Review Article

Can E. coli or thermotolerant coliform concentrations predict pathogen presence or prevalence in irrigation waters?

Yakov Pachepsky1, Daniel Shelton1, Sarah Dorner2, and Gene Whelan3

1USDA-ARS, Environmental Microbial and Food Safety Laboratory, Beltsville, MD, USA, 2Department of Civil, Geological and Mining Engineering, École Polytechnique de Montréal, Montreal, Quebec, Canada, and 3US Environmental Protection Agency, National Exposure Research Laboratory, Athens, GA, USA

Abstract

An increase in food-borne illnesses in the United States has been associated with fresh produce consumption. Irrigation water presents recognized risks for microbial contamination of produce. Water quality criteria rely on indicator bacteria. The objective of this review was to collate and summarize experimental data on the relationships between pathogens and thermotolerant coliform (THT) and/or generic E. coli, specifically focusing on surface fresh waters used in or potentially suitable for irrigation agriculture. We analyzed peer-reviewed publications in which concentrations of E. coli or THT coliforms in surface fresh waters were measured along with concentrations of one or more of waterborne and food-borne pathogenic organisms. The proposed relationships were significant in 35% of all instances and not significant in 65% of instances. Coliform indicators alone cannot provide conclusive, non-site-specific and non-pathogen-specific information about the presence and/or concentrations of most important pathogens in surface waters suitable for irrigation. Standards of microbial water quality for irrigation can rely not only on concentrations of indicators and/or pathogens, but must include references to crop management. Critical information on microbial composition of actual irrigation waters to support criteria of microbiological quality of irrigation waters appears to be lacking and needs to be collected.

Introduction

An increase in food-borne illnesses in the United States has been associated with fresh produce consumption in recent years. Produce-associated outbreaks accounted for 0.7% of all reported US foodborne outbreaks in the 1970s, 6% in 1990s and 46% in the period 1998 through 2008 (Painter et al., 2013; Sivapalasingam et al., 2004). The increased percentage partly reflects improvements in cattle, swine and poultry processing, however, it also indicates an increase in the number and severity of produce-associated outbreaks. Annual costs of food-borne illness in the US have been estimated to range from 4.4 to 39 billion USD (Hoffmann et al., 2012; Scharff, 2010).

Large-scale production of produce typically requires some form of irrigation during the growing season and there is a growing body of research elucidating potential pathways of produce contamination by waterborne pathogens. Both irrigation waters and waters used in agricultural processes can contain pathogenic organisms harmful to human health (Fan et al., 2009; Pachepsky et al., 2011). Pathogens of interest include bacteria Campylobacter spp., enterohemorrhagic Escherichia coli (e.g. E. coli O157:H7), enterotoxigenic Staphylococcus aureus, enterotoxigenic Bacillus cereus, Listeria monocytogenes, Salmonella spp., Shigella spp., Yersinia enterocolitica; protozooa Cryptosporidium spp., Cyclospora cayetanensis, Giardia spp., Entamoeba histolytica; helminths such as Ascaris spp.; and viruses, particularly adenoviruses, enteroviruses, noroviruses and rotaviruses.

With increasing recognition of the importance of microbiological quality of irrigation water and its impact on food safety and public health, the need to regulate it is obvious. Microbiological water quality criteria become standards when adopted into state law (Steele et al., 2005) and are most often based on concentrations of indicator microorganisms which, although not pathogenic, are presumed to signify the presence of pathogens causing illnesses. A large number of microorganisms have been proposed and tested as indicators (Ashbolt et al., 2001), but only a few have been adopted in standards. Water quality criteria recommendations have been published by the USEPA, as discussed in Section 304(a) of the Clean Water Act (USEPA, 2013), and have established recommended levels of a pollutant in ambient waters that protect a water body’s designated use. Recommendations for recreational and drinking waters in the United States were first
adopted in 1976 (USEPA, 1976) and the earliest criteria used “total coliforms” as the indicator organism (USEPA, 1973). Because fecal contamination was considered the most probable source of pathogens in waters, however, micro-organisms occurring in feces were eventually selected as more appropriate. Subsequently, criteria were based on the thermotolerant (fecal) coliforms (USEPA, 1986). Water quality criteria published by EPA in 1986 and 2012 are recommended levels of fecal indicator bacteria in ambient waters to protect a recreational use. Generic E. coli is used as a fecal indicator bacterium for fresh water recreation, although enterococci can also be used (USEPA, 2012). Identical microbial quality criteria have been proposed for irrigation waters used on produce (CSFSGLLGSC, 2009; USFDA, 2013).

Microbiological standards are scrutinized and criticized periodically. A major criticism is the absence of documented relationships between indicator organism and pathogen concentrations. Current EPA criteria for recreational water are based on epidemiological data –i.e. on comparing rates of self-reported gastrointestinal illness for swimmers versus non-swimmers and correlating “excess” illnesses with waterborne E. coli concentrations. These criteria do not rely on correlations between E. coli and any given pathogen.

