

2018

Development of a hierarchical model for predicting microbiological contamination of private groundwater supplies in a geologically heterogeneous region

Jean O'Dwyer

Paul Hynds

Kenneth Byrne

See next page for additional authors

Follow this and additional works at: <https://arrow.tudublin.ie/ehsiart>



Part of the [Medicine and Health Sciences Commons](#), and the [Water Resource Management Commons](#)

This Article is brought to you for free and open access by the ESHI Publications at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, gerard.connolly@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](#)

Authors

Jean O'Dwyer, Paul Hynds, Kenneth Byrne, Michael Ryan, and Catherine Adley



Development of a hierarchical model for predicting microbiological contamination of private groundwater supplies in a geologically heterogeneous region[☆]

Jean O'Dwyer^{a,*}, Paul D. Hynds^b, Kenneth A. Byrne^c, Michael P. Ryan^d, Catherine C. Adley^d

^a School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland

^b Environmental Health and Sustainability Institute, Dublin Institute of Technology, Dublin, Ireland

^c Department of Biological Sciences, University of Limerick, Limerick, Ireland

^d Department of Chemical Sciences, University of Limerick, Limerick, Ireland

ARTICLE INFO

Article history:

Received 19 June 2017

Received in revised form

14 February 2018

Accepted 17 February 2018

Available online 28 February 2018

Keywords:

Groundwater

Contamination

Ireland

Regression modelling

E. coli

ABSTRACT

Private groundwater sources in the Republic of Ireland provide drinking water to an estimated 750,000 people or 16% of the national population. Consumers of untreated groundwater are at increased risk of infection from pathogenic microorganisms. However, given the volume of private wells in operation, remediation or even quantification of public risk is both costly and time consuming. In this study, a hierarchical logistic regression model was developed to 'predict' contamination with *E. coli* based on the results of groundwater quality analyses of private wells ($n = 132$) during the period of September 2011 to November 2012. Assessment of potential microbial contamination risk factors were categorised into three groups: Intrinsic (environmental factors), Specific (local features) and Infrastructural (groundwater source characteristics) which included a total of 15 variables. Overall, 51.4% of wells tested positive for *E. coli* during the study period with univariate analysis indicating that 11 of the 15 assessed risk factors, including local bedrock type, local subsoil type, septic tank reliance, 5 day antecedent precipitation and temperature, along with well type and depth, were all significantly associated with *E. coli* presence ($p < 0.05$). Hierarchical logistic regression was used to develop a private well susceptibility model with the final model containing 8 of the 11 associated variables. The model was shown to be highly efficient; correctly classifying the presence of *E. coli* in 94.2% of cases, and the absence of *E. coli* in 84.7% of cases. Model validation was performed using an external data set ($n = 32$) and it was shown that the model has promising accuracy with 90% of positive *E. coli* cases correctly predicted. The developed model represents a risk assessment and management tool that may be used to develop effective water-quality management strategies to minimize public health risks both in Ireland and abroad.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

It is estimated that private groundwater sources in the Republic of Ireland provide domestic drinking water to approximately 750,000 people or 16% of the population, with many thousands more served on a non-domestic, intermittent basis (CSO, 2012a). Similarly, private wells supply 4.1 million Canadians (11.7%) and 45 million Americans (14.1%), with private wells constituting the

largest proportion of water wells in both countries (Hynds et al., 2014b, Murphy et al., 2016). A further 200–500 million Europeans are reliant on groundwater for their domestic supply (Job, 2010). Private groundwater supplies in Ireland typically comprise two source types, namely, small private supplies (SPS) serving individual households, and private group water schemes (PrGWS), which typically supply <50 persons and/or <10 m³d⁻¹. Both supply types are exempt from the European Commission Drinking Water Directive (DWD) 98/83/EC (EU 1998) and hence water treatment is entirely voluntary, and if employed, the sole responsibility of the owner/caretaker. Hynds et al. (2013), report that 64% of interviewed private well owners ($n = 245$) in the Republic of Ireland did not

[☆] This paper has been recommended for acceptance by Dr. Jorg Rinklebe.

* Corresponding author.

E-mail address: Jean.ODwyer@ucc.ie (J. O'Dwyer).

utilise an appropriate domestic water treatment system.

Rural Ireland is home to approximately 38% (1.76 million) of the national population, and is characterized by a heavily dispersed yet locally dense settlement pattern (Scott and Murray, 2009), with most settlements comprised of individual private or “one-off” dwellings situated outside urban administrative zones (Duffy, 2000, CSO, 2012b). The spatial distribution of private groundwater supplies thus exhibits a marked Urban:Rural divide accredited to the reduction in piped infrastructure, in concurrence with decreased population density (Óhaiseadha et al., 2017). Moreover, SPS reliance in rural Ireland is increasing; during the period 2006–2011, SPS use rose from 74.3% to 76.1%, likely due to accelerated private property development prior to the global recession (CSO, 2012b). Subsequently, due to the combined effects of high private groundwater source reliance, the rural ubiquity of private domestic wastewater treatment systems and pastoral agriculture, a temperature maritime climate, and diverse bedrock and quaternary geology, Ireland may be considered to represent the “perfect storm” with respect to groundwater susceptibility to contamination.

The association between groundwater contamination and waterborne disease has long been acknowledged, with Dr John Snow (1813–1858) establishing the source of a significant cholera outbreak in the London district of Soho during the summer of 1854, as being from a faecally contaminated public well (Donaldson, 2002). More recently, Murphy et al. (2017) have presented clear epidemiological evidence of the transmission of disease through groundwater contamination on a global scale. A review of waterborne outbreaks of enteric infection in the Nordic region over a 15-year period (1998–2012) by Guzman-Herrador et al. (2015) reports that 76% (n = 124) of outbreaks with a confirmed source were associated with groundwater. Overall, Guzman-Herrador et al. (2015) found that a majority of outbreaks were associated with single household (i.e. private) water supplies, while 35% of outbreak clusters were confined to a single household. Moreover, where factors contributing to the waterborne outbreak had been confirmed, 96% of outbreaks associated with private groundwater supplies were deemed to have occurred due to source (well) contamination. Similarly, Pitkänen et al. (2015) note that the presence of *Aeromonas* and *Giardia* among a cohort of 20 vulnerable small (<500 consumers) groundwater supplies in Finland, highlights the significant potential adverse health effects of pathogen ingress to groundwater sources. In the United States (Wallender et al., 2014), have found that 30.3% of 818 drinking water outbreaks reported to the US Centre for Disease Control (CDC) between 1971 and 2008 were attributable to untreated groundwater sources.

Of particular significance within the Irish context is the prevalence of verotoxigenic *E. coli* (VTEC) infection which has been linked to private water supplies (Hynds et al., 2014a, O'Dwyer et al., 2014, Óhaiseadha et al., 2017). Ireland has had the highest crude incidence rates (CIR) of VTEC in Europe, increasing from 3.9/100,000 in 2007 to 15.3/100,000 in 2013 (HPSC, 2015). Óhaiseadha et al. (2017) report that during the period 2008–2013, private well usage was significantly associated (OR 6.896, $p < 0.001$) with the incidence of confirmed primary VTEC O157 infection, while Hynds et al. (2014a) predict an endemic VTEC CIR of 28.3/100,000 private well users per annum; as much as 5–6 times that of the national population. Furthermore, recent work by O'Dwyer et al. (2017) reports that 21.4% of *E. coli* isolates recovered from a cohort of private wells in the mid-west of Ireland exhibited resistance to ≥ 1 human antibiotic, with 100% of isolates presenting resistance to ≥ 1 veterinary antibiotic. Accordingly, there is little doubt that the human health risks posed to private groundwater consumers in the Republic of Ireland are potentially significant, and likely increasing.

However, due to the dispersed, decentralised nature of these sources in Ireland and abroad, in addition to their abundance, appropriate monitoring, maintenance, and remediation is both complex and financially prohibitive.

