In a microcentrifuge tube, 3 µl of sample extracts was gently mixed with 7 µl of sample
711 buffer provided with the kit, followed by heating to 95°C for 3 min. The samples were then incubated
712 on ice for at least 10 min. To each tube, 10 µL of a mixture containing 9 parts reaction buffer and 1-part
713 enzyme mix provided with the kit was added, gently mixed, and put back on ice. Tubes were then
714 incubated at 30°C overnight (approximately 18 h) in a water bath. Finally, inactivation of the
715 polymerase was done by heating the tube to 65°C for 10 min, then cooling to 4°C. Samples were stored

716 at -80°C. To assess the results of WGA, qPCR was performed on the products in 25 µl reactions, as
717 specified in the Methods section titled, RT-qPCR/qPCR for enteric viruses, except that 2 µl of WGA
718 product was added to each reaction.

719

Supplementary Results:

Improving virus recovery efficiency

Reducing the number of steps in sample processing and increasing the total volume of concentrated sample used in extraction has the potential to improve viral recovery efficiency. It was previously shown that celite concentration followed by a large volume (10 ml) extraction procedure, which provides a 10-fold concentration of the sample in addition to nucleic acid extraction, resulted in an increase of 1.4-4.3-fold in recovery efficiency for AdV, EV, and NoV compared to the two-step process of using Vivaspin tertiary concentration and small volume (0.2 ml) extraction procedure (Brinkman et al., 2013). Therefore, the large volume (1 ml) extraction procedure was examined with organic flocculation concentrates seeded with primary effluent to determine if the same increase in viral recovery efficiency was observed in organic flocculation concentrates as in celite concentrates. The use of the large volume extraction procedure allowed for a statistically significant increase in percent recovery for this step, for all 5 virus groups examined (Supplemental Fig. A1) ($p<0.05$; two-way ANOVA) with a mean increase of 2.1-fold relative to the two-step process of tertiary concentration and small volume extraction. Hence, a method that deviated slightly from EPA Method 1615 was used for all DWTP samples in this study, where the large volume nucleic acid extraction replaced tertiary concentration and small volume extraction procedures.

Whole Genome Amplification reveals false negatives by qPCR

As shown in Table 2, source and treated water samples were tested for DNA (AdV and PyV) and RNA viruses (NoV GI, NoV GII, and EV) with qPCR and RT-qPCR assays, respectively. As predicted by IT analysis (Table 3), the potential for false negatives was greater in undiluted samples. Whole genome amplification (WGA) has been used to increase target DNA/RNA concentrations prior to downstream molecular techniques (e.g., deep sequencing) and might be efficacious for the detection of virus in

743 environmental samples. Therefore, WGA was carried out for all samples that were negative in the initial
744 qPCR reactions. For RNA viruses, WGA was not successful after an independent reverse transcriptase
745 step, despite many repeated attempts. We hypothesize this may be due to the sensitivity of reverse
746 transcriptase to inhibitory substances in the water samples that prevented production of cDNA in the
747 samples prior to exponential amplification. Of the samples we analyzed, we found 6 samples that were
748 previously negative by qPCR, to be positive by WGA (Supplemental Table A3). Of these 6 samples, 4
749 were source water samples and 1 was a treated water sample. These data suggest that there are low
750 quantities of AdV and PyV genomic DNA in the water concentrates, at or below the limit of detection for
751 qPCR. WGA, therefore, offers the potential of detecting virus that may be in the samples, but below
752 detection levels using qPCR, thus improving sensitivity. Moreover, these results emphasize the need to
753 model the rate of false negatives to accurately determine viral levels from environmental concentrates.

754 **Table A1:** Primers and probes used for (RT)-qPCR.