The most recent literature review of the accuracy of indicator organisms in predicting pathogen levels was undertaken by Wu et al. (2011). They summarized 540 datasets published from 1970 to 2009 that correlated a range of indicator organisms (14) with waterborne pathogens (18) found in surface or ground water. Their findings show that for any pairing of indicator-pathogen, some studies provide evidence of a correlation while others do not. They noted that the larger the dataset, the greater the chance of finding a correlation, however, even weak correlations may be significant if there are sufficient data points. Interestingly, they found that generic E. coli appeared to be one of the least reliable indicators of pathogen contamination. Despite the importance of irrigation waters, a survey of their microbial contamination levels has not yet been compiled for the US (Stoeckel, 2009) or any other country. Furthermore, we are not aware of any regular monitoring/reporting on the microbial quality of irrigation waters done anywhere in the world. This is partly due to the cost of extensive sampling. In addition, producers/growers who have begun to collect data on microbial water quality may be reluctant to share them (Suslow, 2010).

The objective of our review was to collate and summarize experimental data on the relationships between pathogens and thermotolerant coliform (THT) and/or generic E. coli, specifically focusing on surface fresh waters used in or potentially suitable for irrigation agriculture. The goal was to draw conclusions regarding the type of information about microbial quality of irrigation water that can be obtained from THT coliforms and/or E. coli concentration monitoring.

Materials and methods

The dataset

We collected peer-reviewed publications in which concentrations of E. coli or THT coliforms in natural surface fresh waters were measured with concentrations of one or more of the following waterborne and food-borne pathogenic organisms: Salmonella spp., Campylobacter spp., shiga-toxigenic E. coli, Listeria monocytogenes, Cryptosporidium parvum and Giardia spp. A total of 16 629 data points was compiled from 35 papers and 81 individual datasets on coupled, one-time observations of “THT or E. coli” versus pathogen. The summary of the database is given in the Supplemental Table S1. For each paper, the Table contains the reference, report location, total number of samples, number of samples in which an indicator and pathogen was found, type of water, method used to research the relationship between indicators and pathogens and conclusion about significance of the relationship. Supplemental Table S1 also includes the type of relationship between indicators and pathogens. In the majority of cases, both indicator and pathogen organisms were characterized with two values: prevalence (i.e. percentage of cases when the organism was detected) and concentration, meaning that both values could potentially be compared. Three types of relationships were examined: I – Indicator prevalence as predictor of pathogen prevalence; II – Indicator concentration as predictor of pathogen prevalence; and III – Indicator concentration as predictor of pathogen concentration.

Relationship Types I and III were most often researched by multiple logistical regression. Pathogen absence was considered a dependent variable in the logistic multivariable regression model. A special input variable was introduced to denote the presence of the indicator in the sample; it was set to zero if the indicator was detected and to one if not. The predictive worth of all input variables, including the one coding presence/absence, was analyzed by computing regression coefficient estimates and their p values (Hörmann et al., 2004). In case of multiple indicators, Type I relationships were also researched using discriminant analysis. Results of assays for indicator organisms were converted to a string of binary variables representing presence or absence. The ability of the indicator data string to predict the presence or absence of a pathogen is expressed as the percentage of samples correctly classified into “pathogen present” and “pathogen absent” categories (Harwood et al., 2005). An additional measure of accuracy for the Type I relationship was the percentage of cases when the indicator was absent and the pathogen present (Duris et al., 2013).

Type II relationships were studied with biserial correlations as a generic statistic to establish the relationship between continuous and categorical (presence-absence) variables (Gu et al., 2013). Such relationships were also researched using the Mann–Whitney test and classification and regression trees. The Mann–Whitney U significance test (a non-parametric analog of the two-sample t-test) determines if samples of indicator bacteria, grouped by pathogen presence or absence, came from the same distribution at the set significance level (Arvanitidou et al., 1995). The classification and regression tree were applied to see if there was a threshold indicator concentration that would split samples into two groups with distinctly different prevalence of a pathogen (Wilkes et al., 2009). Type II relationships were also sought using discriminant analysis in the case of multiple indicators (Costán-Longares, 2008). A linear combination of
indicator concentrations that would give the best separation of samples into groups with and without a pathogen was also sought. The water quality criteria value for *E. coli* was tested as a threshold to estimate pathogen prevalence in samples above and below it (Duris et al., 2013).

Type III relationships were estimated using the Pearson correlation coefficient or, more frequently, Spearman rank correlation coefficient. The latter is deemed more appropriate since it does not impose requirements on the type of distribution (Brookes et al., 2005). The partial least squares correlation coefficient. The latter is deemed more appropriate correlation coefficient or, more frequently, Spearman rank above and below it (Duris et al., 2013).

Results and discussion

Occurrence of significant relationships

We found a total of 81 datasets for which attempts to establish relationships between *E. coli* and pathogens (40 datasets) and THT coliforms and pathogens (41 datasets) were reported. The proposed relationships were significant in 28 instances (35%) and not significant in 53 instances (65%). In seven instances a significant relationship was found between *E. coli* and pathogen(s), while in 21 instances a significant relationship was found between THT and pathogen(s).