Groundwater microbial quality may be affected by myriad environmental and source-specific risk factors, including well design, location and maintenance, septic system location and maintenance, local hydrogeological setting, and significant climatic events (e.g. flooding, snowmelt, etc.) (Hynds et al., 2012; Wallender et al., 2014; Atherholt et al., 2017; Andrade et al., 2018). Moreover, groundwater pathogens may derive from multiple human or animal faecal sources such as adjacent septic systems, livestock grazing, manure spreading and/or farmyards (Kozuskanich et al., 2014; Wallender et al., 2014). Recent work has shown that up to 70% of source contamination occurs via “localized” pathways, as opposed to “generalized” aquifer contamination (Hynds et al., 2012, 2014a, b), with groundwater contamination risk (susceptibility) typically increasing in areas characterized by high hazard (i.e. faecal source) densities in concurrence with inappropriate setback distances/gradients, with this relationship mediated by local hydrogeological characteristics e.g. hydraulic conductivity, aquifer productivity, etc. (Bremer and Harter, 2012; Hynds et al., 2014a, b). A relatively recent study of monitoring wells in eastern Ontario detected the presence of both human and animal pharmaceuticals in groundwater, suggesting that faecal indicator bacteria (FIB) were from both human and animal sources (Kozuskanich et al., 2014). While well siting and construction regulations exist in an effort to prevent drinking water contamination, waterborne AGI outbreaks may still occur where infrastructure predates regulation, or when assessments do not properly account for the geological vulnerability of the well area (Bremer and Harter, 2012). Similarly, Atherholt et al. (2017) have shown FIB detection rates are significantly reliant on recent climatic conditions, and particularly local 10-day cumulative antecedent precipitation, in addition to groundwater physical chemistry.

Consequently, there is a need for identifying areas of concern with respect to effective custodianship and development of focused environmental health policy. Accordingly, the current study sought to identify and quantify the associated environmental and social factors affecting the susceptibility of private water wells to contamination. The spatial and temporal distribution of groundwater *E. coli* presence in the mid-western region of Ireland has been examined via a mixed methods approach which comprised a 13-month field sampling programme, a formalised source owner survey, and geostatistical analysis. Collated risk factor variables have been characterised as increasing the Intrinsic (environmental factors), Specific (local (sampling area) features) or Infrastructural (groundwater source and domestic wastewater treatment system characteristics) susceptibility to contaminant ingress, and subsequently employed to develop a predictive hierarchical logistic model of private source contamination. It is considered that the employed approach and subsequent findings are internationally transferable, and may be used by water resource managers, private well owners/users, and local/national governments to support evidence-based risk management and develop quantitative source protection strategies.

2. Materials and methods

2.1. Study area

The study region is situated in the mid-west of Ireland, extends 8248 km² (11% of total area of the Republic of Ireland), and comprises three administrative counties (Limerick, Clare and North Tipperary) (Fig. 1). Regional climate is described as temperate

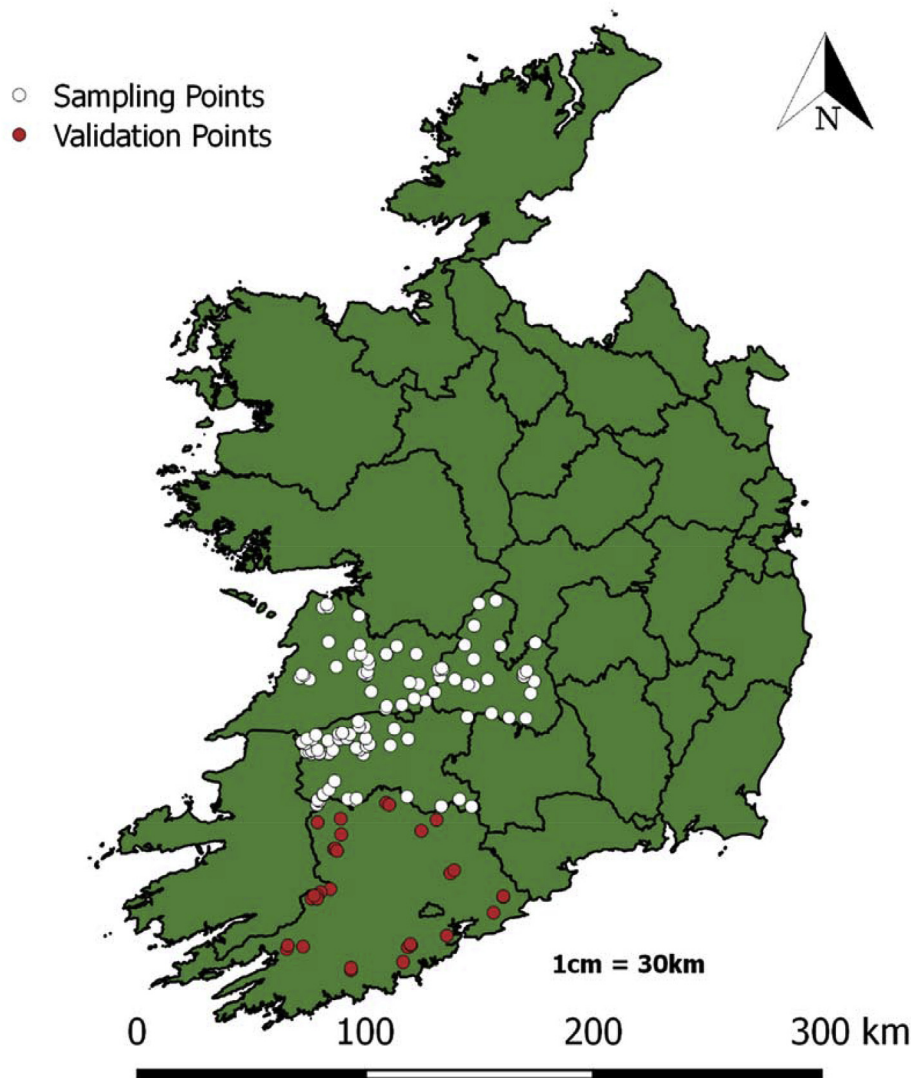


Fig. 1. Map of the study area and distribution of private wells analysed in the primary sampling regime ($n = 132$) and the validation sampling regime ($n = 32$) within the Republic of Ireland.

maritime, similar to that of the rest of the country, however, both annual precipitation (30-year Annual Mean 977.6 mm) and relative humidity (30-year Annual Mean 71.9%) are higher than the national average due to the coastal Atlantic location, in addition to lower mean annual temperatures (10.7 °C) (Met Eireann, 2016). The region is characterised by high private groundwater reliance; as per the 2011 National Census, the area had a total population of 361,028, of which 54.7% are categorically rural ($n = 197,408$), with 23,014 private groundwater sources in operation, most of which (98.4%, $n = 22,651$) are in categorically rural areas (CSO, 2012b).

As previously described by O'Dwyer et al. (2017), the study area bedrock geology is dominated by bedded and un-bedded Dinantian limestones and Devonian sandstones, with some volcanic and shale deposits. This diversity is mirrored by regional Quaternary geology; (limestone, sandstone, and shale) tills dominate in Limerick (South) and North Tipperary (East), with much of Co Clare (West) characterised by thin (subcrop) or absent (outcrop) overburden.

2.2. Groundwater sampling and analysis

As previously set out by O'Dwyer et al. (2014), study participants

were identified and recruited by dissemination of a study overview and recruitment notice on a popular Irish bulletin board website (www.boards.ie) and via the Irish Farmers Association (IFA). Overall, 132 study participants took part in this research.

Sampling was undertaken in three phases during the 13-month period September 2011 to November 2012 (Table 1), with specific attempts made to sample each source after a period of significant rainfall on at least one occasion where possible. Sampling was carried out in accordance with Standard Methods for the Examination of Water and Wastewater methods (APHA, 2005) from all three counties in the research area as follows: Limerick ($n = 66/132$), Clare ($n = 29/132$) and North Tipperary ($n = 37/132$), the geographical distribution of which are shown in Fig. 1.