Virus target	Primers/probes	Sequence 5' to 3' Probe: 6FAM - TAMRA	Reference
Enterovirus	Forward primer	CCC TGA ATG CGG CTA AT	Brinkman et al., 2013
	Reverse primer	TGT CAC CAT AAG CAG CCA	
	Probe	ACG GAC ACC CAA AGT AGT CGG TTC	
Norovirus GI	Forward primer	CGC TGG ATG CGN TTC CAT	Butot et al., 2010
	Reverse primer	CCT TAG ACG CCA TCA TCA TTT AC	
	Probe	TGG ACA GGA GAY CGC RAT CT	
Norovirus GII	Forward primer	ATG TTC AGR TGG ATG AGR TTC TCW GA	Butot et al., 2010
	Reverse primer	TCG ACG CCA TCT TCA TTC ACA	
	Probe	AGC ACG TGG GAG GGC GAT CG	
Adenovirus	Forward primer	GGA CGC CTC GGA GTA CCT GAG	Jothikumar et al., 2005
	Reverse primer	ACI GTG GGG TTT CTG AAC TTG TT	
	Probe (MGM quencher)	CTG GTG CAG TTC GCC CGT GCC A	
Human polyomavirus	Forward primer	AGT CTT TAG GGT CTT CTA CCT TT	McQuaig et al., 2006; 2009
	Reverse primer	GGT GCC AAC CTA TGG AAC AG	
	Probe	TCA TCA CTG GCA AAC AT	
Hepatitis G – EAC Inhibition Test	Forward primer	CGG CCA AAA GGT GGT GGA TG	Schlueter et al., 1996
	Reverse primer	CGA CGA GCC TGA CGT CGG G	
	Probe	AGG TCC CTC TGG CGC TTG TGG CGA G	

Mixed bases in degenerate primers and probes: Y = C/T; R = A/G; W = A/T; I = deoxyinosine; N = any

755

756

757 **Table A2:** Symbols used for math equations and BUGS code

Symbol in Materials and Methods	Symbol or Value in BUGS Code	Occurrence in Equation Number	Description of Symbol
λ_i	u_i	1,2,3,4,13,15	The average concentration of virus genome molecules per ml of sample; indexed with *i*
μ	mu	1	A parameter of the lognormal distribution of enteric virus
τ	tau	1	A parameter of the lognormal distribution of enteric virus
$p_{(EAC(s)+)ij}$ $p_{(EAC(c)+)ij}$ $p_{(virus+)ij}$	ps_{ij} pc_{ij} p_{ij}	4,5,6,7	The Poisson probabilities for undiluted and diluted EAC spike; EAC control; and enteric viruses; all indexed with *ij*
$x\text{-}EAC_{(s)ii}$ $x\text{-}EAC_{(c)ij}$ x_{ij}	pcr_s_{ij}, pcr_c_{ij}, pcr_{ij}	5,7	The PCR count data for undiluted and diluted EAC spike; EAC control; and enteric viruses; all indexed with *ij*
n_{ij}	n_{ij}	5,7	the number of PCR replicates per dilution; indexed with *ij*
d_j	d_j	3,4,6	the number of PCR dilutions for enteric virus or EAC control and spike; indexed with *j*
v	v	3,4,6	Volume of sample added to each PCR reaction in mL; a constant v for enteric virus PCR was 0.005 and for EAC (control and spike) it was 0.0045
f	f	13,15	The total volume (3 ml) of the nucleic acid extract that could be derived from 30 ml of organic flocculate
ε_i	$error_i$	8,13	PCR error; indexed by i
λ_{ci}	uc_i	6,8	The average concentration of control EAC genomes per PCR reaction; indexed with *i*
λ_{si}	us_i	6,8	The average concentration of spiked EAC genomes per PCR reaction; indexed with *i*
ω_i	$loss_i$	9,13,15	The beta distribution of virus loss due to method interference; indexed with *i*
α	a	9,10,11,12	Shape parameter of the beta distribution; Estimated from eq. 11
β	b	9,10,11,12	Shape parameter of the beta distribution; Estimated from eq. 12
c	Calculated from Adenovirus spike	10,11,12	Mean percentage of virus loss
σ^2	Calculated from Adenovirus spike	10	Variance of virus loss

$\frac{vg_i}{L_i}$	u_liter$_i$	13,15	Posterior mean of virus concentration per Liter (*L*) of water; used for Model 1 and Model 2; indexed with *i*
θ_i	re$_i$	14,15	The beta distribution of PCR inhibition using control and spike EAC; indexed with *i*
$\alpha 1$	a1	9,10,11,12	Shape parameter of the beta distribution; the first and second values were used for PCR-negative source and treated samples, respectively
$\beta 1$	b1	9,10,11,12	Shape parameter of the beta distribution; Estimated by using eq. 12; the first and second values were used for PCR-negative source and treated samples, respectively.