In five studies (12 datasets), the authors conducted a Type I analysis comparing the presence/absence of indicator versus pathogen in water samples. The only study reporting a significant relationship was conducted by St-Pierre et al. (2009) in which 2471 river water samples were analyzed for presence/absence of THT and *Campylobacter*: 59% of samples were positive for THT, while 43% were positive for *Campylobacter*, giving an *r* value of 0.22. Payment et al. (1982) compared prevalence of THT and *Salmonella* in reservoir water, while Wilkes et al. (2009) compared prevalence of *E. coli* with *Cryptosporidium, Giardia, Salmonella, Campylobacter, Listeria* and *E. coli* O157:H7 in river water. Since the prevalence of THT and *E. coli* were 100%, it was not possible to establish a correlation. Payment et al. (1982) reported a prevalence of 31% for *Salmonella*, while Wilkes et al. (2009) reported prevalences of 44% (*Cryptosporidium*), 25% (*Giardia*), 22% (*Campylobacter*), 19% (*Listeria*), 9% (*Salmonella*) and 0.6% (*E. coli* O157:H7).

Ahmed et al. (2010) found no relationship between the presence/absence of *E. coli* and *Giardia, Salmonella* or *Campylobacter*. Finally, Harwood et al. (2005) failed to observe a relationship between *E. coli* and *Cryptosporidium* which is unusual in that the prevalence of *Cryptosporidium* (70%) was higher than *E. coli* (27%), and probably reflects the greater relative removal of *E. coli* during treatment.

In 10 studies (29 datasets), authors conducted a Type II analysis comparing the indicator concentration to pathogen prevalence; significant relationships were reported in 14 cases. The most common approach was to distinguish high versus low pathogen prevalence, then identify the indicator concentration (THT or *E. coli*) that best separated the two groups according to a statistical criterion. Usually, significant relationships showed that the *E. coli* and THT median concentration, or statistical distributions of those concentrations, were statistically different between “pathogen present” and “pathogen absent” groups. For example, Wilkes et al. (2009) conducted ten comparisons between THT or *E. coli* and *Giardia, Cryptosporidium, Salmonella, Campylobacter, Listeria* and *E. coli* O157:H7. They were able to identify the indicator concentration that best distinguished the two groups with distinctly different prevalences of pathogen, however, that concentration was unique to each dataset and no single indicator concentration could predict the group to which a particular site belonged.

In 29 studies (40 datasets), data were analyzed using a Type III approach, based on actual concentrations of both indicator and pathogen. They were significant in 13 cases and not significant in 27 cases. Using THT as the indicator (seven for *Cryptosporidium*, six for *Giardia*, two for *Campylobacter* and six for *Salmonella, one for Yersinia*), 23 datasets reported correlation coefficients ranging from −0.05 to 0.91, but no pattern was discernible. Using *E. coli* as indicator (nine for *Cryptosporidium*, two for *Giardia*, three for *Campylobacter*, one for pathogenic *E. coli* and two for *Salmonella*), some of the 17 datasets reported correlation coefficient values that varied widely and no pattern was discernible. For example, McEgan et al. (2013) compared concentrations of *E. coli* to *Salmonella* in samples from 18 different canals, ponds, streams, etc. in central Florida; correlation coefficients ranged from 0.000 to 0.678.

Percentages of significant relationships between indicators and pathogens in potential irrigation water sources (rivers, lakes and ponds) were not substantially different among different pathogens except *Salmonella* (Table 1). For *Campylobacter* spp., *Cryptosporidium parvum, Giardia* spp., *Listeria monocytogenes*, Shiga-toxin-producing *E. coli* and *E. coli* O157:H7, *E. coli* O157 and *Salmonella*, the percentages of significant “indicator-pathogen” relationships were in the range between 19 and 33%. Although *Salmonella* gave a relatively high number of significant relationships (65%), they were generally weak with reported correlation coefficient values ranging from 0.28 to 0.66 (Supplemental Table S1).

Where relationships between coliform indicators and pathogens were established, threshold and range of *E. coli* and THT concentration values appeared to be region- and pathogen-specific (Figure 1). The median value of *Salmonella* detection in northern Greece was substantially higher than in the province of Ontario, Canada (Figure 1a). While median *E. coli* concentrations were different in cases of detection and non-detection of *Salmonella*, no difference was observed for *Campylobacter*. Similar independence of *Campylobacter*

