Evaluation of the protection against norovirus afforded by *E. coli* monitoring of shellfish production areas under EU regulations

A. D. Younger, M. Teixeira Alves, N. G. H. Taylor, J. Lowther,

C. Baker-Austin, C. J. A. Campos, M. Price-Hayward and D. Lees

ABSTRACT

EC Regulation 854/2004 requires the classification of bivalve mollusc harvesting areas according to the faecal pollution status of sites. It has been reported that determination of *Escherichia coli* in bivalve shellfish is a poor predictor of norovirus (NoV) contamination in individual samples. We explore the correlation of shellfish *E. coli* data with norovirus presence using data from studies across 88 UK sites (1,184 paired samples). We investigate whether current *E. coli* legislative standards could be refined to reduce NoV infection risk. A significant relationship between *E. coli* and NoV was found in the winter months (October to February) using data from sites with at least 10 data pairs (51 sites). We found that the ratio of arithmetic means ($\log_{10}$ *E. coli* to $\log_{10}$ NoV) at these sites ranged from 0.6 to 1.4. The lower ratios (towards 0.6) might typically indicate situations where the contribution from UV disinfected sewage discharges was more significant. Conversely, higher ratios (towards 1.4) might indicate a prevalence of animal sources of pollution; however, this relationship did not always hold true and so further work is required to fully elucidate the factors of relevance. Reducing the current class B maximum (allowed in 10% of samples) from 46,000 *E. coli* per 100 g (corresponding NoV value of 75750 ± 103) to 18,000 *E. coli* per 100 g (corresponding NoV value of 29365 ± 69) reduces maximum levels of NoV by a factor of 2.6 to 1; reducing the upper class B limit to 100% compliance with 4,600 *E. coli* per 100 g (corresponding NoV value of 7403 ± 39) reduces maximum levels of NoV by a factor of 10.2 to 1. We found using the UK filtered winter dataset that a maximum of 200 NoV corresponded to a maximum of 128 ± 7 *E. coli* per 100 g. A maximum of 1,000 NoV corresponded to a maximum of 631 ± 14 *E. coli* per 100 g.

Key words | classification, *E. coli*, faecal pollution, oysters, norovirus, regulations

A. D. Younger (corresponding author)
M. Teixeira Alves
N. G. H. Taylor
J. Lowther
C. Baker-Austin
C. J. A. Campos
M. Price-Hayward
D. Lees
Centre for Fisheries, Environment and Aquaculture Science,
Barrack Road, The Nothe, Weymouth, DT4 8UB,
United Kingdom
E-mail: andrew.younger@cefas.co.uk

INTRODUCTION

Production areas for bivalve molluscan shellfish (oysters, mussels, cockles and clams) are required to be classified according to their sanitary quality under EU Regulation 854/2004 on the basis of *Escherichia coli* monitoring (European Communities 2004). The classification is a public health measure and determines the sanitary quality of the production areas and the extent of processing required before shellfish can be placed on the market for human consumption. The classification categories are A, B, and C, with class A being the 'cleanest'. Class A shellfish require no treatment prior to consumption. Class B shellfish require treatment (typically purifying or relaying) whereas class C shellfish require intensive treatment (typically relaying for a longer period or heat processing by an approved method). Production areas with levels of contamination greater than class C cannot be placed on the market and may be designated as Prohibited. Areas are classified based on routine (normally monthly) monitoring of shellfish from representative monitoring points.

Shellfish harvesting areas in England and Wales may be impacted by sewage from both continuous and intermittent sewage discharges (Campos *et al.* 2017a, 2017b). With relatively constant temperature, salinity and food supply, bivalves can process up to 20 L of seawater per hour (Galtsoff 1964) during filter-feeding and pathogens from sewage may also be accumulated. Studies have suggested that *E. coli* may be concentrated by up to 100 times the level found in the growing waters (Kay *et al.* 2008). From a public health perspective, bivalve shellfish can represent a health risk as they tend to be eaten raw (particularly in the case of oysters) or lightly cooked. The published literature documents that outbreaks of disease can occur on a large scale, e.g. around 300,000 infected with hepatitis A from clams in China in 1988 (Xu *et al.* 1992). Whilst hepatitis outbreaks do still occur, the most common infection associated with shellfish consumption in the developed world is currently norovirus (NoV) (Potasman *et al.* 2002; Bellou *et al.* 2013).

Current EU food regulations (EC Regulation 854/2004) do not specify limits for NoV since, until recently, suitable methods have not been available. However, a standardised methodology for quantification of NoV in shellfish has been recently published (ISO 2017) and implemented in many laboratories across Europe. This method uses real-time polymerase chain reaction (RT-PCR) to amplify and detect target sequences within the viral RNA. However, for risk assessment, a significant issue is that this method does not differentiate between infective and non-infective virus particles which could potentially lead to overestimation of risk (EFSA 2012; Hartard *et al.* 2017). Nonetheless, this method has been considered suitable for use within a legislative context by the European Food Safety Authority (EFSA 2012). Previous UK studies have assessed NoV levels in commercial production areas through analysis of >800 samples (Lowther *et al.* 2012). This study concluded that although individual *E. coli* results were poorly predictive of norovirus risk, average *E. coli* levels at a site correlated well with average norovirus levels, particularly in the winter months. In addition, significant differences were found in norovirus levels in class A, B and C sites, supporting the current classification approach. Finally, the authors noted that class B is a very broad category of potential risk accommodating *E. coli* levels up to 46,000 MPN/100 g shellfish (in 10% of samples). In considering NoV risk reduction, an EURL options paper (EURL 2012) suggested that the current class B 10% tolerance upper limit of 46,000 *E. coli*/100 g could be reduced or removed altogether for higher risk species (e.g. oysters). The impact of possible refinements, such as reducing the *E. coli* upper tolerance limits, or determining the maximum *E. coli* levels equivalent to specified possible limits for NoV (EFSA 2012), are explored in this paper.

It is well recognised that no faecal indicator is perfect and each demonstrates shortcomings (Wu *et al.* 2011). Wu *et al.* (2011) also note that results suggest that much of the controversy with regard to indicator and pathogen correlations is the result of studies with insufficient data for assessing correlations. They also add that the most important factors in determining correlations between indicator–pathogen pairs were the sample size and the number of samples positive for pathogens. For this reason, this paper compares shellfish *E. coli* monitoring data against norovirus data obtained from previously published studies (including those assessed by Lowther *et al.* (2012)) across 88 sites (1,184 paired samples) in the UK. It uses a large dataset with a significant number of positive samples (78.4% positive for NoV; 82.8% positive for *E. coli*), thereby addressing one of the shortcomings identified by Wu *et al.* (2011). Further to the Lowther *et al.* (2012) study, our study assesses the level of protection from NoV provided by the current *E. coli*-based standards and investigates whether *E. coli* data can be used to more reliably predict the NoV risk at individual sites. It also considers whether the current legislative standards could be refined to improve public health protection from the potential NoV infection risk associated with consumption of contaminated bivalve shellfish.

