**The impact of meteorology on the occurrence of waterborne outbreaks of vero cytotoxin-producing *Escherichia coli* (VTEC): a logistic regression approach**

Jean O’Dwyer, Margaret Morris Downes and Catherine C. Adley

**ABSTRACT**

This study analyses the relationship between meteorological phenomena and outbreaks of waterborne-transmitted vero cytotoxin-producing *Escherichia coli* (VTEC) in the Republic of Ireland over an 8-year period (2005–2012). Data pertaining to the notification of waterborne VTEC outbreaks were extracted from the Computerised Infectious Disease Reporting system, which is administered through the national Health Protection Surveillance Centre as part of the Health Service Executive. Rainfall and temperature data were obtained from the national meteorological office and categorised as cumulative rainfall, heavy rainfall events in the previous 7 days, and mean temperature. Regression analysis was performed using logistic regression (LR) analysis. The LR model was significant \( (p < 0.001) \), with all independent variables: cumulative rainfall, heavy rainfall and mean temperature making a statistically significant contribution to the model. The study has found that rainfall, particularly heavy rainfall in the preceding 7 days of an outbreak, is a strong statistical indicator of a waterborne outbreak and that temperature also impacts waterborne VTEC outbreak occurrence.

**Key words** | *E. coli*, meteorology, rainfall, temperature, VTEC, waterborne disease

**INTRODUCTION**

Vero cytotoxin-producing *Escherichia coli* (VTEC) are major enteropathogens responsible for causing outbreaks of haemorrhagic colitis and haemolytic-uraemic syndrome (HUS). VTEC strains are named after their ability to produce verotoxins, whose role in the manifestation of bloody diarrhoea and HUS was first reported by Karmali et al. (1983). Although more than 150 serotypes of VTEC have been identified, most outbreaks are caused by the *E. coli* O157 serotype (Pennington 2010), which has an infective dose in the range of 2–45 organisms (Tilden et al. 1996). Clinical features of *E. coli* O157 gastroenteritis include diarrhoea involving five or more bowel motions per day, becoming bloody in 90% of cases, usually after 4 days (Tarr et al. 2005). Approximately 10–15% of patients infected with *E. coli* O157 develop HUS 5–13 days after the onset of diarrhoea, with an estimated 65% of these occurring in children aged less than 5 years (Gould et al. 2009).

*E. coli* O157 and other serotypes of pathogenic *E. coli* are found regularly in the faeces of healthy cattle (Robinson et al. 2004) and may be transmitted to humans through contaminated food (Clarke et al. 1994; Thorns 2000; De Boer & Heuvelink 2001), water (Jackson et al. 1998; Olsen et al. 2002; Johnson et al. 2005) and direct contact (Frank et al. 2008) with infected people (Karmali et al. 1988; Boudailliez et al. 1997; Aslani & Bouzari 2005) or animals (Blanco et al. 1996; Heuvelink et al. 1998; Pritchard et al. 2000). Consequently, VTEC infections are ubiquitous among countries that raise cattle, including the United States (Berkelman 1994), Canada (Johnson et al. 2005), England (Adak et al. 2002), Scotland (Cadwgan et al. 2002) and the Republic of Ireland (Garvey 2010), among others.

Focusing on Ireland, since 2004, changes to infectious disease legislation (S.I. 707 of 2003) have resulted in all
VTEC cases becoming notifiable to the Department of Public Health, Health Protection Surveillance Centre (HPSC) and reported to the European Centre for Disease Control (ECDC). Since then, Ireland has had the highest crude incidence rates of VTEC in Europe; increasing from 3.9/100,000 in 2007 to 12.07/100,000 in 2012 (HPSC 2008, 2013; ECDC 2013) with the exception of 2011, when Germany reported the highest rate due to a large VTEC O104 outbreak linked with fenugreek/sprouted seeds (ECDC 2013).