| Pathogen            | Total number of case studies | Total number of significant relationships |
|---------------------|-----------------------------|------------------------------------------|
| *Campylobacter* spp.| 17                          | 5                                        |
| *Cryptosporidium parvum* | 21                      | 4                                        |
| *Giardia* spp.      | 12                          | 4                                        |
| *Listeria monocytogenes* | 5                        | 1                                        |
| *E. coli* O157      | 3                           | 1                                        |
| *Salmonella* spp.   | 20                          | 14                                       |
| STEC                | 2                           | 1                                        |
| *Yersinia* spp.     | 1                           | 0                                        |
detection on \textit{E. coli} concentrations was observed in summer, but not in winter in the Ontario province. Two studies in Spain (Figure 1b) presented quite different data on \textit{Salmonella} prevalence dependence on thermotolerant coliform concentrations. The site in northeastern Spain apparently had much higher prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). \textit{E. coli} concentration of 300 CFU (100 mL)$^{-1}$ provided a good threshold value between low and moderate concentrations of pathogenic genes in Pennsylvania. Much lower thresholds were found between high and low prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). \textit{E. coli} concentration of 300 CFU (100 mL)$^{-1}$ provided a good threshold value between low and moderate concentrations of pathogenic genes in Pennsylvania. Much lower thresholds were found between high and low prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). \textit{E. coli} concentration of 300 CFU (100 mL)$^{-1}$ provided a good threshold value between low and moderate concentrations of pathogenic genes in Pennsylvania. Much lower thresholds were found between high and low prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). \textit{E. coli} concentration of 300 CFU (100 mL)$^{-1}$ provided a good threshold value between low and moderate concentrations of pathogenic genes in Pennsylvania. Much lower thresholds were found between high and low prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). \textit{E. coli} concentration of 300 CFU (100 mL)$^{-1}$ provided a good threshold value between low and moderate concentrations of pathogenic genes in Pennsylvania. Much lower thresholds were found between high and low prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). \textit{E. coli} concentration of 300 CFU (100 mL)$^{-1}$ provided a good threshold value between low and moderate concentrations of pathogenic genes in Pennsylvania. Much lower thresholds were found between high and low prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). \textit{E. coli} concentration of 300 CFU (100 mL)$^{-1}$ provided a good threshold value between low and moderate concentrations of pathogenic genes in Pennsylvania. Much lower thresholds were found between high and low prevalence of \textit{Salmonella} than the site in southern Spain with the same range of THT concentrations. Different threshold concentrations were found for prevalence of different pathogens (Figure 1c). 

Factors affecting “pathogen–indicator” relationships

Ecology and sources

The relatively low number of established significant “indicator–pathogen” relationships can be explained by different sources, release patterns and environmental fate and transport of the microbes in different indicator and pathogen groups. For example, \textit{Listeria monocytogenes} is known to survive in multiple habitats, and differences in sources might lead to the negative correlation between \textit{E. coli} and \textit{L. monocytogenes} concentrations found by Wilkes et al. (2009) in large datasets from the Ontario province. Duris et al. (2013) noted that \textit{Giardia} occurrence was likely related to nonpoint sources which are highly influential during seasonal overland transport from snowmelt and elevated precipitation in late winter and spring in Pennsylvania. It is unknown whether pathogens and \textit{E. coli} are released from land-deposited animal waste at the same rate, but differences in \textit{E. coli} and enterococci release were documented (Guber et al., 2007). Rates of survival in soil, animal waste, and waters differ among \textit{E. coli} and pathogens (e.g. Jenkins et al., 2011 for \textit{E. coli O157 H7}; Korhonen & Matikainen, 1991, for \textit{Campylobacter}). Similarity in survival rates of \textit{E. coli} and \textit{Salmonella}, that was observed in some freshwater environments (Burton et al., 1987), may be the reason for the relatively large number of cases when the significant relationship between \textit{Salmonella} and \textit{E. coli} concentrations could be established (Table 1).
Differences in transport of *E. coli*, THT, and pathogens of interest differences may affect the relationship between indicators and pathogens. Lemarchand & Lebaron (2003) and Duris et al. (2013), for example, hypothesized that although sources of fecal indicator bacteria and protozoa can be the same, weak correlation to fecal indicator bacteria may be due to differences in settling velocities of protozoa and coliforms and potential differences in entrainment and transport of coliforms and protozoa in soil and in water. Brookes et al. (2005) noted correlations of *Cryptosporidium* spp. and microbial indicators with different particle-size classes and suggested that *Cryptosporidium* spp. tends to be transported similarly to small particles, while bacterial indicators tend to be transported like relatively large particles.

The ecology of pathogens and coliform indicators in surface waters also can compromise the efficiency of *E. coli* and THT as microbial water quality indicators. Many members of the thermotolerant coliform group such as species of *Klebsiella* and *Enterobacter* are not specific to feces (Leclerc et al., 2001). *E. coli* can survive in soils; it was found to grow in soils in both tropical and temperate climate conditions (Ishii & Sadowski, 2008; Nautiyal et al., 2010). *E. coli* has been shown to grow in freshwater aquatic environments, and both direct and indirect evidence suggested that channel bed sediment “stores” closely reflected the land use within their catchments and that there was little die-off of organisms along watercourses (Crowther et al., 2002; Pachepsky & Shelton, 2011). Some higher aquatic organisms and soil organisms can harbor both pathogen and indicator microorganisms (Barker et al., 1999; Bichai et al., 2008), and physiological differences of indicators and pathogens may affect this survival mode. The species *E. coli* per se is incredibly diverse (Ishii & Sadowski, 2008) which can affect their fate and transport: for instance, different *E. coli* strains were shown to have different attachment affinity to particles of different sizes (Pachepsky et al., 2008) and to survive differently in soils (Topp et al., 2003).

**Sampling volume and frequency**

A mismatch in sample volume has been cited as a possible reason for the absence of relationships between indicator bacteria and protozoa (Hörman et al., 2004). *E. coli* and THT may well be present in large volumes of water needed to enumerate oocysts or cysts, but the concentrations of oocysts and cysts in small sample volumes used for bacteria enumeration can be below detection limits. On the other hand, a possible factor affecting low correlation could be different microbial densities in original contamination sources, meaning that failure to detect pathogens was due to sampling volumes that were too small (Geldreich, 1996).