METHODS

Sample collection and microbiological testing

Samples were taken by local authority sampling officers according to an agreed national protocol for official control sampling (Cefas 2017). The key requirements of this are the testing of at least 10 individual animals per sample, commencement of *E. coli* testing within 48 h and maintenance of samples at a temperature below 10 °C whilst in transit to the laboratory prior to testing. In the current study, data were combined from published studies (Lowther *et al.* 2012; Campos *et al.* 2017b) and a smaller amount of previously unpublished data. The dataset covered 88 commercial production areas in the UK and contained 1,184 paired concentrations of *E. coli* and NoV quantified in shellfish. The majority (98%) of samples were oysters with the remainder being mussels (*Mytilus* spp.). For statistical analysis,

E. coli lower censored Most Probable Number (MPN) values (either <20 or <18) were assigned a value of 10. NoV results at the limit of quantification (100 copies per gram) were assigned a value of 50 and those at the limit of detection (40 copies per gram) were assigned a value of 20.

Statistical analysis

The statistical analysis carried out in this study used total NoV counts (GI + GII) as recommended by EFSA for risk assessment purposes (EFSA 2012). The statistical assessment was conducted using the R statistical software (R Core Team 2016). Initial statistical approaches assessed data from all 88 sites (complete data, Figure 1). However, 37 sites had fewer than 10 data points with one site only having one data point. We considered that, for the assessment of mean and maximum value relationships, sites with only a small number of data points might bias or confound outcomes. For this reason, only sites with 10 or more samples were used for the second stage of analysis (filtered data, Figure 1). Within-site arithmetic mean, median and maximal values were calculated for each site. The geometric mean is commonly used in biological analysis (Buckland et al. 2011), but here we found that its use led to extreme values and non-normal distribution. On the other hand, sites are independent of each other and log10 transformation strongly reduces outlier influence, so we considered that the arithmetic mean of log10 values was a suitable representation of each site's central tendency. To identify the strength and direction of the relationship between E. coli and NoV, Pearson correlations were measured for $\log_{10}$ values, within-site mean, median and maximum (Pearson 1895). Both E. coli and NoV data are measured with error that can cause biased parameter estimates when using standard linear regression. Error-in-variables models, however, assume measurement error in both variables, which is relevant when all variables are experimentally observed (Madansky 1959). In such models, the error term is dependent on the slope of the regression and correlated with the explanatory variable. In this study, data were fitted using an error-in-variables model with the function leiv in the R package leiv (Leonard 2015). The leiv function allows us to reject the hypothesis of 'no relationship' if the 95% confidence interval for the slope does not encompass zero. The model validation is based on the posterior density of the slope and intercept estimates. Once the model is validated, predictions of NoV using E. coli (and vice versa) are derived from the slope and intercept estimates. Fitting an error-in-variables model depends on the ratio of standard deviations, which takes into account measurement error in both variables but does not provide solutions for calculating 95% confidence intervals. For this reason, we calculated a mean absolute percentage error (MAPE) and a prediction error percentage (PEP) as a measure of the prediction accuracy at E. coli and NoV thresholds.

RESULTS

As the distribution of data was highly skewed, a $\log_{10}$ transformation was applied (Figure 2). We then filtered the dataset by site, which reduced the number of sites to 51 with an average of 21 observations collected from 2010 to 2015 (Figures 1 and 3). The principal component analysis (PCA) was conducted on annual data and highlighted the positive correlation between E. coli and NoV and the positive relationship with the winter months (Figures 1 and 4). Based on the PCA, a winter season was defined to include the months from October to February.
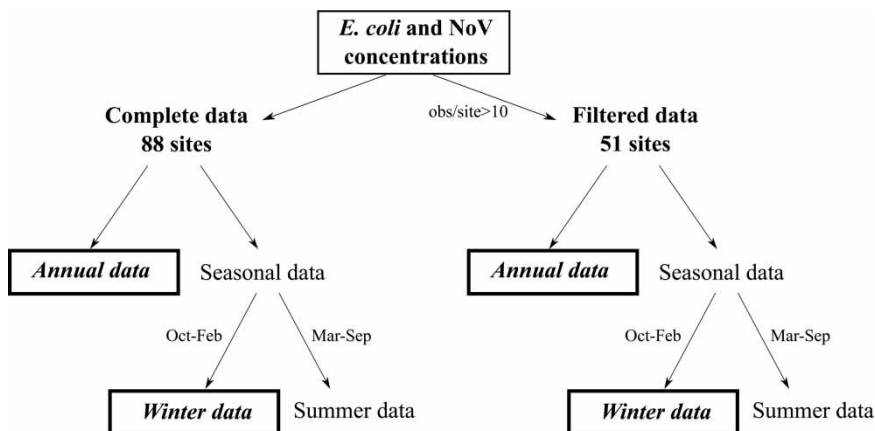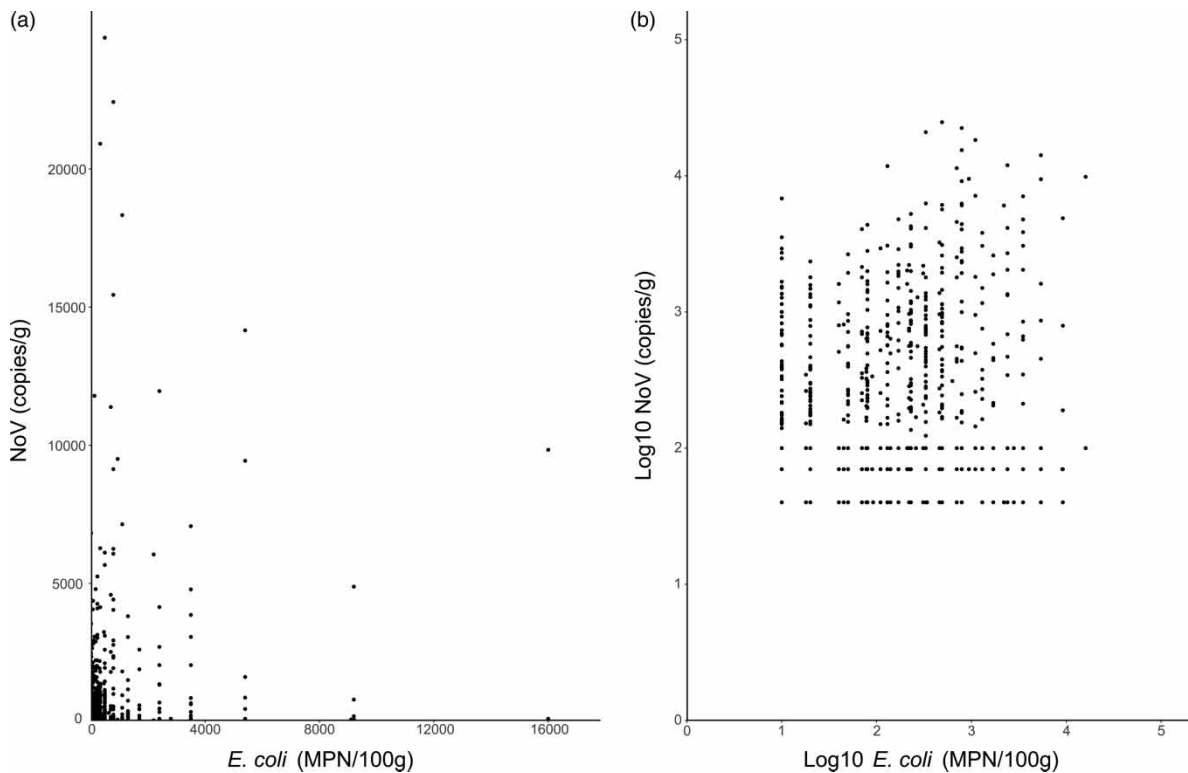

**Figure 1** | Overview of the data filtering and structure.

Figure 2 │ NoV (copies/g) on *E. coli* (MPN/100g) with raw data (a) and log$_{10}$ transformation (b) for annual complete data.