Untreated samples were taken directly from a pre-sterilised (70% ethanol) kitchen tap after a 2-minute flushing period, and collected in sterile sampling bottles. Samples (100 mL) were collected in disposable 120 mL sterile vessels. Samples were immediately transferred to a cool box and transported to a laboratory, with analysis undertaken within 4–6 hrs. All samples were assayed for *E. coli* using a standard US Environmental Protection Agency (EPA) approved commercial culture kit (Colilert, IDEXX

Table 1
Intrinsic Specific Infrastructural (ISI) Vulnerability Assessment: Description of the 15 variables considered and assessed for model inclusion.

Vulnerability Category	Variable	Data Source
Intrinsic	Subsoil Permeability	Geological Survey Ireland (GSI) Mapping
	Aquifer Vulnerability	GSI Mapping
	Aquifer classification	GSI Mapping
	Karst Bedrock	GSI Mapping
	Estimated Recharge	GSI Mapping
	5 Day Rainfall (mm)	Met Eireann
	5 Day mean temp (°C)	Met Eireann
Specific	Land Cover	Environmental Protection Agency Geoportal
	# of Farms per Small Area	CSO ^a
	Livestock Numbers ^c	CSO ^a
	# of DWWTS per Small Area	CSO ^b
	Population Density	CSO ^b
Infrastructural	Well Type	Questionnaire
	Well Depth	Questionnaire
	Well Age	Questionnaire

^a Agricultural Census 2010 (CSO, 2010).

^b National Census 2011.

^c Both Cattle and Sheep; DWWTS, Domestic Waste Water Treatment Systems.

Laboratories Inc., Westbrook, ME, USA) and in accordance with manufacturer's directions. Negative controls (sterile deionised water) were used during all phases of laboratory analyses.

2.3. Intrinsic, specific, and infrastructural (ISI) vulnerability assessment: categories and variables

To aid in the development of an evidence-based and hydrogeologically logical predictive model for contamination of private groundwater sources in the study region, three distinct categories of vulnerability influencing variables were considered; Intrinsic, Specific and Infrastructural, as follows:

- **Intrinsic** variables focused on specific hydrogeological conditions that provide some measure of defence against external contamination. Importantly, intrinsic variables are independent of the nature of the specific contaminants and the contamination source. Intrinsic factors include the hydrological, geological and hydrogeological characteristics of the area which have been shown to impact groundwater vulnerability (Zwahlen, 2004; Hynds et al., 2012; O'Dwyer et al., 2014).
- **Specific** variables were classified as potential sources of faecal contamination as reported in literature. In particular, agricultural practices, particularly livestock density, have been shown to be a significant source of groundwater pollution (Richards et al., 1996; Goss et al., 1998; Close et al., 2008). In addition, sources of human wastes found in groundwater include effluent from on-site sanitation (septic) systems and community wastewater treatment systems (Yates, 1985; Arnade, 1999; Hynds et al., 2012, O'Dwyer et al., 2017).
- **Infrastructural** variables are an important inclusion in this model and facilitate in the evaluation of the design, construction and placement of private water wells which can guide best practice. For this study, based on information provided by participants and sample site assessment, infrastructural considerations including well type, well age and well depth have been considered, which have been shown to influence human vulnerability to contaminated groundwater consumption (Gonzales, 2008)

In total, 15 variables (Table 1) were considered for the development of the predictive model. "Intrinsic" and "Specific" data were

primarily sourced from existing national databases (Table 1). The geographical coordinates of each sampling point were acquired using a global positioning system (GPS) (Garmin nüvi® 3790T) and added to a Geographic Information System (ArcMap10). Where the data were only available as a quantitative measure (e.g. meteorological data), all available synoptic stations ($n = 65$) were added to the basemap, with a buffer created to link sampling sites with the nearest synoptic station. For the current study, the aquifer systems have first been defined as either bedrock or a sand/gravel aquifer. Sand/gravel aquifers have been employed as the indicator variable, and thus all odds ratios refer to this classification. Bedrock aquifers were subsequently subdivided into 'Fissured' and 'Karstic' systems, the aim being to predict groundwater vulnerability as a function of local hydrogeological features. Subsoil permeability was ordinalised (ranked and coded) for model inclusion, ranging from Low permeability (#1) to 'thin or absent (NA)' (#4). Groundwater recharge estimates were collated from the national Groundwater Recharge Map (Williams et al., 2013), which is derived from existing hydrogeological and meteorological data layers, including annual rainfall, annual estimated actual evapotranspiration (AE), soil drainage, subsoil permeability, groundwater vulnerability, peat, sand/gravel aquifer, and bedrock aquifer class. For meteorological data, total five day antecedent rainfall (mm) and temperature (°C) were calculated by geo-locating the sampling point to the nearest synoptic station; five day totals were utilised to allow for contamination ingress and bacterial growth while also allowing for weather variability.

Data pertaining to "Specific" vulnerability themes (e.g. domestic wastewater treatment type and human population density) were extracted from the CSO Census of Ireland 2011 and Census of Agriculture 2009 datasets. CSO Census data have been compiled and spatially indexed using the Irish postal service's GeoDirectory geographic information system (<https://www.geodirectory.ie/>) to 18,488 pre-defined Census enumeration areas ('Small Areas') (mean land area: 3.8 km²), the smallest legally defined administrative areas in Ireland, and are the highest-resolution geographical unit available for statistical compilation at the national (ROI) level, in compliance with current data protection standards (50–200 domestic dwellings). Small Area statistics were joined to a small area shapefile and the data layers were then intersected with the sampling points to create a novel database of the relative data pertaining to the point of sampling. Similarly, data from the Census

of Agriculture, which was completed in 2009 for all agricultural holdings in the State with a 'farmed area' >2.47 acres, were accessed and used to geo-statistically aggregate and calculate cattle and sheep numbers for each Census "Small Area".

Infrastructural data (i.e. source type, depth, and age) were collated via a participant questionnaire, which was completed with all well owners during the first sampling phase. Source depth was coded and entered for the Intrinsic Specific Infrastructural Logistic regression (ISI-LR) model as a discretized ranking variable (1 = <5 m, 2 = 5–20 m, 3 = 21–50 m, 4 = 51–100 m). Well type was treated as a dichotomous categorical variable, with hand-dug wells employed as the indicator or reference variable for model development. A full codebook for the dataset can be found in the [supplementary material](#).

2.4. Statistical analysis

Prior to analyses, all independent variables were evaluated for normality via Q-Q plots and Shapiro-Wilkes tests. Numerous variables exhibited a non-normal distribution, thus non-parametric analyses have been employed for all subsequent analyses. Univariate analyses (risk factor assessment) have been undertaken using Mann-Whitney U or Chi-square tests, as appropriate. Logistic regression (LR) models were developed using *E. coli* presence/absence (0/1) during the study period as the dependent or modelled variable. Parameter entry to initial lumped regression models was undertaken using a "forced entry" approach, with all variables simultaneously analysed. Subsequently, upon development of an efficient model, ISI-LR models were re-run using three (i.e. Intrinsic, Specific, and Infrastructural) vulnerability-based hierarchies (variable entry blocks). This approach permits hierarchical evaluation of each vulnerability category in terms of individual significance within the finalised model, and explained variance within the overall system (Well Cohort).

Hierarchy evaluation was undertaken using the associated Chi-square statistic, with statistically significant increases between hierarchy additions indicative of increasing model efficiency (predictive capacity). Backward elimination ($p > 0.1$) of variables that contributed least to the model was employed for hierarchical model development. The collinearity diagnostic test for tolerance (<0.1) and the variance inflation factor (VIF) (>10) were used to assess collinearity between independent variables prior to model development. The Hosmer Lemeshow test was used to validate model goodness-of-fit, with Nagelkerke's pseudo R^2 used to estimate effect size and explained system variance. IBM SPSS® 22 was employed for all statistical analyses, with a confidence level of 95% ($\alpha = 0.05$) employed throughout by convention.