The total number of samples can affect statistical significance of the relationships between coliform indicators and pathogens. Savill et al. (2001) noted that “a low Spearman’s *r* value does not allow ‘no correlation’ to be inferred but merely denotes that there may be insufficient evidence to conclude that a relationship exists”. The exhaustive review of Wu et al. (2011) demonstrated that as the total number of samples increases, there is an increased chance of observing a relationship between pathogens and indicator.

We tested three statistical hypotheses that compared numbers of samples in Supplemental Table S1 cases, where relationships between coliform indicators were found, with numbers of samples in Supplemental Table S1 cases where relationships were not found. Results are presented in Table 2. It shows low probabilities associated with the absence of the effect of the number of samples in an experimental study on significance of relating coliform concentrations and pathogens. Several reasons for such low probabilities are possible. A purely formal cause may be that statistical significance depends on the number of samples and critical values become smaller as the number of degrees of freedom increase.

Another reason may be that increasing the number of samples generally increases the chance of sampling extremes associated with concentrations, which can increase correlation coefficients. This effect was observed by LeChevallier et al. (1991) who found a relatively strong relationship between *Cryptosporidium* oocyst and THT concentrations, whereas Rose et al. (1988) found none. LeChevallier et al. (1991) suggested that this discrepancy arose from differences in the type of water samples analyzed. Rose et al. (1988) analyzed relatively pristine waters, while LeChevallier et al. (1991) examined a variety of source waters. The increase in sample numbers can also increase probability of having high flow samples known to have higher concentrations of coliform indicators and pathogens (Crowther et al., 2002). The value of individual “indicator–pathogen” data pairs in microbial quality of irrigation water assessment, on the other

| Relationship between coliform indicators and pathogens | Number of case studies | Statistics of numbers of samples |
|------------------------------------------------------|------------------------|----------------------------------|
|                                                      |                        | Average | Median | Geometric mean |
| No significant relationships                         | 51                     | 160     | 86     | 93             |
| Significant relationships                            | 28                     | 302     | 145    | 145            |
|                                                      | Averages are the same*  | 0.15    | 0.07   | 0.12           |
|                                                      | Medians are the sameb   |         |        |                |
|                                                      | Geometric means are the samec |       |        |                |
|                                                      | Distributions are the samed |       |        |                |

* *t*-test.  
* Mann–Whitney test.  
* *t*-test.  
* Kolmogorov–Smirnov test.
hand, must be evaluated since high flow events are relatively rare and may signify an opportunity to stop irrigation.

Wu et al. (2011) suggested that the significance of relationships between indicators and pathogens can be affected by prevalence of pathogens. We computed the average prevalence of pathogens in cases with significant and non-significant relationships between coliform indicators and pathogens; average prevalence values of both were 42% and not statistically different (Supplemental Table S1).

Seasonality and spatial scale

Seasonality appears to be an important factor in the strength of relationships between coliform indicators and pathogens. Wilkes et al. (2009) researched seasonal relationships among indicator bacteria, pathogenic bacteria, Cryptosporidium oocysts, Giardia cysts and hydrological indices for surface waters within agricultural landscapes in Eastern Ontario. They noted that, overall, relationships between indicator bacteria, pathogens and parasite oocysts/cysts were weak, seasonally-dependent and site-specific. At various sampling sites, significant Salmonella versus coliform indicator relationships were all weakly positive, with fall-winter datasets providing somewhat stronger relationships overall. Seasons affected both prevalence and concentrations of coliform indicators and pathogens. In a survey in Finland, the pathogens were detected less frequently during winter than in spring, summer or autumn; this pattern was observed especially with occurrence of Campylobacter spp. (Hörman et al., 2004). A large microbial water quality survey in Pennsylvania revealed seasonal differences in density and frequency of coliform indicators and Giardia, and in detection frequency of the eaeA pathogen gene (Duris et al. 2013). Seasonal variation in shedding rates of Campylobacter and Salmonella were listed as an important contribution to seasonal detection of these organisms. Differences in runoff and stream flows can affect the mobilization of coliform indicators from sediments and soils, thus changing indicator concentrations without relating to pathogen shedding (Pachepsky & Shelton, 2011). Culturability of pathogens is affected by water temperature (Rollins & Corwell, 1986); consequently, higher rates of recovery in summer months may create a seasonality effect on relationships with pathogen concentrations. Yet another cause for the seasonality in these relationships may be differences in prevalence and/or concentrations in feces among shedding animals, as well their migratory patterns (Carter et al., 1987).

Spatial scale can be underestimated as a factor affecting concentrations of coliform indicators and, possibly, prevalence and concentrations of pathogens. Differences may cause scale-dependencies in “indicator–pathogen” relationships: E. coli concentrations in studies in Texas consistently decreased as watershed scale increased from field to small watershed to river basin (Harmel et al., 2010). A nationwide study in Canada (Edge et al., 2012) reported that the annual mean E. coli concentration was significantly higher at agricultural sites on small streams [649 ± 160 CFU (100 mL)\(^{-1}\)] than on large streams/rivers [318 ± 93 CFU (100 mL)\(^{-1}\)]. The same authors found that annual mean frequency of occurrence of Campylobacter spp., Salmonella spp. and Cryptosporidium spp. was higher at agricultural sites on small streams (Strahler stream order ≤3) than on larger streams/rivers (Strahler stream order >3). Beyond lower dilution, such differences could be attributed, at least partly, to the seasonal hydrology of small streams which can experience low and even stagnant waters, affecting microbial ecology in the water courses (Wilkes et al., 2009). He et al. (2007) demonstrated differences in E. coli concentrations in stagnant and flowing parts of low-order streams that can affect estimated relationships with other microorganisms. Interestingly, they found that generic E. coli appeared to be one of the least reliable indicators of pathogen contamination.