758

759

Table A3. Water concentrates positive for AdV and PyV using Whole Genome Amplification (WGA). All samples processed for WGA were PCR negative for AdV & PyV virus in the initial screening.

	DNA Virus	
Water Source	**AdV**	**PyV**
Source Water		
DWTP 11		+
DWTP 24		+
DWTP 26	+	
DWTP 27	+	+
Treated Water		
DWTP 20		+

764 **Table A4.** Deviance Information for Model 1 and Model 2

Model Number, Source (S) or Treated(T) DWTP samples[#]	Deviance Information*			
	Dbar	Dhat	DIC	pD
A) Sensitivity analysis				
Model 1: ST PCR positive data (μ=-2.6)	467.7	440.6	494.7	27.02
Model 1: ST PCR positive data (μ=-0.26)	467.7	440.7	494.7	27.01
Model 1: ST PCR positive data (μ=-0.026)	467.7	440.7	494.7	27.01
B) Model Comparisons				
Model 1: S, PCR positive data	419.1	397.8	440.5	21.32
Model 2: S, PCR positive data	419.7	397.7	441.7	21.96

765 # μ is from Equation 1. Source (S) or Treated (T) DWTP samples.
766 *Dbar, Dhat, DIC and pD indicate how well the models fits the data. DBar is the posterior mean of the
767 deviance, Dhat is a point estimate of the deviance, DIC is the Deviance Information Criterion, and pD is a
768 measure of model complexity.
769

770

Figure A1. Comparison of one-step maxi extraction with two-step vivaspin and maxi-extraction of 5 different virus types. Graph shows differences in percent recovery of nucleic acid from each type of method among the 5 virus types. AdV = adenovirus; EV= enterovirus; NoV GI = norovirus GI; NoV GII = norovirus GII; PyV = polyomavirus. Bars represent standard error of 3 replicate samples. ($p<0.05$; two-way ANOVA)