COMPLETE DATA (88 SITES)

Annual approach

The maximum *E. coli* result in any sample is 16,000/100 g and the corresponding NoV result in this sample pair is 9,836 copies/g. The maximum NoV result is 24,754 copies/g with the corresponding *E. coli* result being 490 *E. coli*/100 g in the paired sample. Using all data pairs, the weak positive correlations that we find between *E. coli* and NoV log$_{10}$ values, within-site arithmetic mean, median and maximum are statistically significant at the 5% level (correlation coefficient of 0.20; 0.29; 0.28 and 0.18, respectively). Slopes from both error-in-variables models (NoV over *E. coli* and *E. coli* over NoV) are different from 0 and their 95% confidence intervals exclude 0 (Figure 5). N.B., if the confidence interval for the slope includes 0, no significant relationship between *E. coli* and NoV levels are likely to occur, while, if the confidence interval excludes 0, the relationship between *E. coli* and NoV levels is significant.
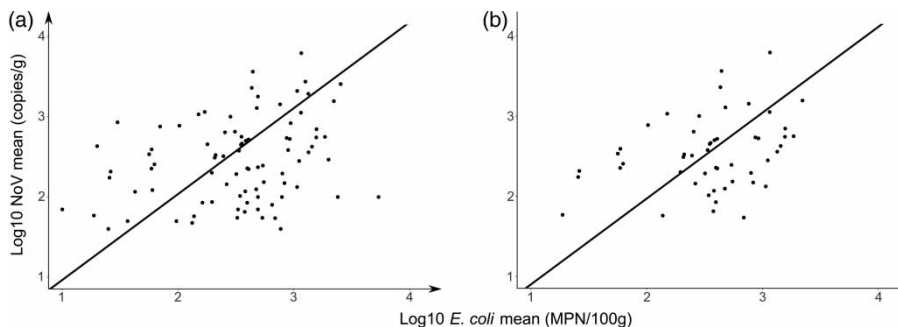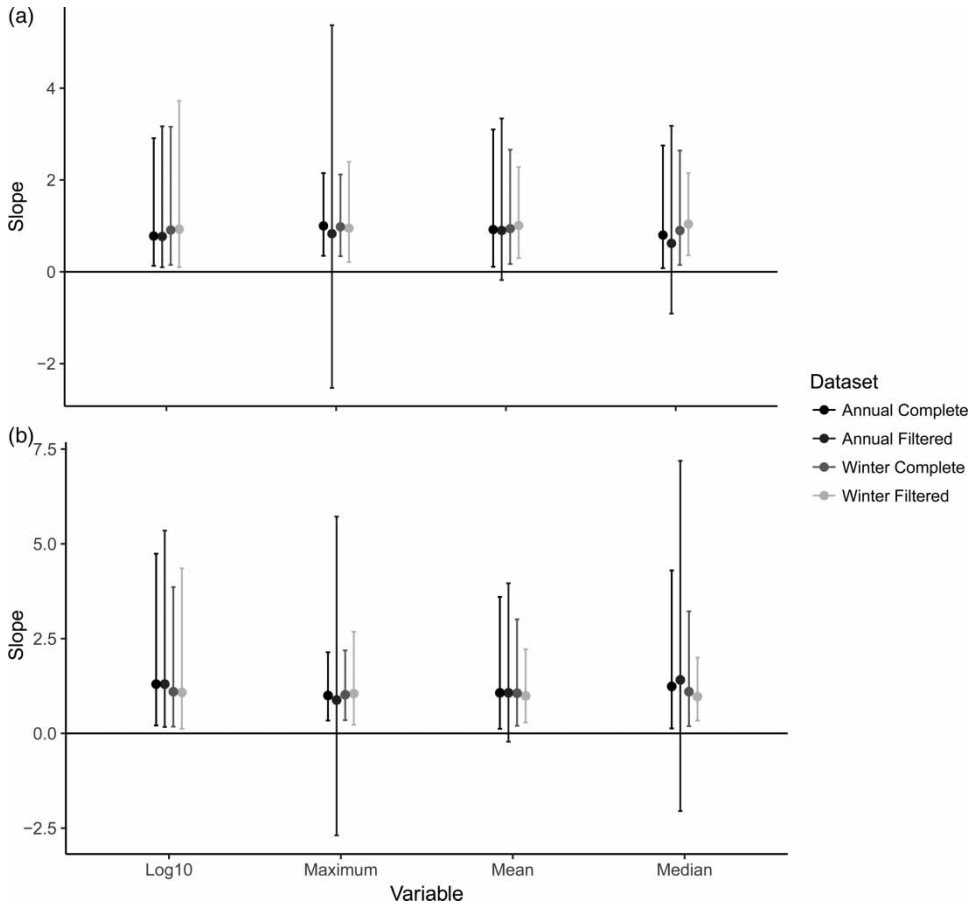

Figure 3 │ Log$_{10}$ NoV within-site arithmetic mean (copies/g) over log$_{10}$ *E. coli* within-site arithmetic mean (MPN/100g) (black points) and the fitted error-in-variable model (black line) for annual complete data (a) and annual filtered data (b).
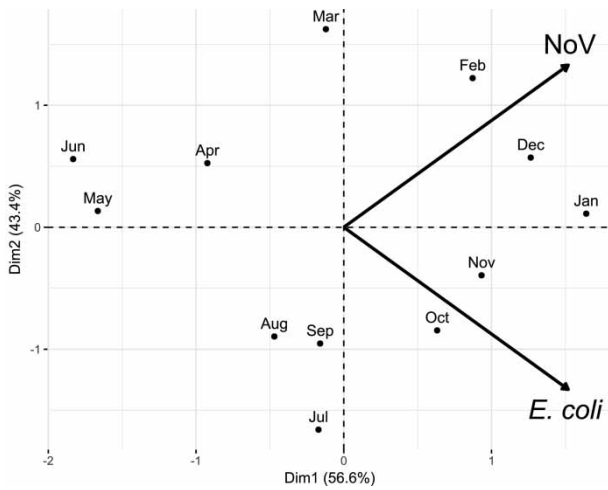
Figure 4 | Principal component analysis (PCA) applied to *E. coli* and NoV log$_{10}$ values with months for complete data. Black circles correspond to months and arrows to *E. coli* and NoV loadings.