Transmission of VTEC infection

While direct contact with infected animals or contaminated food is often the transmission pathway for VTEC infection, waterborne transmission of VTEC is also an area of concern within the public health arena. Accredited to its persistence in the environment, VTEC has the potential to cause large outbreaks, particularly through water. For example, more than 2,300 cases of gastroenteritis were associated with an outbreak of E. coli O157:H7 and Campylobacter jejuni infection following contamination of a municipal water supply in Walkerton, Ontario, in May 2000, in which seven people died (Hrudey et al. 2003). Drinking water supplies that are untreated pose a particular threat to public health and can serve as a reservoir of infectious disease capable of impacting a large population, with a link between VTEC notifications and the use of private water supplies (PrWS) identified by Health Service Executive investigations. A PrWS is defined as a supply serving less than 50 households and producing less than 10 m³ per day (EPA 2012). These supplies are exempted from the European Commission Drinking Water Directive (DWD) 98/83/EC and hence the treatment of supplies, if any, is the sole responsibility of the owner (Council Directive 1998). Approximately 13% of households in Ireland are reliant on an unregulated PrWS (CSO 2012).

Outbreaks of VTEC infection (an episode in which two or more people, thought to have a common exposure, experience a similar illness or proven infection) associated with drinking water have a strong epidemiological presence in Ireland, and with all notified drinking water-associated outbreaks linked with PrWS in 2012 (HPSC 2013), there has never been a VTEC outbreak associated with a public, treated supply in Ireland. The sources and pathway of water contamination are varied, particularly in relation to groundwater supplies. The underlying geology, soil characteristics, land usage and local infrastructure are all contributing factors to groundwater vulnerability, but meteorological phenomena also are likely to play a significant role in contamination. Persistent or heavy rainfall, for example, can mobilise pathogens within the environment increasing runoff from agriculture and transporting this microbiologically diverse medium into rivers, coastal waters and groundwater wells (Semenza & Menne 2009); the latter being noted as a key player in VTEC epidemiology by the Irish Environmental Protection Agency (O’Reilly 2013). A previous study of selected PrWS in Ireland by O’Dwyer et al. (2014), demonstrated that E. coli contamination increased with an increase in rainfall. Similarly, temperature, which will have a direct effect on the reproductive rate of the organism (Semenza & Menne 2009) could potentially be a contributing factor to waterborne VTEC outbreaks.

In Ireland, however, the impact of meteorology on waterborne VTEC outbreaks, if any, has not been quantified. Therefore, assessing meteorological factors that may ‘trigger’ or enable a waterborne outbreak of VTEC is an important public health service. This study has used data from a variety of sources to create a database of waterborne outbreaks over an 8-year period (2005–2012) in the Republic of Ireland and has examined the relationship between cumulative rainfall, heavy rainfall in the previous 7 days of infection notification, the mean monthly temperature and outbreaks of waterborne infectious disease. The aim is to assess if specific meteorological parameters precede waterborne VTEC outbreaks.

METHODS

Data collection and collation

Data were collected from multiple sources and are summarised in Table 1. Waterborne outbreaks are classified in the event that there is a strong epidemiological link or if there is bacteriological evidence implicating a source. As a result, there is high confidence that the outbreaks are associated with water. However, where the source of
contamination is undetermined, the cause is stated as 'unknown'; as a result, it is possible that the number of waterborne-derived outbreaks is, in fact, underreported. The date of onset of the primary outbreak case was utilised as the index case in this study so as to ensure confidence in the waterborne nature of the outbreak; subsequent cases could potentially be by person to person. To create a consistency between the three data types (rainfall, temperature and number of VTEC outbreaks), the data were categorised based on month (January to December) for the 8 years (2005–2012 inclusive). Where the onset date was close to the start of a month, i.e., within 3 days of the first day of the month, the previous month’s cumulative rainfall and mean temperature were used as the comparative total. Total monthly cumulative rainfall was utilised over weekly cumulative rainfall in order to reduce collinearity between rainfall and heavy rainfall events. In terms of heavy rainfall events, the previous 7 days prior to onset of symptoms was analysed. This allows for the incubation period of VTEC infection which is typically between 2 and 4 days (Slutsker et al. 1997; World Health Organization 2011; O’Reilley 2013). Heavy rainfall was defined as days where rainfall exceeded 30 mm over a 24-hour period. This measurement was used in accordance with the Irish Meteorological Office’s weather warning categories, whereby a ‘category orange’ alert is issued when rainfall exceeds the aforementioned limit.