Uses of coliform indicators

In general, coliform indicators alone cannot provide conclusive, non-site-specific and non-pathogen-specific information about the presence and/or concentrations of most important pathogens in surface waters suitable for irrigation. Nonetheless, E. coli and THT concentrations are currently used by regulatory agencies to assess the presence of fecal contamination and other pathogens in freshwaters. These organisms are monitored primarily as an index of suspicion: if fecal contamination is present, pathogens may also be present. Thermotolerant coliforms and E. coli are used because (i) there are too many potential pathogens to monitor and (ii) indicators are shed by the entire populations of warm blooded animals in higher numbers than any of the pathogens and should still be present after the pathogens have died off. False-positives (indicators present in the absence of detectable pathogens) are expected to occur fairly often. Although they cause false alarms, these false positives are tolerated. Public health officials are willing to accept reasonable numbers of false-positives because they err on the side of public health. These false positives alone will cause correlation coefficients to be low. False-negatives (pathogens present in the absence of indicators) are considered to be a greater problem. The papers written on correlations between indicators and pathogens normally do not take the expected false-positives into account and do not tell us how much of the lack of correlation is due to false-positives and how much is due to false-negatives.

There is a wealth of data on E. coli and THT concentrations. How, then, can these concentration data be used to inform on microbiological quality of irrigation waters? Four possibilities have been explored in literature so far.

(1) Pathogen prevalence and concentrations were estimated using coliform indicators not as sole predictors in empirical equations but as predictors along with other indicator organisms and/or with hydrological, land use and other parameters.

(2) Pathogen prevalence and occurrence were estimated using fate-and-transport process modeling. In these models, coliform concentration data were used to estimate transport parameters whereas the release and fate parameters were specific for pathogens.

(3) Coliform concentrations were used to define microbial quality of irrigation waters in terms other than a specific pathogen concentration and occurrence.
Concentrations of E. coli in irrigation waters were related directly to health hazards from consumption of irrigated crops. Development of pathogen prevalence or concentration predictions with regressions that have coliform concentration or prevalence as one of several independent variables has been demonstrated in various regional studies. Concentration of E. coli and absence of C. perfringens were found to have significant predictive value for the absence of pathogens in a logistic regression model in a study in southern Finland (Hörmann et al., 2004). LeChevallier et al. (1991), for example, presented multiple linear regressions to estimate Cryptosporidium oocyst and Giardia cyst concentrations in which E. coli concentration served as an input variable along with turbidity, temperature and degree of watershed protection. Vereen et al. (2007) presented a multivariate model to estimate Campylobacter concentrations from temperature, total monthly rainfall before sample collection and fecal coliform counts with fixed effects of season and site. In the latter work, however, when other predictive variables such as sampling site, rainfall, and summer season were included in the multivariate model, thermotolerant coliforms counts ceased to appear in the list of predictors. The authors suggested that thermotolerant coliforms might be influenced by the same environmental variables as Campylobacters.

Regressions with E. coli among predictors of pathogen occurrence and/or concentrations appear to be valuable in evaluating relative importance of factors and controls. Since these regression equations are purely empirical, their applicability outside the development region is unknown. Also, data used to develop these equations reflect climatic conditions and climate change may affect their applicability (Nnane et al., 2011).

Using process-based fate and transport models to estimate pathogen occurrences and concentrations was suggested as a way to overcome the complexity of accounting for multiple controls of pathogen dissemination. Wilkes et al. (2009) noted that the strength and direction of hydrology–microorganism relationships appear to depend on seasonal characteristics, type of microorganism, sample site disposition (e.g. stream order), upstream land use and differences in specific hydrological loading/transport processes. The authors emphasized that these complex relationships underscore the potential benefits that process-based approaches might have over qualitative techniques in predicting fecal pollution and pathogen risk in agricultural watersheds (i.e. reducing the uncertainty of indicator density versus pathogen relationships, with respect to dilution effects). Haydon & Deletic (2006) and Dorner et al. (2006) reported good performance of an E. coli fate-and-transport model for catchments with areas from 10 to 100 km². Whelan et al. (2014) demonstrated the application of an integrated modeling system to generate various exposure scenarios in a quantitative microbial risk assessment framework. Fate and transport were simulated for Salmonella enterica, Cryptosporidium spp., and E. coli O157:H7 released from typical rural sources. McBride et al. (2011) developed and applied a set of models to evaluate the presence of Campylobacter in food and the environment, and its link to public health. In this work, four coupled models simulated source attribution using genotype information; pathway attribution, including various exposures; carriage and transmission by farmed animals; and catchment dynamics and associated risk models. Brookes et al. (2005) showed that modeling tools can be beneficial in designing a sampling strategy that accounts for biological, physical and chemical controls of transport, distribution and inactivation in natural source water and this provides a systematic way to sample and analyze water supply reservoirs for risk.