Seasonal approach

A strong seasonal effect is identified with the PCA that highlights a difference between 'winter' months from October to February, and 'summer' months from March to September (Figure 4). No correlation is found in the summer season, whereas the individual log$_{10}$ result values, within-site arithmetic mean, median and maximum for *E. coli* and NoV are significantly correlated in the winter season at the 5% level (correlation coefficient of 0.23; 0.35; 0.34 and 0.47, respectively). The error-in-variables model outcomes are consistent with the above correlations. The 95% confidence intervals of the slopes do not include 0 (Figure 5) confirming a statistically significant correlation. Models using the winter data generally show narrower slope distributions than those using the full annual dataset. The exception in this case is found to


Figure 5 | Slopes from error-in-variables regression ((a) NoV over *E. coli*, (b) *E. coli* over NoV) with 95% confident interval (CI) with log$_{10}$ values, within-site arithmetic mean, median and maximum for annual complete data, annual filtered data, winter complete data and winter filtered data (black, dark grey, medium grey and light grey, respectively).

be the individual $\log_{10}$ result values where no such difference is apparent. This indicates that slope estimates are more precise when using regression for within-site arithmetic mean, median and maximum in the winter season than throughout the whole year. NoV predictions at

E. coli thresholds and *E. coli* predictions at NoV thresholds; however, they show strong variations depending on the data and model type (Figures 6 and 7). The lowest prediction errors are obtained with the within-site arithmetic mean models (Figure 8).
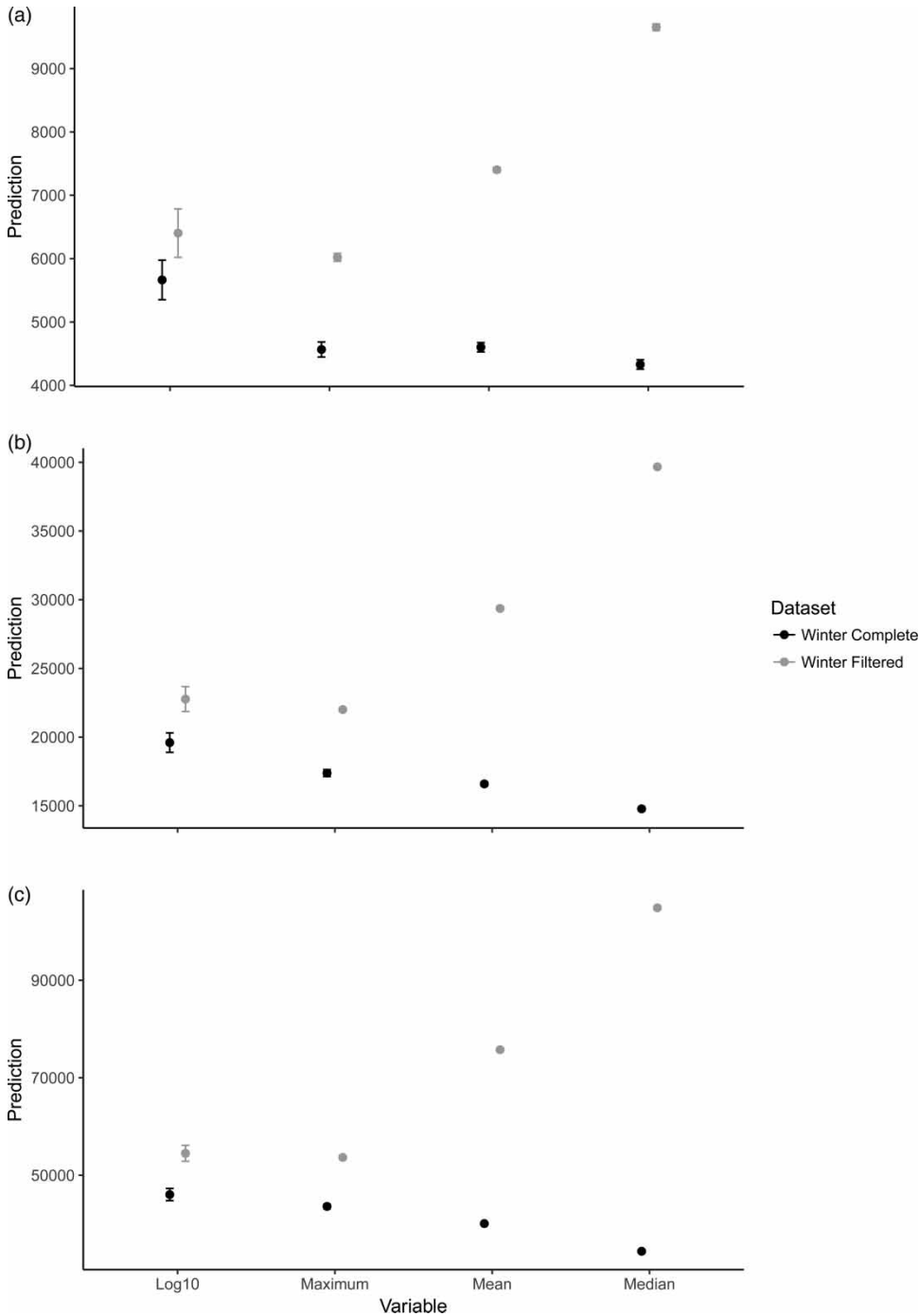
Figure 6 │ NoV prediction ± prediction error (copies/g) at 3 *E. coli* levels ((a) 4,600 MPN/100g; (b) 18,000 MPN/100g; (c) 46,000 MPN/100g) with winter complete data (black) and winter filtered data (grey).
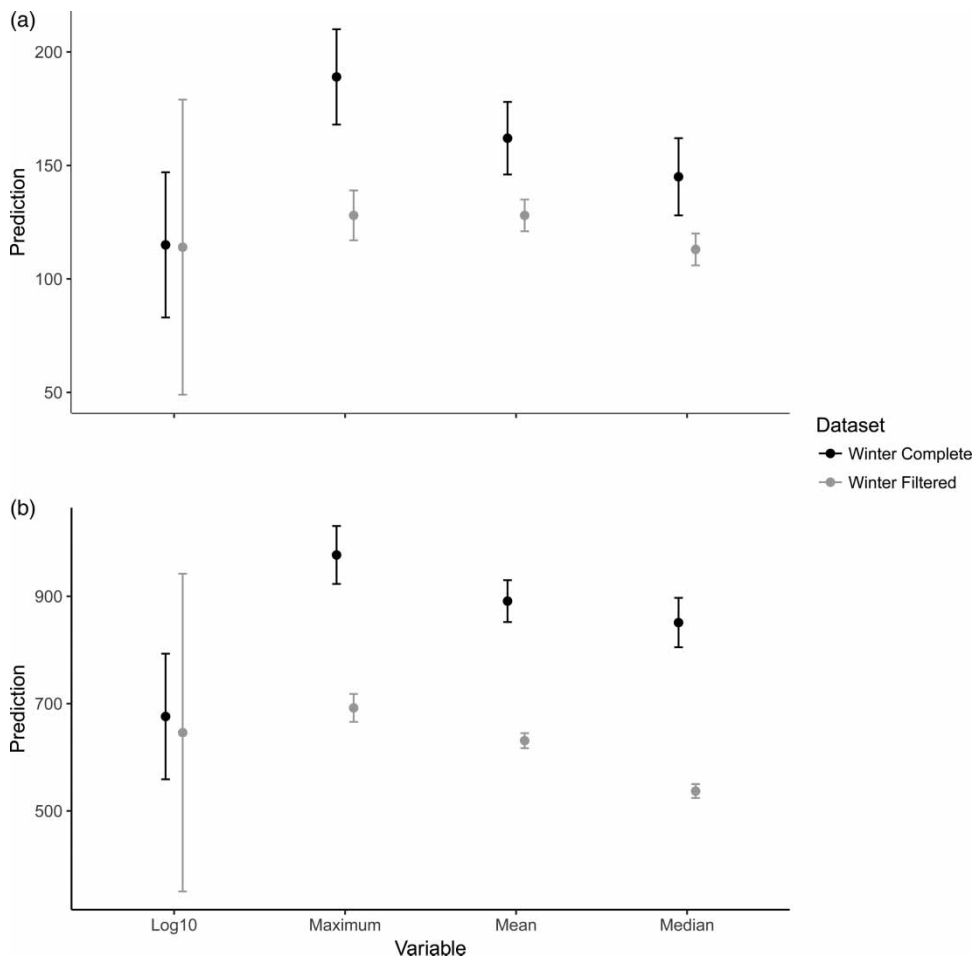
Figure 7 │ *E. coli* prediction ± prediction error (MPN/100g) at 2 NoV levels ((a) 200 copies/g; (b) 1,000 copies/g) with winter complete data (black) and winter filtered data (grey).