Data analysis

Multivariate analysis

Prior to regression modelling, precursory analysis was undertaken to identify relationships between the independent variables (cumulative monthly rainfall, heavy rainfall in the previous 7 days and mean monthly temperature) and the dichotomous dependent variable (month without outbreak (0), month with outbreak (1)). Continuous variables were tested for normality using the Shapiro–Wilk test of normality. Standard Pearson’s chi-square tests of independence were used in comparing proportions within groups of the study. Mann–Whitney U tests were used to investigate significant mean differences between continuous and dichotomous variables. All statistical analyses were

### Table 1 | Description of the dependent and independent variables inputted into the LR model

<table>
<thead>
<tr>
<th>Data description</th>
<th>Data source</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterborne VTEC outbreaks as reported by the Health Protection Surveillance Centre</td>
<td>Computerised Infectious Disease Reporting (CIDR) system and from the Zoonotic and Vector borne Disease, and Outbreaks of Infectious Disease Quarterly Reports by the Health Protection Surveillance Centre</td>
<td>Outbreak data were categorised based on month number as per the ISO-8601 standard (ISO 8601:2004). Thus, outbreak data were transformed to a binary variable where 1: outbreak and 0: no outbreak</td>
</tr>
<tr>
<td>Cumulative rainfall (total monthly rainfall)</td>
<td>Irish Meteorological Office: Met Eireann: To facilitate homogeneity between rainfall and outbreak data, the station(s) in closest proximity to the area of outbreak was utilised. Where there were two or more stations in the area, the average cumulative rainfall was derived and utilised</td>
<td>Continuous variable. The total daily and subsequent total monthly rainfall was summed for each station for each month of the 8-year period</td>
</tr>
<tr>
<td>Heavy rainfall events (rainfall in excess of 30 mm in a 24-hour period (Met Eireann 2004)) in previous 7 days prior to outbreak</td>
<td>Irish Meteorological Office: Met Eireann: To facilitate homogeneity between rainfall and outbreak data, the station(s) in closest proximity to the area of outbreak was utilised</td>
<td>Binary variable where 1: heavy rainfall event and 0: no heavy rainfall event</td>
</tr>
<tr>
<td>Temperature (mean monthly temperature)</td>
<td>Irish Meteorological Office: Met Eireann</td>
<td>Continuous variable. The mean daily temperature and subsequent mean monthly temperature was summed for each station for each month of the 8-year period</td>
</tr>
</tbody>
</table>
performed in IBM SPSS Statistics 22. The significance level was set at 5% \((p < 0.05)\) for all analyses.

**Logistic regression**

Binary logistic regression (LR) was utilised as the statistical analysis in this study as the dependent variable (waterborne VTEC outbreaks) is dichotomous in nature, i.e., the presence or absence of an outbreak. LR has been used in the health sciences since the late 1960s to predict a binary response from explanatory variables (Lemeshow et al. 1988). The goal of LR is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest and a set of independent, predictor or explanatory variables, which in this study are cumulative monthly rainfall, heavy rainfall and mean monthly temperature. LR differs from classical linear regression in that the modelled response is the probability of being in a category, rather than the observed quantity of a response variable. The LR model has the following form:

\[
P = \frac{e^{b_0 + bX}}{1 + e^{b_0 + bX}}
\]

where \(P\) is the probability that an outbreak is detected, \(b_0\) is a scalar intercept parameter, \(X\) is a vector of \(n\) explanatory variable values, and \(b\) is vector of slope coefficient values so that \(bX = b_1X_1 + b_2X_2 + \ldots + b_nX_n\). Detailed descriptions of LR are available in Helsel & Hirsch (2002) and Hosmer et al. (2013).

Explanatory variables in the LR model were checked for statistical significance using the Wald statistic (Hosmer et al. 2013). The rescaled \(R^2\) value and the Hosmer–Lemeshow (HL) goodness-of-fit test statistic were used to evaluate model discrimination and calibration, respectively (Hosmer et al. 2013). The predictor variables were also assessed for collinearity issues, which can affect the integrity of the LR model.

**RESULTS**

For the 8 years (2005–2012) analysed in this study, a total of 32 cases of waterborne-associated VTEC outbreaks were notified in the Republic of Ireland, which were allocated to 25 months over a 96-month period (2005–2012). These 32 outbreaks represent a total of 137 confirmed cases as represented in Table 2.