Applying models to evaluate microbial quality of irrigation waters has been done infrequently. Wilkes et al. (2009) noted that the current constraints on applying such models in the absence of monitoring data would not allow their calibration and validation with microorganisms other than E. coli.

Several ideas have been proposed to define microbial quality of irrigation waters with metrics other than specific pathogen concentration and occurrence, and to estimate these metrics using E. coli concentrations. One approach is to evaluate water quality by the total number of waterborne pathogen species. Edge et al. (2012) developed a linear regression expression for all agricultural sites in the form:

\[
\text{[Total annual number of detected pathogens] = -0.424 + 0.288} \times \log \left( \frac{\text{annual average E. coli concentration in CFU (100 mL)}^{-1}}{} \right)
\]

with an R² of 0.398. A similar concept of estimating pathogen presence–absence without differentiating species was adopted by Staley et al. (2012), who used Bayesian network modeling. Another method of quantifying microbiological water quality is to estimate the probability of enumerating a pathogen at or below several pre-set levels (i.e. estimating the empirical probability distribution of the pathogen concentration). An example is given in a recent study in southern Florida (McEgan et al., 2013) where probabilities of enumerating Salmonella at specified concentration levels appeared to be well-correlated with E. coli concentrations. This method could be combined with relating prevalence to the fecal coliform concentration range. It has yet to be established how these new water quality metrics will indicate potential health hazards from using this water for irrigation.

Relationships between E. coli and THT concentrations and illness incidence require epidemiological studies. Earlier, epidemiological studies have been carried out for recreational waters (Prüss, 1998) and provided the science base for recreation water quality criteria (Wade et al., 2003). We are not aware of any systematic epidemiological studies that relate E. coli concentrations in irrigation waters to health hazards from consuming irrigated crops. Attempts to establish a direct relationship of E. coli concentrations in water to E. coli content in produce have not been successful so far (Won et al., 2013). For example, very high concentrations of THT in irrigation water translate to a wide range of E. coli contents in produce, but no correlation is observed (Figure 2). Data on rates of pathogen survival on specific plants under defined environmental conditions and in the field, method of water application and the irrigation-to-harvest interval are needed to provide the scientific justification for establishment of standards for microbial quality of irrigation water by regulatory agencies.
Outlook

Our main conclusion is that there is a need for research to support regulatory criteria of microbiological quality of irrigation waters. The database for this review (Supplemental Table S1) contains relatively few research efforts on waters actually used for irrigation, and yet, the database covers most available peer-reviewed work on coupled coliform indicator and pathogen data in surface fresh waters. Therefore, data on microbial composition of actual irrigation waters are needed. Longitudinal studies must be performed and a large number of samples collected to assure sufficient probabilistic coverage of prevalence and concentrations of coliform indicators and pathogens. Hydrological-based fate and transport models must be used to interpret observations since they can estimate relative input of indicator and pathogen sources and transport pathways.

The longitudinal data collection is needed at the state and/or region scale. Regional efforts could lead to different critical values: for example, Vermont (Vermont Water Agency, 2009) established E. coli criteria of 77 CFU (100 mL)\(^{-1}\) which is close to thresholds found in Ontario province survey (Wilkes et al., 2009), while almost two times greater E. coli concentration was suggested for California (CSFSGLLGSC, 2006). Even larger differences can be found in standards for reclaimed wastewater to irrigate raw eaten crops in the United States (Pachepsky et al., 2011). So far, regional efforts focusing on pathogens of primary importance have provided the compelling results (Till et al., 2008).

Information on pathogen concentration in irrigation waters will continue to be limited. Therefore, information about indicators must be generated for risk assessment frameworks to prevent illnesses associated with consuming raw vegetables and fruits.

Standards of microbial water quality for irrigation cannot rely only on concentrations of indicators and/or pathogens, but must include references to crop management. The presence and concentration of pathogens in or on consumed irrigated fruits and vegetables are affected not only by their presence and concentration in irrigation water, but also by the water’s delivery method (drip, surface, or overhead), by duration of and weather conditions between irrigation and harvest, and by storage conditions before packaging and delivering to markets. Much more needs to be learned about pathogen occurrence and fate in irrigated agricultural systems to provide reliable guidance on this public health issue of paramount importance.

Declaration of interest

The United States Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here under contract DW-12-92348101 to the USDA-ARS. It has been subjected to agency review and approved for publication.

References

Ahmed W, Goonetilleke A, Gardner T. (2010). Implications of faecal indicator bacteria for the microbiological assessment of roof-harvested rainwater quality in southeast Queensland. Can J Microbiol 479:471–9.

Arvanitidou M, Stathopoulos GA, Constantinidis TC, Katsouyannopoulos V. (1995). The occurrence of Salmonella, Campylobacter and Yersinia spp. in river and lake waters. Microbiol Res 150:153–8.

Ashbolt NJ, Grabow WOK, Snoozi M. (2001). Indicators of microbial water quality. In: Fewtrell L, Bartram J, eds. Water quality: guidelines, standards and health: risk assessment and management for water related infectious diseases. London: IWA Publishing, 289–316.