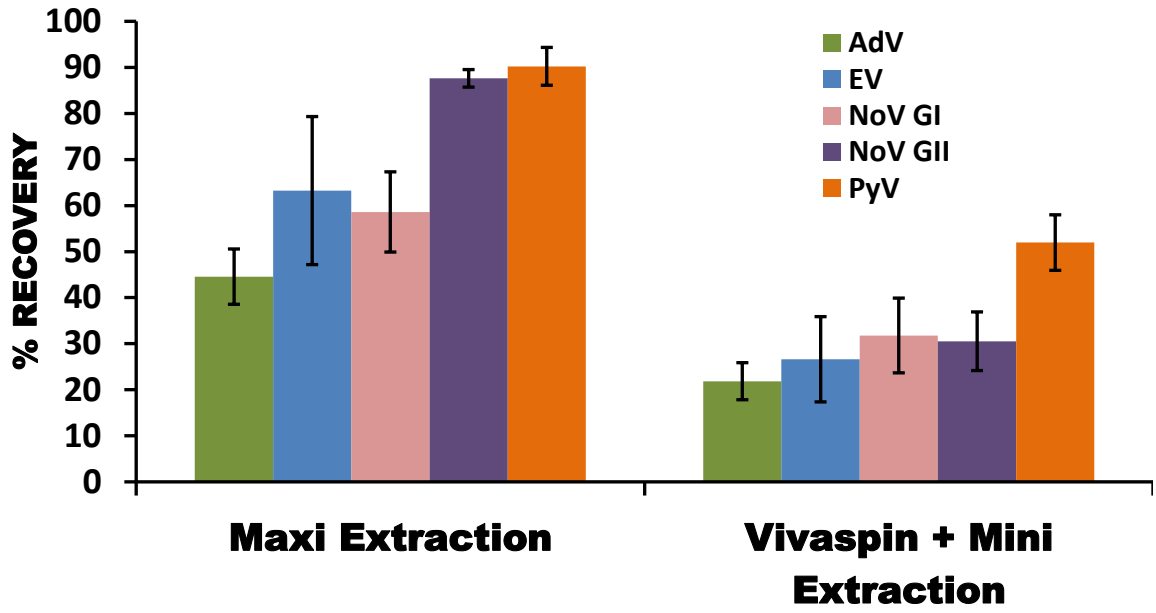

Figure A2. Bayesian credible interval plots of EAC (exogenous amplification control) spiked in samples from various DWTPs. Numbers represent DWTP identification. S and T suffixes represent source and treated water samples, respectively. Bars represent 95% credible intervals.

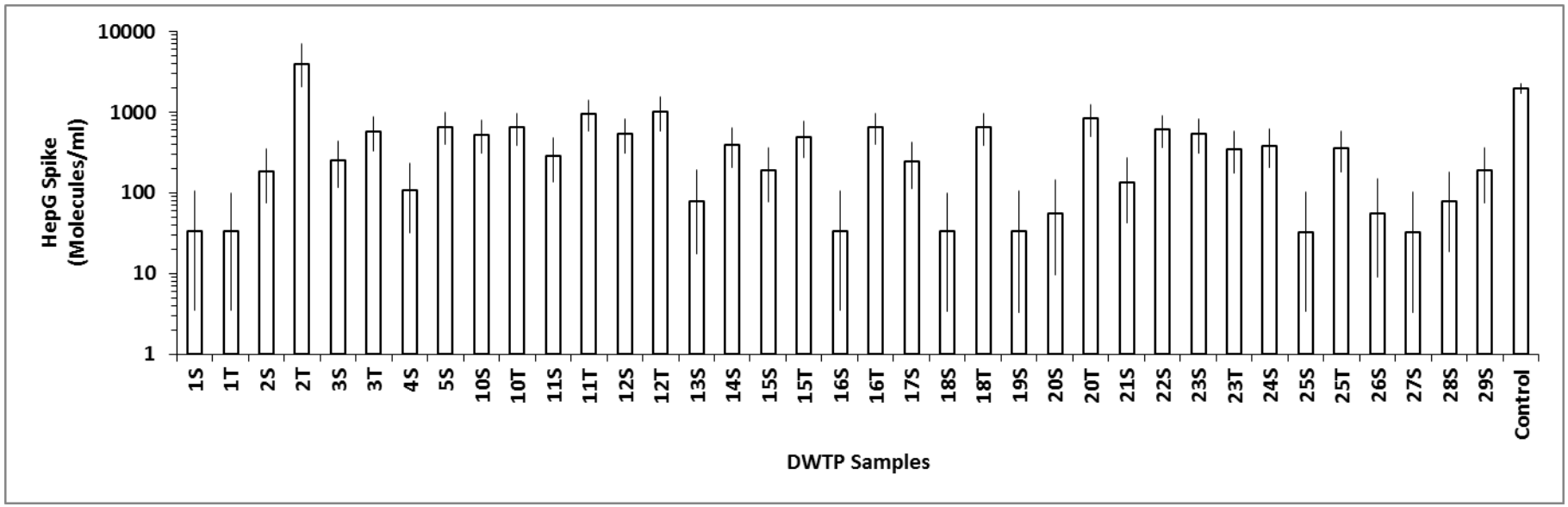

Figure A3. Comparison of the Brooks-Gelman-Rubin (BGR) statistic for 5 different virus types in DWTP 3S. AdV = adenovirus; EV= enterovirus; NoV GI = norovirus GI; NoV GII = norovirus GII; PyV = polyomavirus. The BGR ratio was determined for the $\frac{vg_i}{L_i}$ parameter in Equation 13.