There are an estimated 17 million human diarrhoea cases annually in the United Kingdom. In 2008 and 2009, enteroaggregative *E. coli* (EAEC) were identified in 1.9% of stools. However, it remains unclear whether there is a causal link between presence of EAEC and disease. This study used bacterial load, the presence of co-infections and demographic data to assess if EAEC was independently associated with intestinal infectious disease. Quantitative real-time PCR data (Ct values) generated directly from stool specimens for several pathogen targets were analysed to identify multiple pathogens, including EAEC, in the stools of cases and healthy controls. Sensitivity and specificity using Ct value (60% and 60%) was not useful for identifying cases or controls, but an independent association between disease and EAEC presence was demonstrated: multivariate logistic regression for EAEC presence (odds ratio: 2.41; 95% confidence interval: 1.78–3.26; p<0.001). The population-attributable fraction was 3.3%. The group of bacteria known as EAEC are associated with gastrointestinal disease in at least half of the cases with EAEC positive stools. We conclude that the current definition of EAEC, by plasmid gene detection, includes true pathogens as well as non-pathogenic variants.

**Introduction**

Measuring the burden of infectious disease is essential for the rational design of public health intervention strategies and for the allocation of resources. For intestinal infectious diseases (IID) there is a massive global burden; the World Health Organization (WHO) estimates around 2 billion cases every year [5]. Detailed surveillance studies have shown that there are up to 17 million sporadic community cases of IID and one million general practitioner (GP) consultations annually in the United Kingdom (UK) [2]. Routine investigations of IID in the UK include salmonellosis, shigellosis, campylobacteriosis, cholera, infection with verotoxin-producing *Escherichia coli* O157 (VTEC), rotavirus, norovirus and parasitic infections and yet no cause is identified for over half of the laboratory-investigated diarrhoeal episodes [3]. One, often undiagnosed, potential pathogen is enteroaggregative *E. coli* (EAEC). In England, this pathotype of *E. coli*, defined by the ability to aggregate to HEp-2 cells [4], has been associated with cases of gastrointestinal infection [2,5,6] at a level comparable to *Salmonella* [6,7]. EAEC gained notoriety during a recent outbreak in Germany and France caused by an *E. coli* strain that was both a verotoxin-producing and enteroaggregative [8]. This outbreak was unusual due to the scale of morbidity and mortality, high even for VTEC infection, and the acquisition of the EAEC plasmid which may have played an important role in adherence to the human gut; the *E. coli* strain that caused the outbreak lacked the attachment and effacement (eae) gene for intimate adherence to human gut epithelium normally associated with severe disease caused by VTEC [9]. The emergence of this hybrid pathogen has been described before in 1996, when an O111:H2 strain had caused an outbreak of haemolytic uraemic syndrome (HUS) in France [10], in 1999, when an O86:H strain associated with HUS was isolated in Japan [11], and most recently in 2011, when an O111:H21 strain was associated with a family outbreak in Ireland [12]. All of these cases were associated with severe disease. It is likely that there are more cases of IID caused by EAEC and VTEC hybrids, but the EAEC pathotype is not routinely looked for.
Although EAEC itself has been associated with disease globally [13-19] including outbreaks (most notably a large outbreak in Japan involving 2,697 children [20]), a considerable proportion of healthy controls in case–control studies (16–31%) also harbour this pathotype [21-23]. Furthermore, research data describing the association of genetic factors with virulence are contradictory [21,24,25]. The reliability of virulence factors to identify EAEC for diagnostic purposes is therefore unclear [16]. The situation is further complicated by the presence of co-infections in IID [7]. When multiple pathogens are present in a diarrhoeic stool, defining which are causing the symptoms can be problematic, and as diagnostic tools improve, mixed infections in the gut are being recognised more frequently [26]. This is especially true in studies looking at EAEC infection; in Peru, for instance, multiple pathogens are found in 40% of infants with diarrhoea and with EAEC in their stool [27].

The successful completion of two IID burden studies in the UK [2,6] using quantitative PCR, presented the opportunity to investigate the causal link between gastrointestinal disease and the presence of EAEC in the stool. We estimated bacterial load for EAEC and the presence of co-infection in a well-defined population in the UK and tested the independent association between EAEC presence and disease.

**Methods**

**Datasets**

Data from two IID studies were used in this analysis: the IID1 case–control study (August 1993–January 1996) [6,7,28] and the IID2 case-only study (April 2008–March 2009) [29]. The data had been generated by testing stool samples by real-time PCR for the presence of a range of pathogens and recording the number of PCR cycles (Ct) needed before detection of product, to give a semi-quantitative estimate of bacterial load. The EAEC probe was the anti-aggregation protein transporter gene CVD432/aatD [30].