Barker, J, Humphrey TJ, Brown MWR. (1999). Survival of Escherichia coli O157 in a soil protozoan: implications for disease. FEMS Microbiol Lett 173:291–5.

Biscaro GA, Gimaraes-Tomazela AB, Cruz RL, Lopes MDC. (2008). Aspectos sanitarios do cultivo da alface americana, irrigada com agua de esguiches urbanos [Sanitary aspects of lettuce (Lactuca sativa L.) culture irrigated with receiving waters of urban effluent]. Cienc Agrotecnol 32:295–301.

Bichai F, Payment P, Barbeau B. (2008). Protection of waterborne pathogens by higher organisms in drinking water: a review. Can J Microbiol 54:309–24.

Borrego JJ, Igo MAM, Vicente ADE, et al. (1987). Coliphages as an indicator of faecal pollution in water. Its relationship with indicator and pathogenic microorganisms. Water Res 21:1473–80.

Briancesco R, Libera S, Semproni M, Bonadonna L. (1999). Relationship between Cryptosporidium and microbiological water quality parameters in raw water. J Prev Med Hyg 40:39–42.

Brookes JD, Hipsey MR, Burch MD, et al. (2005). Relative value of surrogate indicators for detecting pathogens in lakes and reservoirs. Environ Sci Tech 39:8614–21.

Burton GA, Gunnison D, Lanza GR. (1987). Survival of pathogenic bacteria in various freshwater sediments. Appl Environ Microb 53:633–8.

Carter AM, Pacha RE, Clark GW, Williams EA. (1987). Seasonal occurrence of Campylobacter spp. in surface waters and their correlation with standard indicator bacteria. Appl Environ Microb 53:523–6.

Chauret C, Armstrong N, Fisher J, et al. (1995). Correlating Cryptosporidium and Giardia with microbial indicators. J Am Water Works Assoc 87:76–84.

Costán-Longares A, Montemayor M, Payán A, et al. (2008). Microbial indicators and pathogens: removal, relationships and predictive capabilities in water reclamation facilities. Water Res 42:4439–48.

Crowther J, Kay D, Wyer MD. (2002). Faecal-indicator concentrations in waters draining lowland pastoral catchments in the UK: relationships with land use and farming practices. Water Res 36:1725–34.

CSFSGLLGSC. (2006). Commodity Specific Food Safety Guidelines for the Lettuce and Leafy Greens Supply Chain. 1st ed. Available
Townsend SA. (1992). The relationships between Salmonella and faecal indicator concentrations in two pools in the Australian wet/dry tropics. J Appl Bacteriol 73:182–8.

USEPA. (1973). Water Quality Criteria. Washington, DC: National Academy of Sciences Report to the United States Environmental Protection Agency (USEPA), 350–66.

USEPA. (1976). Quality criteria for water. Washington, DC: United States Environmental Protection Agency. Available from: http://water.epa.gov/scitech/swguidance/standards/criteria/current/upload/2009_01_13_criteria_redbook.pdf [last accessed 26 Aug 2014].

USEPA. (1986). Ambient water quality criteria for bacteria – 1986. Washington, DC: United States Environmental Protection Agency. Available from: http://water.epa.gov/scitech/swguidance/standards/criteria/aqlife/upload/2009_01_13_criteria_goldbook.pdf [last accessed 26 Aug 2014].

USEPA. (2012). Recreational water quality criteria. Office of Water 820-F-12-058. Available from: http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/RWQC2012.pdf [last accessed 26 Aug 2014].

USFDA. (2013). FSMAFacts. Available from: http://www.fda.gov/downloads/Food/GuidanceRegulation/FSMA/UCM360242.pdf [last accessed 26 Aug 2014].

Vereen E, Lowrance RR, Cole DJ, Lipp EK. (2007). Distribution and ecology of Campylobacters in coastal plain streams (Georgia, United States of America). Appl Environ Microb 73:1395–403.

Vermile A, Nabi AQ, Bonadonna L, et al. (2009). Occurrence of Giardia and Cryptosporidium in Italian water supplies. Environ Monit Assess 152:203–7.

Wade TJ, Pai N, Eisenberg JNS, Colford JM. (2003). Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ Health Persp 111:1102–9.

Whelan G, Kim K, Pelton MA, et al. (2014). An integrated environmental modeling framework for performing quantitative microbial risk assessments. Environ Modell Softw 55:77–91.

Wilkes G, Edge T, Gannon V, et al. (2009). Seasonal relationships among indicator bacteria, pathogenic bacteria, Cryptosporidium oocysts, Giardia cysts, and hydrological indices for surface waters within an agricultural landscape. Water Res 43:2209–23.

Won G, Schlegel PJ, Schrock JM, LeJeune JT. (2013a). Absence of direct association between coliforms and Escherichia coli in irrigation water and on produce. J Food Protect 76:959–66.

Wu J, Long SC, Das D, Dorner SM. (2011). Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. J Water Health 9:265–78.

Vermont Water Agency. (2009). Water quality criteria for agricultural water sources and microbial testing guidelines. Available from: http://www.gaps.msue.msu.edu/ref3_022.pdf [last accessed 26 Aug 2014].