FILTERED DATA (51 SITES)

Annual approach

The weak correlation found between annual *E. coli* and NoV individual $\log_{10}$ result values is statistically significant at the 5% level (correlation coefficient of 0.18). Within-site arithmetic mean, median and maximum are, however, not significantly correlated. The error-in-variables models confirm these results with 95% confidence intervals for the slope excluding 0 only in the case of the individual $\log_{10}$ result values (Figure 5).

Seasonal approach

The PCA analysis shows a strong seasonal effect on *E. coli* and NoV values with a similar winter and summer pattern as is found with the annual dataset. Moderate Pearson correlations are found for the individual $\log_{10}$ result values and within-site arithmetic mean, median and maximum in the winter season and are statistically significant at the 5% level (correlation coefficient of 0.19; 0.48; 0.52 and 0.42, respectively), whereas no significant correlation is found in the summer season, except for the within-site median. The error-in-variables models also confirm a positive relationship between *E. coli* and NoV in the winter season (Figure 5). Confidence intervals for the slope exclude 0 for both individual $\log_{10}$ result values and arithmetic mean, median and maximum values. Moreover, slope distributions in all cases, except when using the individual $\log_{10}$ result values, are taller and narrower in the winter season than in the annual dataset. NoV and *E. coli* predictions, however, vary considerably depending on the actual data used for the error-in-variables models (Figures 6 and 7). The possible reasons for this are discussed later. Within-site arithmetic mean models provide predictions associated with the lowest error (Figure 8).
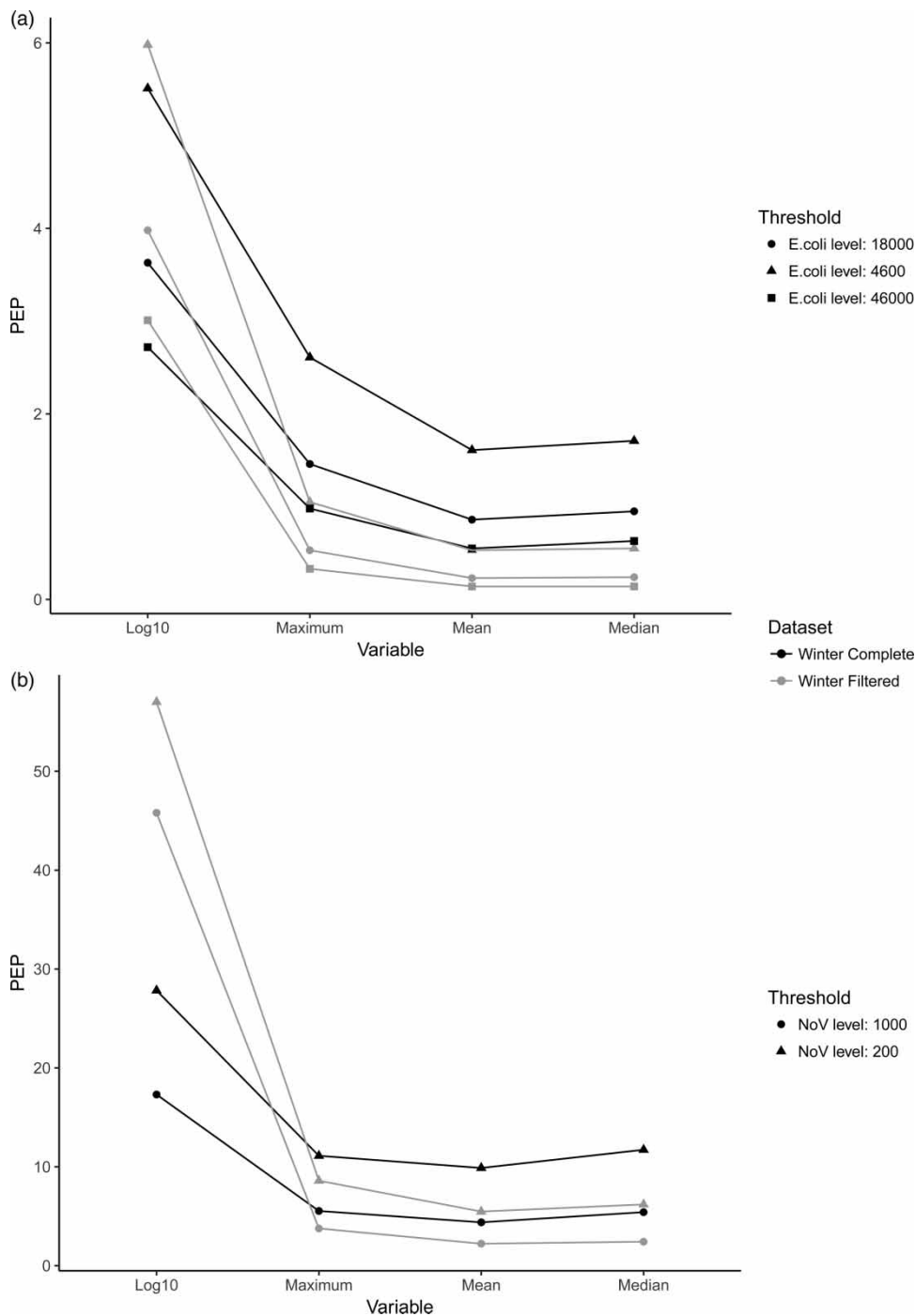
Figure 8 │ Prediction error percentage (PEP) (%) at 3 *E. coli* levels ((a) circle points: 4,600 MPN/100g; triangle points: 18,000 MPN/100g; square points: 46,000 MPN/100g) and at 2 NoV levels ((b) circle points: 200 copies/g; triangle points: 1,000 copies/g) from the error-in-variables models with winter complete data (black) and winter filtered data (grey).