Cases of IID were defined in the same way in both studies as having had more than one loose stool, or clinically significant vomiting, over a two week period with no underlying non-infectious cause, followed by a symptom-free period of three weeks [2]. Healthy controls (IID-free) were only recruited in IID1 and were selected from the study cohort, matched for age and sex, and asked to submit a stool specimen.

The dataset for the IID1 case–control study contained 4,664 stool specimens (2,443 cases, 2,221 controls); EAEC was detected, by PCR, in 113 cases and in 38 controls but real-time Ct values (for the EAEC probe) were only available for 102 cases and 31 controls; in this study, all 151 positive cases were used for descriptive comparisons, and the 133 with Ct values for quantitative analysis.

The dataset for the IID2 case-only study [29] contained PCR Ct values from 3,966 stools (all of which were from individuals with diarrhoea); EAEC was detected in 83 of them. These data were used for burden estimations and comparisons of demographic data for cases; there had been no controls recruited in IID2 case-only study, and so IID1 data only were used for comparison of cases with controls.

**Statistical methods**

One aim of this study was to assess the methods for estimating burden of EAEC in England from the current IID2 study results. However, no controls were recruited to the IID2 study and so a receiver-operating characteristic (ROC) analysis was constructed from the case–control data (IID1), and used to look for a cut-off between case and control in the Ct values. We compared the distribution of Ct values from EAEC-positive cases and controls using Student’s t-test.

It is clear that the relationship between presence of EAEC and disease is not absolute and so several methods were used to investigate the association of EAEC with disease:

- **Carriage rates of EAEC in healthy controls, compared to other pathogens**
  - For each infection, the chi-squared test was used to test if the distribution of the pathogen between cases and controls was as expected by chance.

- **Association of disease with individual pathogens in persons with multiple pathogens in their stool**
  - For all EAEC-positive individuals with multiple pathogens (both cases and controls), we tested whether individual pathogens were equally distributed between cases and controls using chi-squared tests for independence. Because norovirus was the most common pathogen, we also compared by chi-squared test co-infection in all individuals positive for EAEC and all individuals positive for norovirus to see if the presence of other individual pathogens was dependent on infection with EAEC or norovirus.

- **Independent association of EAEC presence with disease**
  - A logistic regression of univariate and multivariate analysis was carried out using case or control as outcome, and infecting agent and age as independent variables. In this way we assessed the independent association between EAEC and disease, while controlling for other pathogens. Model results were then used to calculate the population attributable fraction (PAF):

\[
P_{a} = \frac{P_{e}(R_{a}-1)}{R_{a}}
\]

where \(P_{e}\) is the proportion of cases with the exposure (EAEC) and \(R_{a}\), the relative risk of disease. This form allows for confounding of the exposure if an adjusted RR is used, as recommended in Rockhill et al. [31]. In that case, adjusted odds ratios (OR) are substituted
Results

Defining diagnostic cut-off values for Ct values in EAEC infection

In order to investigate the link between Ct value and disease, the sensitivity and specificity of the Ct value was assessed in EAEC-positive specimens from the case–control study (dataset IID1); Ct values were obtained and included 102 cases and 31 controls. Figure 1 shows the resulting ROC curve, and Figure 2 the distribution of Ct values in cases and controls. The cut-off was chosen to balance sensitivity and specificity and was set at a Ct value of 31 (Figure 1). The ratio of false positives versus false negatives with this cut-off point was 1.09 (95% confidence interval (CI): 0.79–1.53) (Figure 2), so the total number of test-positives, although not a good diagnostic for the individual, was a reasonable estimate of the total number of cases. Importantly however, in the population studied, there was a significant association between bacterial load and disease state (p=0.039), and further investigations were carried out using the point of <40 to indicate presence of EAEC.

Descriptive statistics

To test if the analysis of data from the IID1 case–control study remained relevant in 2009, we compared the demographic data from the two periods. There was no significant difference between the rate of EAEC in the IID1 case–control study (1993–96) and IID2 case-only study (2008–09), with 1.4% and 1.9%, respectively; individuals with EAEC present in their stool were distributed evenly across all age groups in both IID1 and IID2 (chi-squared p value for non-independence: 0.253). For EAEC-positive individuals, there was no significant difference in age between cases and controls (p=0.237). We therefore believe that the epidemiology did not change significantly for EAEC infection between the two periods. Cases tended to be slightly older than controls in IID1 (mean age of cases: 30.1 years, standard deviation (SD): 24.7 years; mean age of controls 28.7 years, SD: 23.9 years; p value for difference: 0.051).