2.5. Model validation

To validate the model, an external model validation protocol was undertaken using the results of a groundwater analysis regime ($n = 32$) which took place in County Cork, Ireland ([Fig. 1](#)) during the months September 2013 to February 2014 and included Infrastructural descriptions in line with the developed model (well depth, age and type). Variables relative to the Intrinsic and Specific categories were geospatially linked to the geographical coordinates of each sampling location, as previously described. The external model validation was undertaken by substituting the secondary sampling data (validation data) into the developed ISI-LR equations as model coefficients. A classification cut-off of 0.5 was used, with a test model output of <0.5 recorded as a negative prediction and an output of >0.5 recorded as a positive prediction. The observed and predicted frequencies were then compared in order to provide an estimate of model validity.

3. Results

3.1. Temporal groundwater sampling

Results of microbial groundwater sampling and analyses over the 13-month sampling regime are presented in [Table 2](#). As shown, during the first sampling period, *E. coli* was detected in 59.8% ($n = 78$) of supplies. During the second sampling period, the prevalence of *E. coli* decreased, with 51.5% ($n = 68$) of supplies testing positive; a reduction of approximately 8% ($n = 10$). Between sample periods 2 and 3, several ($n = 30$) cohort households which had previously been informed of a positive *E. coli* sample had domestic water treatment systems installed and were thus excluded from further sampling. Accordingly, due to domestic intervention, these households have been omitted from the cumulative contamination percentage, reducing the total number of supplies to 102 and number of samples to 366. During the third sample period, *E. coli* was detected in 31.8% ($n = 42$). Overall, 51.4% ($n = 188$) of private well samples taken over the 13-month sampling duration tested positive for *E. coli*, while 70.45% (93/132) of sampled wells tested positive on at least one occasion.

3.2. Risk factor (univariate) analysis of microbial contamination

Prior to logistic regression model development, all variables within each vulnerability category underwent preliminary statistical analysis to determine their level of univariate association with the dependant variable (*E. coli* presence/absence), and thus prioritise their inclusion and entry for modelling as a likely contamination risk factor. This method of "purposeful variable selection" via pre-analysis of independent variables was proposed by [Hosmer et al. \(2013\)](#). The full descriptive statistics of the sampling results across variable categories are detailed within the [supplementary material](#).

As shown ([Table 3](#)), all variables within the Intrinsic vulnerability category were significantly associated ($p < 0.05$) with *E. coli* presence, and were thus included in LR modelling. Highest levels of univariate significance were exhibited by subsoil permeability ($p < 0.001$), precipitation ($p = 0.001$), and the presence of karstified bedrock ($p = 0.001$). Three categorically "Specific" vulnerability variables were found to have significant associations with the dependent variable, namely, the number of farms ($p = 0.002$), the number of cattle ($p = 0.001$), and the number of DWWTSS ($p = 0.001$) per Small Area. District. Accordingly, these variables were prioritised for inclusion in LR model development. With respect to infrastructural vulnerability, two of the three categorised variables exhibited a significant association with *E. coli* presence; well depth ($p = 0.003$) and well type ($p = 0.032$) were thus included for model development. Overall, eleven of the fifteen variables analysed within the vulnerability categories (Intrinsic (6), Specific (3), and Infrastructural (2)) were deemed statistically robust for model development.

Table 2
E. coli presence during 13-month temporal sampling regime.

Sample Period	<i>E. coli</i>	
	N	%
One (Oct-Dec 2011) ^a	78	59.8
Two (Mar-Jun 2012) ^a	68	51.5
Three (Jul-Nov 2012) ^b	42	41.5
Total ^c	188	51.4

^a 132 samples.

^b 102 samples.

^c 366 samples.

Table 3
Univariate risk factor analysis of independent (input) vulnerability variables (n = 15).

Variable Name	Variable Coefficient	Significance (p) ^c
Subsoil Permeability	19.95 ^a	<0.001**
5 Day Rainfall (mm)	3.193 ^b	0.001**
5 Day mean temp (°C)	2.139 ^b	0.032*
Aquifer classification	4.181 ^a	0.041*
Karst Bedrock	10.835 ^a	0.001**
Estimated Recharge	1.987 ^b	0.047*
Land cover	0.029 ^a	0.866
# Cattle per Small Area	3.134 ^b	0.002**
# Sheep per Small Area	-0.496 ^b	0.620
# of Farms per Small Area	3.182 ^b	0.001**
# DWWTS per Small Area	3.48 ^b	0.001**
Population Density	-0.808 ^b	0.419
Well Type	4.589 ^a	0.032*
Well Depth	14.278 ^a	0.003**
Well Age	2.607 ^a	0.272

^a Pearson's Chi Square.

^b Mann Whitney U, * p < 0.05, ** p < 0.01; DWWTS, Domestic Waste Water Treatment Systems.

^c Variables with a P value < 0.05 were deemed suitable for inclusion in the hierarchical logistic model.

3.3. Intrinsic, Specific and Infrastructural Logistic Regression (ISI-LR) model development

Following univariate analysis, 11 variables were initially included in the ISI-LR model prior to the application of backwards elimination, whereby variables which contributed least to the model ($P > 0.1$) were removed ($n = 3$). In total, 8 of the 11 variables were deemed statistically satisfactory for inclusion in the final model (Table 4). Overall, model predictive sensitivity was high, with *E. coli* presence correctly classified in 94.2% of samples, while *E. coli* absence was correctly classified in 84.7% of samples. The Hosmer and Lemeshow test for goodness of fit produced an insignificant p-value ($p = 0.620$); thus, as the p-value was (significantly) greater than alpha (0.05), the null hypothesis that the observed and expected event rates (*E. coli* present/absent) are matched within subgroups of the sample population is accepted. The final Nagelkerke coefficient of determination (cumulative R^2) was 0.81, thus the model input variables explain 81% of system variability. Accordingly, the developed ISI-LR model was well calibrated and considered an effective predictive tool for forecasting *E. coli* presence in private groundwater sources. (Multi)Collinearity diagnostics including Tolerance ($1 - R^2$) and calculation of the Variance Inflation Factor ($1/\text{Tolerance}$) were undertaken for all variables after model development. Results show that no individual variable was associated with a Tolerance < 0.4 or Variance Inflation

Factor > 2.5, thus indicating that (multi)collinearity was not an issue within the final ISI-LR model.

A hierarchical approach was taken for model development to specify a prioritised order of variable entry based on risk factor analysis, in addition to permitting testing the effects of specific predictor groups, independent of the influence of others (Section 2.3). The first predictor hierarchy employed for model development was the 'Intrinsic' vulnerability block, which comprised soil permeability, aquifer classification, presence of karstified bedrocks and precipitation; both temperature and recharge were eliminated as they did not contribute significantly. As shown (Table 4), this predictor hierarchy was significant within the model ($\chi^2 = 45.910$, $p = 0.001$), with intrinsic variables alone capable of correctly classifying (*E. coli* presence/absence) 77.6% of contaminated samples. Within the "Intrinsic" hierarchy, ordinal (ranked) subsoil permeability played a significant role ($p = 0.016$) in dependent variable prediction, with the calculated $\text{Exp}(\beta)$ (odds ratio) indicating that with each ranked increase in subsoil permeability e.g. low to moderate, moderate to high etc., the probability of *E. coli* presence increased by a factor of 3.252 (95% C.I. 1.250–8.463). Similarly, aquifer classification contributed significantly to the model ($p = 0.044$), with a negative (<1) odds ratio found (OR 0.231, 95% C.I. 0.101–0.707), indicating that wells situated in sand/gravel aquifers were substantially less likely to be contaminated. As might be expected, the presence of karstified bedrocks variable contributed significantly to the model ($p = 0.042$), exhibiting an odds ratio of 2.801 (95% C.I. 0.030–3.451), thus indicating that bedrock aquifers comprising karstic features are significantly more vulnerable to contamination. Precipitation was also a significant predictor variable ($p = 0.030$); for each 1 mm increase in 5-day antecedent precipitation, the probability of *E. coli* presence increased by a factor of 1.247 (95% C.I. 1.022–1.522). The second "Specific" predictor hierarchy contained geo-derived livestock (cattle) numbers and DWWTS numbers. Addition of this second model block was found to increase overall model effect size and model efficiency (87.9%; +10.3%), with the model remaining significant ($\chi^2 = 81.557$, $p < 0.001$). Total number of cattle per Census "Small Area" was associated with an odds ratio of 1.01 (95% C.I. 1.00–1.20). The second potential contaminant source within the final ISI-LR model was DWWTS number per Census "Small Area" ($p = 0.006$), with an associated odds ratio of 1.029 (95% C.I. 1.008–1.050).