From the slope and intercept estimated with the error-in-variables models, i.e. from the linear relationship found between NoV and *E. coli*, we derived NoV predictions using *E. coli* and calculated the associated percentage prediction error. In terms of risk management for NoV using *E. coli*

(Figure 6) using the filtered UK winter dataset, we find that reducing the current class B maximum from 46,000 *E. coli* per 100 g (corresponding NoV value of 75,750 ± 103) to 18,000 *E. coli* per 100 g (corresponding NoV value of 29,365 ± 69) reduces maximum levels of NoV by a factor of

2.6. Further reduction to 4,600 *E. coli* per 100 g (corresponding NoV value of $7,403 \pm 39$) reduces maximum levels of NoV by a factor of 10.2. In predicting maximum *E. coli* values from maximum NoV values (Figure 7) using the filtered winter dataset, we found using the UK filtered winter dataset that a maximum of 200 NoV corresponded to a maximum of 128 ± 7 *E. coli* per 100 g. A maximum of 1,000 NoV corresponded to a maximum of 631 ± 14 *E. coli* per 100 g.

Ratio of means (*E. coli*/NoV)

Finally, the variability in the relationship between *E. coli* and NoV incidence in paired samples can be expressed as a ratio of means (we used arithmetic mean of $\log_{10}$ values to be consistent with the approach described above). Using data from sites with 10 or more paired results (51 sites in all), the ratio of means ($\log_{10}$ *E. coli*/$\log_{10}$ NoV) was found to range from 0.6 to 1.4 across the 51 sites (Figure 9).

DISCUSSION

This study focused principally on oysters as these represent the highest risk for shellfish-associated NoV infection (Potasman *et al.* 2002; Lees 2010). The analysis of both complete and filtered data confirmed a significant relationship between *E. coli* and NoV levels. This relationship was stronger in the winter season, allowing a better prediction of NoV from *E. coli* levels. Moreover, good correlation and low prediction error for the within-site arithmetic means suggested that the trend of NoV can be best predicted from *E. coli* means. However, it was apparent from our analysis that the relationship between *E. coli* and NoV varied with the amount of data used. For example, assessing the general relationship between maximum or average results across all sites had the potential to be considerably biased by sites where the number of data pairs was small (e.g. <10), particularly where extreme results had occurred in either the *E. coli* or NoV dataset. In this way, a small number of sites with extreme or potentially unrepresentative values can significantly influence the overall estimation of the relationship between *E. coli* and NoV. Factors such as poorly representative and/or reliable data at some sites, or differential inactivation and variable pollution source (animal vs. human) inputs may all affect the relationship assessment.

We noted in all cases, except for individual $\log_{10}$ result values, that no significant correlation was found between *E. coli* and NoV levels when using the annual filtered data (i.e. sites with 10 results or more) but significant correlations were identified with the winter filtered data. However, correlations were always found to be significant with both annual and winter complete data from all sites (i.e. unfiltered). Our conclusions from this observation were that filtering the data to ensure sites had at least 10 data pairs removed under-represented sites that could strongly bias the relationships between *E. coli* and NoV due to their potentially atypical data. Furthermore, we concluded that the apparently significant relationship found with the complete dataset is probably due to these specific sites. This result shows that our approach is sensitive to outlier data.
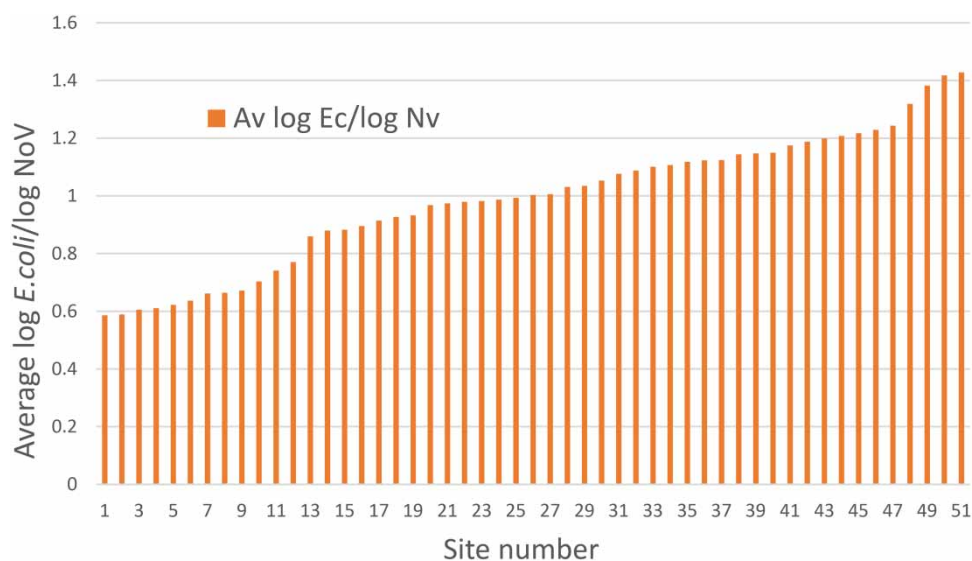

Figure 9 | Ratio of average log *E. coli*/average log NoV by site.

Focusing the study on the winter season (only) improved the strength of the correlations and therefore our confidence in the model predictions with both complete and filtered data. Since no significant correlation was found in the summer season, we determined that the use of this data by assessing the full annual dataset may reduce the strength of the relationship between *E. coli* and NoV levels. Excluding the data for the 'summer' period, during which NoV is not traditionally so prevalent in the community, thus allowed us to improve the model quality and strengthen the prediction robustness. The low prediction error obtained with the within-site arithmetic mean from filtered data indicates that NoV levels could be predicted from *E. coli* mean with a reasonably good level of confidence in the winter season.

We investigated the correlation of within-site mean values in this study. The nature of the data available, *E. coli* and NoV results from samples taken at the same time within sites, meant that this was the most practical option. We accept that this approach does not directly address the variance in the between-site NoV and *E. coli* correlation. From the shellfish consumer perspective, it is clearly important to know whether *E. coli* monitoring of shellfish can give adequate protection from NoV risk in any given situation. This study highlights that *E. coli* is not able to give such assurances for all sites and at all times. Nevertheless, we consider that our approach confirms that *E. coli* data may be useful in assessment of potential NoV risk at many sites, particularly in the winter months. Conversely, it also confirms that there are some sites where other approaches (e.g. direct testing for NoV) may be necessary.