To assess the relationship between meteorological parameters and waterborne VTEC outbreaks prior to LR modelling, the data were analysed for distribution and subsequently multivariate analyses were employed. The results of the Shapiro–Wilk analysis confirmed that the data were not normally distributed and thus non-parametric analyses were utilised. Two methods of association were applied: Pearson’s chi-square test of independence and Mann–Whitney \(U\) analysis. Chi-square was utilised where associations were analysed between the dependent variable (waterborne outbreaks, yes (1), no (0)) and the categorical independent variable (heavy rainfall in the preceding 7 days, yes (1), no (0)). Where the independent variable was continuous (cumulative rainfall and temperature), Mann–Whitney \(U\) analysis was performed, which assesses differences in means based on a categorical dependent variable. The results of the multivariate analysis are presented in Table 3. Significant \((p < 0.05)\), positive associations were

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of outbreaks</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>15</td>
<td>75</td>
</tr>
<tr>
<td>2011</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2009</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>2008</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>2007</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>2006</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2005</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3 | Multivariate analysis between the dependent and independent variables

<table>
<thead>
<tr>
<th>Variable name</th>
<th>(N)</th>
<th>Test statistic</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative monthly rainfall</td>
<td>96</td>
<td>4.638&lt;sup&gt;a&lt;/sup&gt;</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Heavy rainfall (in previous 7 days)</td>
<td>96</td>
<td>26.214&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mean monthly temperature</td>
<td>96</td>
<td>3.779&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<sup>a</sup>Mann–Whitney \(U\).

<sup>b</sup>Pearson’s chi-square.
found to exist between all three independent variables and the dependent variable, and thus are suitable for inclusion in the LR through purposeful selection as proposed by Hosmer et al. (2013).

Consequently, using all three independent variables, LR analysis was performed, the results of which are shown in Table 4. A positive regression coefficient indicates a positive correlation between a significant explanatory variable and a waterborne VTEC outbreak, while a negative coefficient suggests an inverse or negative correlation. Overall, the results indicate a strong association between all three predictor variables: cumulative monthly rainfall, heavy rainfall events in the preceding 7 days and mean monthly temperature.

The LR model, inclusive of the three predictor variables demonstrated an overall significance (p) of <0.001 indicating that the model has good predictive capacity in relation to waterborne VTEC outbreaks. Cumulative monthly rainfall and heavy rainfall events in the preceding 7 days made the most significant contribution to the LR model (p < 0.001), returning odds ratios (Exp(B)) of 1.047 and 22.890, respectively. In terms of cumulative rainfall, these results indicate that for every 1 mm increase in rainfall, the probability of a waterborne VTEC outbreak increases by a factor of 1.047. For heavy rainfall events, in which the variable had a binary categorisation (1: heavy rainfall event, 0: no heavy rainfall event), the odds demonstrate that where there is a heavy rainfall event in the previous 7 days, the probability of a waterborne VTEC outbreak increases by 22.890, a significant finding. Similarly, temperature played a significant role in the LR model (p = 0.005), demonstrating that for every 1 degree Celsius increase in temperature, the probability of a waterborne VTEC outbreak increases by a factor of 1.370.

In terms of calibration, p-values >0.05 for the HL statistic, indicate good calibration and thus the high p-values for the HL statistic in this model (p = 0.501) confirm good model calibration. Hence, we can say the models discriminate and calibrate well for the target-dependent variable: waterborne VTEC outbreaks. Collinearity diagnostics were run on all the predictor variables utilised in the model and thus the tolerance and variance inflation factor (VIF) was examined for each variable. No variable has a tolerance less than 0.4 or a VIF greater than 2.5 and subsequently this indicates that the variables do not have collinearity issues.

DISCUSSION

This study represents the first quantitative analysis of the relationship between meteorological phenomena, namely: cumulative rainfall, heavy rainfall events and temperature, and waterborne VTEC outbreaks at the national level and over an extended period. Our findings show a statistically significant association between all three variables and disease outbreaks. However, we acknowledge that multiple factors are involved, which must occur simultaneously in time and space. Elements of an outbreak event include: (1) a source of contamination (infected humans, domestic animals or wildlife); (2) fate and transport of the contaminant from source to water ecosystems; (3) inadequate treatment of drinking water or inadequate monitoring of recreational water; and (4) investigation into the source of outbreaks (Curriero et al. 2001). Given the variability of these factors across the Republic of Ireland, the robustness of our findings demonstrates the important role of meteorological events in