The final developed ISI-LR model (Table 4), inclusive of all variable blocks remained statistically significant ($p = 0.003$), and was associated with correct classification of *E. coli* presence/absence in 90.7% of samples (i.e. final predictor block increased classification by 2.8%). Just two variables associated with the infrastructural hierarchy remained within the final model, namely, well type ($p = 0.047$) and well depth ($p = 0.021$), both of which contributed

Table 4
Final ISI-LR model of source contamination vulnerability.

	B	Sig. (p)	Exp(B)	95% C.I. for Exp(B)		
				Lower	Upper	
Intrinsic	Subsoil Permeability ^a	1.179	0.016	3.252	1.250	8.463
	Aquifer class ^b	-1.465	0.044	0.231	0.101	0.707
	Karst ^b	1.030	0.042	2.801	0.030	3.451
Specific	5 Day Rainfall (mm)	0.221	0.030	1.247	1.022	1.522
	# Cattle per Small Area	0.001	0.047	1.010	1.000	1.200
	# DWWTS per Small Area	0.028	0.006	1.029	1.008	1.050
Infrastructural	Well Depth ^a	-1.026	0.021	0.358	0.150	0.859
	Well Type ^b	2.174	0.047	8.797	6.856	30.433
	Constant	-9.804	0.027	0.000		

ISI-LR, Intrinsic, Specific and Infrastructural Logistic Regression; B, coefficient of predictor variables; Exp(B), Odds Ratio; C.I., Confidence Interval.

^a Categorical variable (binary).

^b Categorical variable (ordinal); DWWTS, Domestic Waste Water Treatments Systems.

significantly to the ISI-LR model. Well depth was associated with an odds ratio of 0.358 (95% C.I. 0.150–0.859). Accordingly, this indicates that as well depth increases between discretized categories, the likelihood of contamination decreases by a factor of 0.358. Results indicate that samples from hand-dug wells were over 8 times more likely to have *E. coli* present (OR 8.797, 95% C.I. 6.856–30.433).

3.4. External validation of the ISI-LR model

In total, 32 external private water well samples were used to validate the ISI-LR model. As outlined in Fig. 2, the ISI-LR model predicted that 12 of the samples would test negative for *E. coli* while 14 negative samples were observed; an accuracy of 85.7%. Conversely, the ISI-LR predicted 20 positive *E. coli* samples while 18 were observed; a 90% accuracy rate, which is comparable with the 90.7% of samples correctly classified within the final stage of the ISI-LR model. The model exhibited a higher level of accuracy with regard to the prediction of *E. coli* presence than absence and thus, is conservative in its estimation. However, it is important to note that the accuracy of the ISI-LR when applied to this validation dataset should not be overestimated as the dataset is limited due to a small sampling size and due to sampling being confined to the winter months when rainfall is higher than average in Ireland.

4. Discussion

The current study utilised a hierarchical approach in order to predict the likelihood of *E. coli* presence relative to distinct variable categories, namely intrinsic, specific, and infrastructural with the aim of improving current knowledge of contamination mechanisms while aiding well protection and maintenance.

As shown (Table 2), 51.4% of groundwater samples were positive for the presence of *E. coli* over the duration of the study period, well water contamination ranged from 41.5% to 59.8% over three sampling periods, and 70.5% were contaminated at least once over the 13-month period. This is in line with the most recently available

findings from the EPA groundwater monitoring network; during 2012, positive counts were detected at 104 (51%) of 205 monitoring locations (Bradley et al., 2015). Conversely, two previous studies of Irish groundwater quality have reported lower levels of contamination; Hynds et al. (2014a, b) found thermotolerant coliforms (TTC) present in 28.9% of 262 private wells from 5 study areas, while Bacci and Chapman (2011) report a contamination rate of 24% among 75 private wells in the south of the country (Co Cork). Notably, neither of these studies included regions characterised by karstified bedrock, which are significantly more vulnerable to contaminant ingress. This is also demonstrably higher than domestic well contamination rates reported internationally, for example, 33% of domestic wells analysed from 1993 to 2004 as part of a large-scale study in the United States tested positive for total coliforms (Embrey and Runkle, 2006). The higher levels reported in this study is testament to the variability within groundwater systems in Ireland and highlights the need for more routine sampling and monitoring. The developed ISI-LR model indicates that subsoil permeability plays a significant role ($p = 0.016$) in predicting *E. coli* presence; each categorical increase in subsoil category (e.g. low to moderate, moderate to high etc.) was shown to concur with a significantly increased (OR 3.252) probability of *E. coli* contamination. This finding supports the overarching concept of groundwater vulnerability classification currently employed in the Republic of Ireland; low groundwater vulnerability is assigned to areas with relatively thick (>10 m) layers of low-permeability subsoils, recharge rates are typically low, leading to increased attenuation of contaminants through subsoil horizons. Conversely, high/extreme vulnerability is typified by shallow layers of high-permeability subsoils, and subsequently, limited attenuation capacities (Daly and Warren, 1998; Fealy and Green, 2009). Previous work by Hynds et al. (2012) has shown that subsoil permeability and thus, by extension, groundwater vulnerability category, is a useful indicator of aquifer susceptibility to contamination; however its suitability with regard to well contamination is less clear, particularly in geologically homogenous regions. Accordingly, findings from the current study suggest that generalised (i.e. non-preferential aquifer

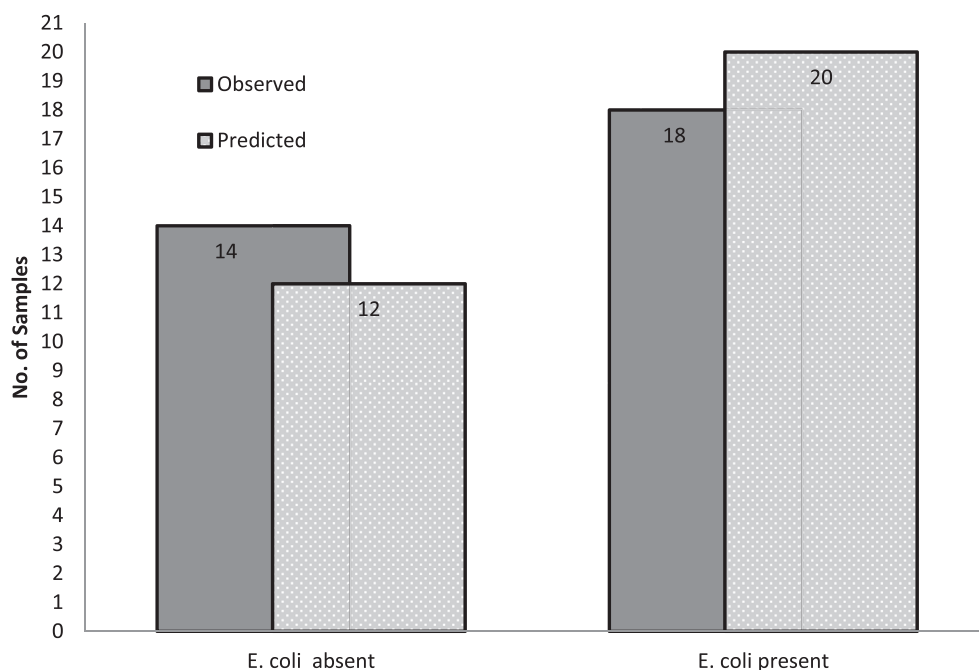


Fig. 2. ISI-LR Model Validation Classification (validation dataset) (n = 32).

recharge) contaminant mechanisms are likely responsible for a proportion of groundwater contamination in the study area.