We found that reducing the current class B maximum from 46,000 to 18,000 *E. coli* per 100 g reduces NoV risk by a factor of 2.6. Reducing it further to an absolute limit of 4,600 *E. coli* per 100 g (currently 90% compliance with this value is allowed) reduces NoV risk by a factor of 10.2. We assumed for the purposes of this exercise that higher NoV values equated to higher risk although it is recognised that, given the inability of the current NoV method to distinguish between viable and non-viable NoV, it is currently not possible to draw a direct link between NoV copies/g in shellfish and consumer risk.

A maximum NoV level of 200 copies/g (combined GI and GII) has been tentatively proposed as a possible end product standard value, with a suggested maximum acceptable value of 1,000 for raw product from shellfish production areas prior to treatment (EURL 2012). In order to determine what this might equate to in terms of maximum *E. coli* values, we used the errors-in-variables model to predict maximum *E. coli* values from maximum NoV values

(Figure 7) using the filtered UK winter dataset. We found that a maximum of 200 NoV corresponded to a maximum of 128 ± 7 *E. coli* per 100 g. A maximum of 1,000 NoV corresponded to a maximum of 631 ± 14 *E. coli*. per 100 g. It should be noted that the current upper limit for class A sites under Regulation 854/2004 is 700 *E. coli*/100 g which, according to the predictions from our model, could allow a NoV level of up to 1,000 copies/g – a limit proposed for raw shellfish prior to treatment and higher than that proposed for end product. EFSA recently noted (EFSA 2012) that an end product standard limit of 200 NoV would have meant 61.1% non-compliant batches according to data taken from the UK during January–March 2010 (24.4 to 83.3% in France and Ireland, respectively). Whilst the intention of introducing any NoV standards would be to improve consumer protection levels, the above figures clearly indicate the potential for a significant adverse impact on the shellfish industry. At the time of writing, EFSA are currently undertaking a two-year baseline survey of NoV in oysters to assess Europe-wide prevalence of NoV, with a view to potentially establishing a legislative standard for NoV in shellfish. This survey will generate a large dataset which will contribute significantly to the evidence base for this area of regulation.

Observations from the data: It is obvious from the data that the strength of relationship between *E. coli* and NoV varies from site to site. It has been reported that differential inactivation can occur at some sites due to the effects of different forms of sewage treatment. In particular, UV disinfection on sewage discharges has been found to produce a 5 log reduction in *E. coli* compared with only a 2–3 log reduction in NoV (Campos *et al.* 2016). UV disinfection can therefore lead to situations where there are low *E. coli* counts in shellfish but high NoV levels. Campos *et al.* (2016) also noted a lesser differential reduction in *E. coli* vs NoV levels according to some forms of secondary treatment. These authors also noted a high degree of variability in the efficiency of treatment at the treatment works studied. On the basis of the results from our study it would appear that this potential for differential reduction, perhaps combined with intervening distance between discharge and shellfish sampling point, may also lead to situations where high NoV and low *E. coli* shellfish results could be observed. An effect with increased distance from sewage discharges generally may be due to a combination of UV from sunlight and other environmental effects such as predation from protozoa or other microfauna. One example from this study supporting this observation is a site known to be impacted by large sources of secondary treated effluent some distance (4–5 km) upstream of the shellfish

sampling point. This particular site returned a consistent trend of marked high NoV compared with low *E. coli* results. One example pair was a NoV result of 6,815 copies per gram vs. a result <18 *E. coli*/100 g.

The current European *E. coli* 5×3 tube MPN reference method, as used in this study, cannot differentiate between animal and human inputs (Walker *et al.* 2018). GI and GII NoV are predominantly from human sources, whereas the *E. coli* detected in the MPN method can be from both animal and human sources. One example of a sample pair from this study from a site known to be exposed to predominantly faecal contamination from an animal source would be results of 9,200 *E. coli*/100 g and undetected NoV. Thus, the ratio of *E. coli* to norovirus at a site is also likely to be impacted by the extent of non-human sources of *E. coli*.

A consideration with the current PCR test for NoV is that it provides no indication of NoV viability. A proportion of the NoV count may therefore be non-viable virus. This proportion may itself be variable depending on environmental factors (e.g. sunlight) and sewage treatment processes (e.g. UV disinfection). The relationship between the number of infectious virus particles and the number of virus genome copies detected by quantitative PCR is not a constant and EFSA has identified that the infectious risk associated with low level positive oysters as determined by RT-PCR may be overestimated (EFSA 2012; Hartard *et al.* 2017). NoV infection incidence in England (Public Health England 2017) is very seasonal and, as in most other temperate regions, is variable in the community generally (unlike *E. coli*) particularly in small community situations. Consequently, NoV may not be present during low risk periods, even in polluted sites. This scenario could also lead to high *E. coli* shellfish results but absence of NoV.

The variability in the relationship between *E. coli* and norovirus incidence in paired samples from sites can be expressed as a ratio of means. Using data from sites with 10 or more paired results (51 sites in all) the ratio of mean logged data (*E. coli*/NoV) ranged from 0.6 to 1.4 across the 51 sites (see Figure 9). The lower ratios (towards 0.6) might typically indicate situations where the contribution from UV disinfected sewage discharges was more significant. The site in this study with the lowest ratio of 0.6 is known to be impacted by two separate UV disinfected sewage treatment works. Conversely, higher ratios (towards 1.4) might typically indicate a prevalence of animal sources of pollution. Two of the three sites with these highest ratios are known to be significantly impacted by animal sources of contamination. Curiously, however, the second highest value was returned from a predominantly urban site thought to be impacted by a number of sewage discharges (including UV disinfection). Clearly the relationship is not a simple one and will require further work to fully elucidate the factors of relevance. It is worth noting that UV disinfection of sewage discharges is increasingly being adopted across the USA and Europe.

Many environmental factors can influence *E. coli* and NoV concentrations and their relationship at specific sites. These include location, rainfall, water temperature, current flows and types of pollution sources. In particular, ultraviolet disinfection of sewage discharges can lead to a greater degree of inactivation of *E. coli* compared with NoV and this will limit the usefulness of *E. coli* as an indicator at sites where the contribution from UV disinfected discharges is significant. Conversely, significant inputs of pollution from animal pollution sources could lead to an overestimation of NoV risk if based solely on *E. coli* monitoring. Nevertheless, in general, assessment of longer term data (e.g. 3 years or more of monthly monitoring) from representative monitoring points, combined with information on relevant local environmental factors and pollution sources can ensure a greater robustness to our approach. Overall, we suggest a long-term winter dataset for *E. coli* at a site can give a valuable indication of the likely risk for NoV.