Investigation of the association of EAEC presence with disease

Carriage rates of EAEC, compared to other pathogens, in healthy controls

Submitting a stool specimen that was positive for EAEC was positively associated with having disease (Figure 3). However, one quarter of all EAEC positive individuals were asymptomatic (38/151).

Association of disease with individual pathogens in persons with multiple pathogens in their stool

The presence of co-infection was almost three times higher in EAEC-positive cases (74/113, 66%) than in EAEC-positive asymptomatic controls (9/38, 24%) (Figure 4). Cases had more multiple co-infections...
From a PCR-based case-control study (IID1). Using the Ct value as an indicator of bacterial load, we were only able to define a cut-off with 60% sensitivity and specificity. These values suggest that estimation of bacterial load by the Ct value of a quantitative PCR for virulence factors is not a useful diagnostic test for EAEC infection.

However, there was a strong association between higher load (low Ct) and being a case, so we tried to define more accurately in which positive individuals EAEC was the causal agent of diarrhoea. The bacterial load data revealed the presence of two overlapping normally distributed data sets for EAEC: one representing the load in health (controls) and one in disease (cases) (see Figure 2). We further addressed any possible confounding effects of age (i.e. acquired immunity) and co-infection using logistic regression confirmed by univariate analysis; the results showed that an individual was 2.5 times more likely to be a case than a control if they had EAEC. Therefore we concluded that EAEC was the causal agent of diarrhoea.

Investigation of the independent association of EAEC presence with disease

The logistic regression of EAEC status (but not Ct value) in univariate analysis gave an OR of 2.55 (95% CI: 1.91–3.39, p=0.001); in multivariate analysis, the OR was 2.41 (95% CI: 1.78–3.26, p<0.001). This means that among IID cases, the odds of EAEC infection were 2.5 times higher compared with asymptomatic controls. The resulting adjusted PAF was 0.033% (95% CI: 0.024–0.039), suggesting that around 3.3% of cases of IID in the UK were attributable to EAEC. This confirmed that EAEC was an independent cause of IID.

A comparison of co-infections with the most common cause of IID, norovirus, is presented in Figure 5.

Discussion

Although described as a pathogenic group of E. coli, it is well documented that EAEC may be associated with asymptomatic infection [21-23]. In this study we asked the question how much disease EAEC is responsible for. In an attempt to remove healthy carriers from the case definition (a lower bacterial load might be expected in carriers than in cases), we analysed data from a PCR-based case-control study (IID1). Using the Ct value as an indicator of bacterial load, we were only able to define a cut-off with 60% sensitivity and specificity. These values suggest that estimation of bacterial load by the Ct value of a quantitative PCR for virulence factors is not a useful diagnostic test for EAEC infection.
Figure 5
Comparison of co-infections with enteroaggregative Escherichia coli (n=113) or norovirus (n=715), United Kingdom, August 1993–January 1996

EAEC: enteroaggregative E.coli.

Co-infection with EAEC was more common than with norovirus (66% versus 43%).

The p values for individual agents are indicated on the right.

Independently associated with disease, we investigated the factors influencing this association.

Our results suggest that EAEC is common in the absence of disease. This situation is similar for gastrointestinal viral infection where post-infection levels of virus particles, although reduced, persist up to 56 days after symptoms have cleared [32,33]. Another possibility is pre-existing immunity to the infection at the time of exposure, which could result in reduced viral replication and a failure to develop symptoms. If pre-existing immunity was the cause of symptomless EAEC carriage we would expect to find an age distribution where adults are less frequently infected (older individuals have a higher chance of exposure and therefore a higher chance of immunity). The age distribution was even across the age groups and, as seen in the ROC analysis, the association between bacterial load and symptoms was not strong. Therefore we investigated an alternative explanation, the presence of a co-infecting pathogen.

The presence of increased co-infection in cases raises the possibility that the co-infecting pathogen rather than the EAEC, or a combination of both, is causing disease. To test this hypothesis we took norovirus, an infectious agent known to be present in both symptomatic and asymptomatic infection, as a comparator. As norovirus was a very common infection, we removed cases infected simultaneously with both norovirus and EAEC from the calculation: there were slightly more co-infections in EAEC-positive cases than in norovirus-positive cases (66% versus 43%). For EAEC co-infection, 12.6% were explained by enterotoxigenic E. coli (ETEC) and Shigella co-infections (Figure 5). This suggests that a proportion of EAEC cases can be explained by other pathogens (ETEC and Shigella are associated almost exclusively with symptomatic infection), but by no means all cases.