Aquifer classification also contributed significantly to the model, with a negative effect exhibited (i.e. OR <1) ($p = 0.044$); ISI-LR modelling predicts that private wells situated within delineated sand/gravel aquifers were approximately 80% less likely to have *E. coli* present. This finding is of particular relevance in the Republic of Ireland as the two most important and common subsoil types in Ireland are glacial deposits (tills) and glaciofluvial sand and gravel deposits (Hynds et al., 2014a), with the sand and gravel aquifers which underlie approximately 2% of the country being the only aquifers with intergranular permeability. Furthermore, the presence of karstified bedrocks, which are associated with a large proportion of productive aquifers in Ireland (GSI, 2000), were also found to aid classification of *E. coli* presence ($p = 0.042$). Within the study area, the developed model predicts that wells located in characteristically karstic areas were almost three times more likely to have *E. coli* present (OR 2.801). This finding agrees with previous work; i.e. previous groundwater monitoring in Ireland has shown that groundwater monitoring locations in Karst limestone areas show the greatest degree of microbiological pollution (Bradley et al., 2015), thus reflecting the inherently vulnerable nature of dynamic flow systems, in concurrence with the lack of attenuation capacity associated with shallow soils or subsoils (Pronk et al., 2009). This is particularly the case with regard to diffuse (non-point) contamination in carboniferous limestone regions (Drew et al., 1996; Wallender et al., 2014).

Several previous studies have found a significant relationship between antecedent rainfall periods (mm) and the presence of thermotolerant coliforms (TTC) or *E. coli* (Howard et al., 2003; Hynds et al., 2012, 2014a; O'Dwyer et al., 2016); for example, Howard et al. (2003) have reported that increased rainfall over the 48-hr period prior to sampling increased the likelihood of TTC and faecal streptococci presence in shallow groundwater in Kampala, Uganda, while O'Dwyer et al. (2016) report an association between heavy rainfall events and outbreaks of VTEC in Ireland; a pathogen statistically associated with private well ownership (O'haiseadha et al., 2017). In both cases, the authors attribute the association between short-term rainfall to rapid (preferential) recharge and direct ingress at the wellhead due to overland flow, thus highlighting the role of meteorology in general, and rainfall in particular as an important driver of groundwater contamination. In the current study, 120hr (5d) precipitation was found to make a significant contribution to the model ($p = 0.030$), while, conversely, mean temperature over the same period was not efficient as a predictor variable ($p = 0.678$). Based on previous work by Hynds et al., (2012, 2014a) that reports this antecedent period as being significantly associated with the presence of TTC, this association is indicative of microbial bypass mechanisms, namely elevated overland flow (i.e. direct wellhead ingress) and increased shallow groundwater infiltration (i.e. preferential subsurface flow paths), with preferential subsurface flow paths the more likely mechanism. However, the authors acknowledge that the inclusion of rainfall as a predictive parameter may have limited applications. For example, the predictive capacity of antecedent rainfall is arguably localised; i.e. more accurate during specific and targeted sampling regimes as opposed to a broader assessment strategy at aquifer level. As a result, the inclusion of rainfall within the model should be considered on a case-to-case basis depending on the geographical scope and data availability.

Estimated recharge (mm) was not found to significantly contribute to the ISI-LR model ($p > 0.05$), albeit it was univariately significant ($p = 0.047$ (Table 3)). It is considered that this may be due to a number of data-related issues; firstly, the recharge data employed in the current study was not calculated specifically, but

instead retrieved from the national recharge data file which estimates recharge based upon historical effective precipitation. Secondly, where lower productivity aquifers underlie the land surface, a recharge cap is applied for recharge estimation to simulate rejected recharge, thus theoretically reflecting the limited ability of these aquifers to accept and transmit recharging waters (Williams et al., 2013). The application of 100 mm and 200 mm recharge caps in *Locally Important and Poorly Productive* aquifers, respectively, which comprise a significant proportion of the study area, may be too general, and thus unrealistic. Finally, as estimated recharge is inherently related to both precipitation and subsoil characteristics, it is reasonable to suggest that the presence of these variables in the ISI-LR model served to negate the predictive effect of recharge.

Potential contaminant sources were investigated via the "Specific" model hierarchy, which comprised GIS-derived, spatially distinct continuous measurements of cattle numbers, farm numbers, and DWWTS numbers per Census "Small Area". The total number of cattle per Census "Small Area" associated with sampling sites returned an odds ratio of 1.01 ($p = 0.047$); while this may appear modest, the effect size of this variable is based upon individual livestock units i.e. the addition of 1 livestock unit (cow) per "Small Area" increased the probability of *E. coli* presence by approximately 1%. The geo-derived variable "number of farms" did not contribute significantly to the final model ($p = 0.780$). The agri-food sector is characteristically diverse, ranging from tillage to mixed horticulture to pasture, however, results from the current study indicate that livestock farming, and specifically cattle husbandry, is the primary agricultural driver of private groundwater microbial contamination in Ireland. This is reflective of numerous international studies (Schets et al., 2005; Close et al., 2008; Kabore et al., 2010; Lenaker et al., 2017), for example, Schets et al. (2005) sampled 144 private groundwater supplies in The Netherlands, and found via molecular techniques (PFGE) that grazing cattle were the likely cause of contamination by *E. coli* O157:H7. It is estimated that 80% organic wastes generated in Ireland are derived from agricultural activities (Daly, 2000), with a 40ha dairy or beef farm producing waste volumes equivalent to that of a small town, with little, if any, treatment undertaken.

The number of DWWTSs per sampling Census "Small Area" represented the second major potential pollution source within the ISI-LR model ($p = 0.006$) (Table 4); this was found to be the most significant individual predictor within the finalised model. As for cattle, the "DWWTS/Small Area" variable was associated with a high mean input value, thus resulting in a seemingly low odds ratio (OR 1.029), however this is reflective of the nature of the input values, as opposed to the magnitude of their effects i.e. each additional DWWTS per Census "Small Area" increased the likelihood of *E. coli* presence by 2.9%. At present, there are as many as 450,000 domestic systems annually discharging 80 million m³ of effluent to the Irish subsurface (Daly, 2000; CSO, 2012b). Further, a recently initiated National Inspection Programme has found that DWWTSs are a primary source of faecal contaminants and are of poor standard in Ireland with over half (53%) of inspected DWWTSs failing first inspections (EPA, 2014).

Based upon findings from the "Specific" hierarchy, which improved model predictive efficiency by 10.3%, it may be concluded that multiple delivery mechanisms are causative in terms of *E. coli* contamination among the studied well cohort. This assertion agrees with a recent study by O'Dwyer et al. (2017) in the same study area; *E. coli* isolates ($n = 42$) from 125 untreated groundwater wells were analysed for the presence of human and veterinary antimicrobial resistance. Approximately one in five isolates (21.4%) exhibited resistance to ≥ 1 human antibiotic, while all isolates demonstrated resistance to ≥ 1 veterinary antibiotic. Thus, it may be

concluded that groundwater *E. coli* in this region are derived from both (point) human and (diffuse) animal sources, and likely gain ingress to wells both at the surface (wellhead) and in the subsurface (i.e. the production zone of the well). Similar findings have been reported by Lenekar et al., (2017) in the United States; analysis of 290 samples from eight rivers in the Great Lakes Basin showed that human and bovine viruses were present in 16.9% and 14.8% of runoff-event samples, and 13.9% and 12.9% of low-flow samples, thus pointing to multi-modal pathogen delivery within catchments.