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>p</th>
<th>Odds ratio</th>
<th>95% C.I. for odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative rainfall</td>
<td>0.046</td>
<td>0.013</td>
<td>12.868</td>
<td>&lt;0.001</td>
<td>1.047</td>
<td>1.021 - 1.073</td>
</tr>
<tr>
<td>Heavy rainfall</td>
<td>3.131</td>
<td>0.861</td>
<td>13.211</td>
<td>&lt;0.001</td>
<td>22.890</td>
<td>4.231 - 23.826</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.315</td>
<td>0.112</td>
<td>7.854</td>
<td>0.005</td>
<td>1.370</td>
<td>1.099 - 1.707</td>
</tr>
<tr>
<td>Constant</td>
<td>-10.729</td>
<td>2.577</td>
<td>20.379</td>
<td>&lt;0.001</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

B = the coefficient of the predictor variables; S.E. = standard error; Wald = value of the coefficient divided by standard error; Exp(B) = the base of natural logarithms for the predictor coefficients; C.I. = confidence interval.
microbial fate and transport of VTEC as a contributing factor in Irish waterborne disease outbreaks. However, it must be asserted that meteorology is likely one of many factors that lead to an outbreak.

Our study is limited by the availability of waterborne disease outbreak data, for which reporting was only legislated in 2004. As determining the route of infection is a difficult task, it is estimated that we are only highlighting a small fraction of the actual waterborne outbreak figures (Frost et al. 1996). Future studies incorporating more data will aid in an increased understanding of the relationship between rainfall and VTEC epidemiology.

The results presented in this paper are consistent with findings from other studies relating to waterborne infectious diseases. For example, Wilkes et al. (2009) found that concentrations of pathogenic bacteria increased with an increase in rainfall. Similarly, Atherholt et al. (1998) found that Cryptosporidium oocysts and Giardia cysts in the Delaware River were positively correlated with rainfall. Literature relating E. coli organisms and rainfall is scarce, as E. coli is typically seen as a foodborne pathogen. However, the largest reported outbreak of E. coli O157:H7 occurred at a fairground in the state of New York (USA) in September 1999 and was linked to contaminated well water, and was preceded by an unusually heavy rainfall event (Patz et al. 2000). The mechanisms whereby rainfall might contribute to outbreaks through increased contamination of source waters are straightforward. Heavy ‘flash’ rainfall or periods of prolonged rainfall can mobilise pathogens within the environment increasing runoff from agriculture and transporting this microbiologically contaminated medium into rivers, coastal waters and groundwater wells. Similarly, temperature has been shown to play a dual role in infectious disease. Transmission of enteric disease can possibly be increased by high temperatures by the direct effect on the growth rate of an organism in the environment (Semenza & Menne 2009) and this is upheld by the results of this study, which demonstrated a relationship between increased temperature and waterborne VTEC outbreaks. The significance of temperature also may signify a link, with greater faecal shedding in warmer months associated with agricultural practice (Michel et al. 1999; Money et al. 2010). This is particularly important as the farming of ruminants is implicated as a major transmission route (O’Brien et al. 2001; Voetsch et al. 2007).

In summary, there is mounting evidence that meteorological parameters contribute to the risk of waterborne VTEC outbreaks in Ireland. The results offer useful insight into the links between waterborne disease and meteorology, a particular area of interest as waterborne diseases are predicted to increase in relation to future climate change scenarios (Rose et al. 2001). The information gathered can be used to inform public health professionals on the causes and factors contributing to outbreaks, to target prevention strategies and to monitor and direct the effectiveness of prevention programmes into the future.

CONCLUSION

Waterborne outbreaks of VTEC are a worrisome transmission route of the pathogenic organism. The objective of this paper was to assess the relationship between rainfall events (cumulative and heavy), temperature and the incidence of waterborne outbreaks. The study has found that rainfall, particularly heavy rainfall in the preceding 7 days of an outbreak, is a strong statistical indicator of an outbreak event and that rising temperature also plays a role in VTEC outbreaks. Utilising the information within this paper, we aim to inform public health professionals on the role of meteorology in the epidemiology of infectious disease.

ACKNOWLEDGEMENTS

We wish to acknowledge the many microbiologists, environmental health specialists and surveillance scientists who have investigated and reported waterborne outbreaks. Without their efforts, valuable insight into epidemiology and disease prevention and control would be lost. We would particularly like to thank our partners at the Health Protection Surveillance Centre for allowing us access to data and for their expertise. There are no conflicts of interests to declare.

REFERENCES

H7 associated haemolytic uraemic syndrome in a nurse. Lancet 331 (8584), 526.


First received 16 January 2015; accepted in revised form 26 May 2015. Available online 7 July 2015.