The logistic regression of co-infection univariate and multivariate was statistically significant and again confirmed that EAEC was independently associated with disease; the odds of disease were 2.4 times higher if EAEC was present than if not and were still highly significant after controlling for co-infections. The PAF adjustments indicated that EAEC would be responsible for disease in 3.3% of cases, a significant proportion in gastrointestinal disease, higher than for Salmonella [2]. Although age was an independent predictor for disease overall, controlling for age did not change the association of disease with EAEC, and there was no interaction between EAEC and age.

This study did not directly address causality over association, but we believe that bacterial variation best explains the observed association of EAEC with disease for the following reasons. There are two common arguments for EAEC being found in high levels in healthy individuals: (i) Low levels of EAEC are present in a symptomless commensal relationship in the human gut and only increase to detectable levels after infection with a true pathogen because adherence of EAEC to the gut epithelium is stronger than for other commensals; an independent association of EAEC with disease argues against this for at least half of the infections in this study. (ii) Post-infection immunity leads to carriage in apparently healthy individuals; lack of any detectable trends in age distribution and no clear association between pathogen load and disease, as seen in norovirus infection [34], suggest that acquired immunity against EAEC does not protect against infection and is therefore unlikely to lead to symptomless carriage. Transient passage, as with plant viruses, is also unlikely, as there is no known reservoir for exposure to EAEC from outside the human gut.

It seems therefore clear that some, but not all, EAEC cause disease. The explanation for this may be that EAEC are defined by in vitro phenotype rather than by the ability to cause disease: non-pathogenic EAEC, able to agglutinate cells in the laboratory but unable to cause disease in the human host, are found in controls and in co-infections with true pathogens, but pathogenic variants are found as the sole pathogen detected in diarrhoeic stools. Attempts to define genetic markers for EAEC using alternative probes still do not define those EAEC capable of causing disease: the presence of the oat (anti-aggregative transporter) [35] or aggR (a transcriptional activator) [13,18,35] does not correlate
precisely with disease, but rather with the ability to agglutinate cells in the laboratory.

It may be that the genetic factors used for EAEC diagnostics are not true virulence factors and that they rather encode the ability to adhere to human intestinal cells and allow colonisation (especially during infection with a true pathogen). It is likely that a combination of the EAEC-associated adherence factors and a true virulence factor allows EAEC to cause primary infection. This was seen in the German ST678 (O104:H4) outbreak [36], where the EAEC adherence genes were present in the same bacterial host as the Shiga-like toxin gene (stx). We suggest that an appropriate diagnostic test for pathogenic EAEC should look for the EAEC plasmid genes and other virulence factors. More work is still needed to define those other virulence factors in diarrhoeagenic EAEC.

The main limitation of this study is the lack of controls in the IID2 study. Although there were 20 years between the IID1 and IID2 studies, the demographic data for cases suggest that the epidemiology has not changed during that period. Although there may have been some change in co-infection rates, we believe the data to be relevant in 2013. Another limitation, but also a strength, of the study is the range of infectious agents identified. Small numbers in some groups of cases with co-infections (six cases or less for EAEC co-infections) means that the ability to detect statistical differences between cases and controls was limited. However, the study allowed us, for the first time, to explore the association between EAEC and all potential co-infecting agents as well as the more common pathogens norovirus (n=29) and Campylobacter (n=12).

Conclusion

This study highlights the importance of EAEC as a pathogenic group of bacteria which caused disease in more than 1% of all IID cases in the UK in 2008–09. The EAEC group is most likely to be a mixture of pathotypes which needs to be split into rational subgroups before tests for detection and typing can be implemented. Detailed studies of the genetic content of EAECs from case-control studies are warranted.

Acknowledgements

We would like to thank the study teams for allowing us access to the data generated from the two IID studies and in particular Dr Corrine Amar from the Laboratory of Gastrointestinal Pathogens (LGP) Unit, Health Protection Agency (HPA), London and Sarah O’Brien, University of Liverpool, Liverpool. We would also like to thank Prof Iruka Okeke from the Molecular Microbiology Department, Haverford College, Pennsylvania, USA and Kathie Grant from LGP, HPA, London for valuable advice with this manuscript.

Conflict of interest

None declared.

Author’s contributions

Conception, design of study, interpretation of data, drafting and revising manuscript: MA Chattaway and J Wain; Acquisition and analysis of statistical data: R Harris; Drafting manuscript and interpretation of data: T Clarence, M Iturriaza-Gomara, Claire Jenkins and John Coia.

References