CONCLUSIONS

It should be emphasised that the required evidence-base, particularly in terms of comparative *E. coli* and NoV data, in this particular area of regulation is lacking. This study seeks to make the best use of the limited data that are available in the UK but recognises that, whilst a general relationship appears to exist, this cannot be assumed with any certainty for all sites and at all times of year.

The main conclusions from this study are that no significant relationship was found between *E. coli* and NoV in individual sample pairs. The best relationships were found when using site-specific mean values using data from the winter months (October–February) at sites with at least 10 data points. We found higher correlations for the mean values but lower correlations for the individual $\log_{10}$ result and maximal values. We found that our models were sensitive as removing data led to different outcomes. We focused on the winter season, the traditional period of higher NoV incidence in the community, and sites with at least 10 data pairs, which improved both correlations and model predictions. We found that reducing the current class B

maximum from 46,000 to 18,000 reduces NoV risk by a factor of 2.6 to 1. Reducing it further to an absolute limit of 4,600 (currently 90% compliance with this value is allowed) reduces NoV risk by a factor of 10.2 to 1. We found that a maximum of 200 NoV corresponded to a maximum of 128 ± 7 *E. coli*. A maximum of 1,000 NoV corresponded to a maximum of 631 ± 14 *E. coli*.

In general, this study supports the use of *E. coli* as currently employed in the classification of harvesting areas using longer term (e.g. 3 years) datasets (as recommended in the European Community Guidance and Good Practice Guide in support of EC Regulation 854/2004) as this is more representative of NoV risk (particularly when assessed over the winter months) than individual data points. However, given the variability of the *E. coli*/NoV ratio of means observed between sites in this study, a standard based on *E. coli* alone to represent NoV risk (as determined by PCR) will not be protective at all sites, in particular, those impacted by UV disinfected sewage discharges.

Further work is needed to assess the significance of NoV levels and to address the question of virus viability. Alternative approaches may need to be developed to better assess levels of risk at sites where the *E. coli*/NoV correlation is poor, e.g. sites subject to influence from UV disinfected sewage discharges or animal faecal inputs. Knowledge of the sources of pollution and their relative contribution is key to interpretation of *E. coli* monitoring data and subsequent risk management measures.

REFERENCES

Bellou, M., Kokkinos, P. & Vantarakis, A. 2013 Shellfish-borne viral outbreaks: a systematic review. *Food and Environmental Virology* **5** (1), 13–23.

Buckland, S. T., Studeny, A. C., Magurran, A. E., Illian, J. B. & Newson, S. E. 2011 The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere* **2** (9), 100.

Campos, C. J. A., Avant, J., Lowther, J., Till, D. & Lees, D. N. 2016 Human norovirus in untreated sewage and effluents from primary, secondary and tertiary treatment processes. *Water Research* **105**, 224–232.

Campos, C. J. A., Kershaw, S., Morgan, O. C. & Lees, D. N. 2017a Risk factors for norovirus contamination of shellfish water catchments in England and Wales. *International Journal of Food Microbiology* **241**, 318–324.

Campos, C. J. A., Goblick, G., Lee, R., Wittamore, K. & Lees, D. N. 2017b Determining the zone of impact of norovirus contamination in shellfish production areas through microbiological monitoring and hydrographic analysis. *Water Research* **124**, 556–565.

Cefas 2017 Protocol for the collection of shellfish under the microbiological classification monitoring programme (EU Regulation 854/2004). Version 8, April 2017. https://www.cefas.co.uk/media/52551/201704-cefas-classification-sampling-protocol-for-local-authorities-final.pdf.

EFSA 2012 Scientific opinion on Norovirus (NoV) in oysters: methods, limits and control options. *EFSA Journal* **10** (1), 2500.

EURL 2012 *Discussion Paper on Live Bivalve Molluscs (LBM) and Human Enteric Virus Contamination: Options for Improving Risk Management in EU Food Hygiene Package*. https://eurlcefas.org/media/13637/ws12_15.pdf.

European Communities 2004 Regulation (EC) No. 854/2004 of the European Parliament and of the Council of 29 April 2004 Laying Down Specific Rules for the Organisation of Official Controls on Products of Animal Origin Intended for Human Consumption. *Official Journal of the European Union* L226, 25.6.2004, 83–127.

Galtsoff, P. S. 1964 *The American Oyster Crassostrea Virginica Gmelin*. Fishery Bulletin of the Fish and Wildlife Service, United States Government Printing Office, Washington, DC, USA, p. 64.

Hartard, C., Leclerc, M., Rivet, R., Maul, A., Loutreul, J., Banas, S., Boudaud, N. & Gantzer, C. 2017 F-Specific RNA bacteriophages, especially members of subgroup II, should be reconsidered as good indicators of viral pollution of oysters. *Applied and Environmental Microbiology* **84** (1), e01866–17.

ISO 2017 ISO 15216-1:2017. *Microbiology of the Food Chain – Horizontal Method for Determination of Hepatitis A Virus and Norovirus Using Real-Time RT-PCR – Part 1*. Method for quantification, Geneva, Switzerland.

Kay, D., Kershaw, S., Lee, R., Wyer, M. D., Watkins, J. & Francis, C. 2008 Results of field investigations into the impact of intermittent sewage discharges on the microbiological quality of wild mussels (*Mytilus edulis*) in a tidal estuary. *Water Research* **42**, 3033–3046.

Lees, D. N. 2010 Viruses and bivalve shellfish. *International Journal of Food Microbiology* **59** (2000), 81–116.

Leonard, D. 2015 leiv: Bivariate Linear Errors-In-Variables Estimation, R package version 2.0-7.

Lowther, J. A., Gustar, N. E., Powell, A. L., Hartnell, R. E. & Lees, D. N. 2012 Two-year systematic study to assess norovirus contamination in oysters from commercial harvesting areas in the United Kingdom. *Applied and Environmental Microbiology* **78** (16), 5812–5817.

Madansky, A. 1959 The Fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association* **54** (285), 173–205.

Pearson, K. 1895 Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* **58**, 240–242.

Potasman, I., Paz, A. & Odeh, M. 2002 Infectious outbreaks associated with bivalve shellfish consumption: a worldwide perspective. *Clinical Infectious Diseases* **35**, 921–928.

Public Health England 2017 *PHE National Norovirus and Rotavirus Report*. Summary of surveillance of norovirus and rotavirus, 19 October 2017 – data to week 40. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/652744/Norovirus_update_2017_week_40.pdf.

R Core Team 2016 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Walker, D. I., Younger, A., Stockley, L. & Baker-Austin, C. 2018 *Escherichia coli* testing and enumeration in live bivalve shellfish – present methods and future directions. *Food Microbiology* **73**, 29–38.

Wu, J., Long, S. C., Das, D. & Dorner, S. M. 2011 Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. *Journal of Water and Health* **9** (2), 265–278.

Xu, Z., Li, Z., Wang, J., Xiao, Z. & Dong, D. 1992 Ecology and prevention of a shellfish-associated hepatitis A epidemic in Shanghai, China. *Vaccine* **10** (S1), S67–S68.