Private well infrastructure (well type and depth) were deemed significant within the ISI-LR model (Table 4). Hynds et al., (2014a, b) have previously found that hand-dug well users in the Republic of Ireland are substantially more likely to contract waterborne enteric infections, a result which was borne out in the current study with hand-dug wells almost nine times more likely (OR 8.797) to have *E. coli* present ($p = 0.047$), primarily due to lower design and construction specifications. Similarly, well depth was an important predictor in terms of *E. coli* presence ($p = 0.021$); every ordinalised (10 m) increase in well depth concurred with a 35.8% decrease in the likelihood of contamination. Several previous studies have shown that well depth is representative of the required recharge distance for colloidal transport (Richards et al., 1996; Goss et al., 1998; Tabbot and Robson, 2006; Bahram et al., 2012). For example, Gonzales (2008) found that 71% of sampled wells <60 m tested positive for bacterial contamination, while an equivalent figure of <10% was encountered among wells >60 m depth.

The ISI-LR model was externally validated and demonstrated promising predictive capacity of 85.7% and 90% for *E. coli* absence and presence, respectively. The model was shown to have a greater predictive capacity for the presence of *E. coli*; overestimating the occurrence of the faecal organism. While this is a Type I error, it is also a conservative error, which, in the context of public health is preferable to a false negative (Type II). Consequently, it is considered that the ISI-LR model can be used in hydrogeologically similar regions in both Ireland and further afield as a management and risk assessment tool. While the relatively small sampling cohort presents a study limitation, the authors have sought to negate this by undertaking multiple sampling phases at each well and developing the presented model for temporal groundwater samples, as opposed to individual supplies. However, we do consider that the predictive capacity of the model, as corroborated through validation, may serve as evidence to suggest that a relatively small, yet varied, sampling regime may be sufficient for groundwater risk assessment at a local scale. Moreover, the sequential nature of the model allows for improved usability, whereby individuals can 'choose' which variables they wish to include on a hierarchical basis (which will affect accuracy) and are thus, less confined by data availability; offering a more simplistic approach which is not over-parameterised and thus potentially useful for non-experts.

5. Conclusion

The ISI-LR model developed in the current study represents a potentially useful and internationally transferable tool for predicting bacterial contamination of private domestic water wells in geologically heterogeneous regions with high groundwater contamination rates. The presented model was shown to be highly efficient, with the presence of *E. coli* being correctly classified in 94.2% of cases, and the absence of *E. coli* being correctly classified in 84.7% of cases, thus improving on previous similar studies. A further, external model validation step ($n = 32$) corroborates the efficacy of the model with a 90% predictability rate (20 predicted, 18 observed) for the presence of *E. coli* from a sampling cohort outside the study area.

When viewed through the lens of the characteristic agricultural, infrastructural, and hydrogeological profile of rural Ireland and other similar regions, findings serve to highlight the ubiquity and magnitude of private groundwater source hazards (both point and diffuse), in addition to the inherent complexities associated with effective source protection and maintenance. With no prospect of private well regulations imminent, there is an exigency for effective communication initiatives, particularly as private source reliance is expected to increase due to recent water charges, high urban rents, and housing demands in urban centres far outstripping supply. In conclusion, the developed model represents a comprehensive risk assessment and management tool that may be used by local authorities, water managers, and perhaps even well owners/users to develop effective water-quality management strategies to minimize pathogen exposure risks in Ireland and further afield.

Acknowledgements

This research was funded through the Irish Research Council Post-doctoral Fellowship Scheme (2017) (grant no. GOIPD/2017/721) and the Department of Chemical Sciences of the University of Limerick.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.envpol.2018.02.052>.

References

- Andrade, L., O'Dwyer, J., O'Neill, E., Hynds, P., 2018. Surface water flooding, groundwater contamination, and enteric disease in developed countries: a scoping review of connections and consequences. *Environ. Pollut.* 236, 540–549.
- APHA (American Public Health Association), 2005. Standard Methods for the Examination of Water and Wastewater, twenty-first ed. American Public Health Association, Washington, D.C., USA.
- Arnade, L.J., 1999. Seasonal correlation of well contamination and septic tank distance. *Groundwater* 37 (6), 920–923.
- Atherholt, T.B., Procopio, N.A., Goodrow, S.M., 2017. Seasonality of coliform bacteria detection rates in New Jersey domestic wells. *Groundwater* 55 (3), 346–361.
- Bacci, F., Chapman, D.V., 2011. Microbiological assessment of private drinking water supplies in Co. Cork, Ireland. *J. Water Health* 9, 738–751.
- Bahram, A., Hamid, B., Akram, N., 2012. Estimation of coliform contamination rate and impact of environmental factor on bacterial quality of tube well water supplies in Khorramdarreh County, Iran. *Afr. J. Biotechnol.* 11 (31), 7912–7915.
- Bradley, C., Byrne, C., Craig, M., Free, G., Gallagher, T., Kennedy, B., McDermott, G., 2015. Water Quality in Ireland 2010–2012. Environmental Protection Agency, Wexford, Ireland. Available at: <http://www.epa.ie/pubs/reports/water/waterqua/wqr20102012/>. (Accessed 16 January 2017).
- Bremer, J.E., Harter, T., 2012. Domestic wells have high probability of pumping septic tank leachate. *Hydrol. Earth Syst. Sci.* 16 (8), 2453–2467.
- Central Statistics Office, 2010. Statbank database – livestock and farm numbers listing. Accessed online. http://www.cso.ie/px/pxeirestat/Database/eirestat/Livestock%20and%20Farm%20Numbers/Livestock%20and%20Farm%20Numbers_statbank.asp?SP=Livestock%20and%20Farm%20Numbers&Planguage=0. (Accessed 16 December 2016).
- Central Statistics Office, 2012a. This is Ireland. Government of Ireland Publication, Molesworth street, Dublin. <http://www.cso.ie/en/census/census2011reports/census2011thisisirelandpart1/>. (Accessed 16 December 2016).
- Central Statistics Office, 2012b. The roof over our heads – housing in Ireland. Government of Ireland Publication, Molesworth street, Dublin. <http://www.cso.ie/en/census/census2011reports/census2011profile4theroofoverourheadshousinginireland/>. (Accessed 16 December 2016).
- Close, M., Dann, R., Ball, A., Pirie, R., Savill, M., Smith, Z., 2008. Microbial groundwater quality and its health implications for a border-strip irrigated dairy farm catchment, South Island, New Zealand. *J. Water Health* 6 (1), 83–98.
- Daly, D., 2000. Practical approaches to preventing pollution of wells. *Geol. Surv. Ireland* 14. Available online. <http://nfgws.ie/kfmgget-full/fckeditor/File/Protectinggroundwater.pdf>. (Accessed 21 November 2016).
- Daly, D., Warren, W.P., 1998. Mapping Groundwater Vulnerability: the Irish Perspective, vol. 130. Geological Society, London, pp. 179–190. Special Publications.
- Donaldson, L., 2002. Health services and the public health. *J. Epidemiol. Community*

- Health 56, 835–840.
- Drew, D., Burke, A., Daly, D., 1996. Assessing the extent and degree of karstification in Ireland. In: Proceedings of International Conference on Karst Fractured Aquifers—vulnerability and Sustainability. University of Silesia, Katowice-Ustron, Poland, pp. 37–47.
- Duffy, P., 2000. Trends in nineteenth and twentieth century settlement. In: Barry, T. (Ed.), A History of Settlement in Ireland. Routledge, London, pp. 206–227.
- Embrey, S.S., Runkle, D.L., 2006. Microbial quality of the Nation's ground-water Resources, 1993–2004. <https://pubs.usgs.gov/sir/2006/5290/>. (Accessed 8 September 2016).
- EPA, 2014. National Inspection Plan - Domestic Wastewater Treatment Systems 2013–An Interim Review Environmental Protection Agency, Ireland, p. 4.
- Fealy, R., Green, S., 2009. Teagasc-EPA Soils and Subsoils Mapping Project: Final Report V. 1. Teagasc; Environmental Protection Agency.
- Gonzales, T.R., 2008. The effects that well depth and wellhead protection have on bacterial contamination of private water wells in the Estes Park Valley, Colorado. *J. Environ. Health* 71 (5), 17.
- Goss, M.J., Barry, D.A.J., Rudolph, D.L., 1998. Contamination in Ontario farmstead domestic wells and its association with agriculture. Results from drinking water wells. *J. Contam. Hydrol.* 32 (3), 267–293.
- GSI, 2000. Groundwater sand and gravel aquifers, dataset produced by geological survey Ireland. <http://spatial.dcenr.gov.ie/GeologicalSurvey/Groundwater/index.html>. (Accessed 8 October 2016).
- Guzman-Herrador, B., Carlander, A., Ethelberg, S., De Blasio, B.F., Kuusi, M., Lund, V., Löfdahl, M., Macdonald, E., Nichols, G., Schönning, C., 2015. Waterborne outbreaks in the Nordic countries, 1998 to 2012. *Euro Surveill.* 20, 24.
- Health Protection Surveillance Centre (HSPC), 2015. Annual Epidemiological Report 2013. Dublin, Ireland.
- Hosmer, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression. Wiley.
- Howard, G., Pedley, S., Barrett, M., Nalubega, M., Johal, K., 2003. Risk factors contributing to microbiological contamination of shallow groundwater in Kampala, Uganda. *Water Res.* 37 (14), 3421–3429.
- Hynds, P.D., Misstear, B.D., Gill, L.W., 2012. Development of a microbial contamination susceptibility model for private domestic groundwater sources. *Water Resour. Res.* 48 (12).
- Hynds, P.D., Misstear, B.D., Gill, L.W., 2013. Unregulated private wells in the Republic of Ireland: consumer awareness, source susceptibility and protective actions. *J. Environ. Manag.* 127, 278–288.
- Hynds, P.D., Gill, L.W., Misstear, B.D., 2014a. A quantitative risk assessment of verotoxigenic *E. coli* (VTEC) in private groundwater sources in the Republic of Ireland. *Hum. Ecol. Risk Assess. Int. J.* 20, 1446–1468.
- Hynds, P.D., Thomas, M.K., Pintar, K.D.M., 2014b. Contamination of groundwater systems in the US and Canada by enteric pathogens, 1990–2013: a review and pooled-analysis. *PLoS One* 9, 93301.
- Job, C.A., 2010. Groundwater Economics. CRC Press, Boca Roton, FL.
- Kabore, H., Levallois, P., Michel, P., Payment, P., Dery, P., Gingras, S., 2010. Association between potential zoonotic enteric infections in children and environmental risk factors in Quebec, 1999–2006. *Zoonoses Public Health* 57 (7–8), 195–205.
- Kozuskanich, J.C., Novakowski, K.S., Anderson, B.C., Crowe, A.S., Balakrishnan, V.K., 2014. Anthropogenic impacts on a bedrock aquifer at the village scale. *Groundwater* 52 (3), 474–486.
- Lenaker, P.L., Corsi, S.R., Borchardt, M.A., Spencer, S.K., Baldwin, A.K., Lutz, M.A., 2017. Hydrologic, land cover, and seasonal patterns of waterborne pathogens in Great Lakes tributaries. *Water Res.* 113, 11–21.
- Met Eireann, 2016. Climate of Ireland. Accessed online. <http://www.met.ie/climate/climate-of-ireland.asp>. (Accessed 7 August 2016).
- Murphy, H., Thomas, M., Schmidt, P., Medeiros, D., Mcfadyn, S., Pintar, K., 2016. Estimating the burden of acute gastrointestinal illness due to Giardia, Cryptosporidium, Campylobacter, E. coli O157 and norovirus associated with private wells and small water systems in Canada. *Epidemiol. Infect.* 144, 1355–1370.
- Murphy, H.M., Prioleau, M., Borchardt, M., Hynds, P.D., 2017. Groundwater and enteric disease: a global review of the epidemiological evidence 1948–2015. *Hydrogeol. J.* 25 (4), 981–1001. Special Issue “Hydrogeology and Human Health”.
- O'Dwyer, J., Downes, M.M., Adley, C.C., 2016. The impact of meteorology on the occurrence of waterborne outbreaks of vero cytotoxin-producing *Escherichia coli* (VTEC): a logistic regression approach. *J. Water Health* 14 (1), 39–46.
- O'Dwyer, J., Dowling, A., Adley, C.C., 2014. Microbiological assessment of private groundwater-derived potable water supplies in the Mid-West Region of Ireland. *J. Water Health* 12 (2), 310–317.
- O'Dwyer, J., Hynds, P., Pot, M., Adley, C.C., Ryan, M.P., 2017. Evaluation of levels of antibiotic resistance in groundwater-derived *E. coli* isolates in the Midwest of Ireland and elucidation of potential predictors of resistance. *Hydrogeol. J.* 1–13. Issue “Hydrogeology and Human Health”.
- O'haiseadha, C., Hynds, P.D., Fallon, U.B., O'Dwyer, J., 2017. A geostatistical investigation of agricultural and infrastructural risk factors associated with primary verotoxigenic *E. coli* (VTEC) infection in the Republic of Ireland, 2008–2013. *Epidemiol. Infect.* 145 (1), 95–105.
- Pitkänen, T., Juselius, T., Isomäki, E., Miettinen, I.T., Valve, M., Kivimäki, A.-L., Lahti, K., Hänninen, M.-L., 2015. Drinking water quality and occurrence of Giardia in Finnish small groundwater supplies. *Resources* 4, 637–654.
- Pronk, M., Goldscheider, N., Zopfi, J., 2009. Microbial communities in karst groundwater and their potential use for biomonitoring. *Hydrogeol. J.* 17 (1), 37–48.
- Richards, R.P., Baker, D.B., Creamer, N.L., Kramer, J.W., Ewing, D.E., Merryfield, B.J., Wallrabenstein, L.K., 1996. Well water quality, well vulnerability, and agricultural contamination in the Midwestern United States. *J. Environ. Qual.* 25 (3), 389–402.
- Schets, F.M., During, M., Italiaander, R., Heijnen, L., Rutjes, S.A., Van der Zwaluw, W.K., de Roda Husman, A.M., 2005. *Escherichia coli* O157: H7 in drinking water from private water supplies in The Netherlands. *Water Res.* 39 (18), 4485–4493.
- Scott, M., Murray, M., 2009. Housing rural communities: connecting rural dwellings to rural development in Ireland. *Hous. Stud.* 24, 755–774.
- Tabbot, P.N., Robson, M.G., 2006. The New Jersey residential well-testing program—a case study: randolph township. *J. Environ. Health* 69 (2), 15.
- Wallender, E.K., Ailes, E.C., Yoder, J.S., Roberts, V.A., Brunkard, J.M., 2014. Contributing factors to disease outbreaks associated with untreated groundwater. *Groundwater* 52, 886–897.
- Williams, N.H., Misstear, B.D.R., Daly, D., Lee, M., 2013. Development of a national groundwater recharge map for the Republic of Ireland. *Q. J. Eng. Geol. Hydrogeol.* 46 (4), 493–506.
- Yates, M.V., 1985. Septic tank density and ground-water contamination. *Groundwater* 23 (5), 586–591.
- Zwahlen, F., 2004. Vulnerability and Risk Mapping for the Protection of Carbonate (Karst) Aquifers, Final Report (COST Action 620). European Commission, Directorate XII Science, Research and Development, Report EUR 20912, p. 297. Brussels.