2018

Importance of case age in the purported association between phylogenetics and hemolytic uremic syndrome in Escherichia coli O157:H7 infections

G.A.M. Tarr

S. Shringi

H.N. Oltean

J. Mayer

P. Rabinowitz

See next page for additional authors

Authors
G.A.M. Tarr, S. Shringi, H.N. Oltean, J. Mayer, P. Rabinowitz, J. Wakefield, P.I. Tarr, T.E. Besser, and A.I. Phipps

Epidemiology and Infection

cambridge.org/hyg

Original Paper

Cite this article: Tarr GAM, Shringi S, Oltean HN, Mayer J, Rabinowitz P, Wakefield J, Tarr PI, Besser TE, Phipps AI (2018). Importance of case age in the purported association between phylogenetics and hemolytic uremic syndrome in *Escherichia coli* O157:H7 infections. *Epidemiology and Infection* 1–6. https://doi.org/10.1017/S0950268818001632

Received: 9 November 2017
Revised: 13 February 2018
Accepted: 18 May 2018

Key words:
Epidemiology; *Escherichia coli* (*E. coli*); phylogenetics; Shiga toxin-producing *E. coli*

Author for correspondence:
G. A. M. Tarr, E-mail: gillian.tarr@ahs.ca

Importance of case age in the purported association between phylogenetics and hemolytic uremic syndrome in *Escherichia coli* O157:H7 infections

G. A. M. Tarr[1], S. Shringi[2], H. N. Oltean[3], J. Mayer[4,5], P. Rabinowitz[6], J. Wakefield[7], P. I. Tarr[8], T. E. Besser[2] and A. I. Phipps[4]

[1]Department of Pediatrics, University of Calgary Cumming School of Medicine, Calgary, Alberta, CA; [2]Veterinary Microbiology and Pathology, Washington State University, Pullman, Washington, USA; [3]Washington State Department of Health, Shoreline, Washington, USA; [4]Department of Epidemiology, University of Washington, Seattle, Washington, USA; [5]Department of Geography, University of Washington, Seattle, Washington, USA; [6]Department of Environmental and Occupational Health Sciences and Center for One Health Research, University of Washington, Seattle, Washington, USA; [7]Departments of Biostatistics and Statistics, University of Washington, Seattle, Washington, USA and [8]Department of Pediatrics, Washington University in St. Louis School of Medicine, St. Louis, Missouri, USA

Abstract

Escherichia coli O157:H7 is the largest cause of hemolytic uremic syndrome (HUS). Previous studies proposed that HUS risk varies across the *E. coli* O157:H7 phylogenetic tree (hypervirulent clade 8), but the role of age in the association is unknown. We determined phylogenetic lineage of *E. coli* O157:H7 isolates from 1160 culture-confirmed *E. coli* O157:H7 cases reported in Washington State, 2004–2015. Using generalised estimating equations, we tested the association between phylogenetic lineage and HUS. Age was evaluated as an effect modifier. Among 1082 *E. coli* O157:H7 cases with both phylogenetic lineage and HUS status (HUS $n = 76$), stratified analysis suggested effect modification by age. Lineages IIa and IIb, relative to Ib, did not appear associated with HUS in children 0–9-years-old. For cases 10–59-years-old, lineages IIa and IIb appeared to confer increased risk of HUS, relative to lineage Ib. The association reversed in ⩾60-year-olds. Results were similar for clade 8. Phylogenetic lineage appears to be associated with HUS risk only among those ⩾10-years-old. Among children <10, the age group most frequently affected, lineage does not explain progression to HUS. However, lineage frequency varied across age groups, suggesting differences in exposure and/or early disease manifestation.

Introduction

Although significant progress has been made in reducing the incidence and impact of *Escherichia coli* O157:H7, it remains the largest cause of post-diarrhoeal hemolytic uremic syndrome (HUS) [1]. HUS incidence varies by age, with the greatest burden among children <5-years-old [2–5].

Beyond age, pathogen characteristics are an important factor in determining progression to HUS. Shiga toxin (Stx), *E. coli* O157:H7's cardinal virulence factor, can be encoded by multiple genes (*stx1* and *stx2*), with some genotypes more frequently associated with HUS than others [6–9]. A study in 2008 identified a subtype of *E. coli* O157:H7, termed clade 8, associated with increased risk of HUS [10]. Although numerous studies have investigated virulence factor expression that may be responsible for this association [11–14], studies confirming the association have been limited, suggesting an effect of varying strength and specificity [15–17].

The phylogenetic definition of the *E. coli* O157:H7 serotype has advanced since the 2008 discovery of the putatively hypervirulent clade 8 and it is unknown whether branches of the updated phylogenetic tree are also associated with HUS. In contrast to the earlier tree, Bono *et al.* [18] reported a tree of phylogenetic lineages that drew on a large pool of systematically chosen single nucleotide polymorphisms (SNPs) and incorporated isolates from a diverse set of sources. This tree was further developed by Jung *et al.* [19].

In a population-based cohort of 1160 *E. coli* O157:H7 cases in Washington State, reported 2005–2014, we sought to use the updated lineages to confirm and refine the role of phylogenetics in HUS risk proposed by Manning *et al.* [10]. Given the higher incidence of HUS among young children and the preponderance of clade 8 strains isolated from children by Manning *et al.* [10], we also investigated the role of age in the association between phylogenetic lineage and HUS, evaluating it as a potential confounder and effect modifier.

CAMBRIDGE UNIVERSITY PRESS

Methods

Study setting and design

We conducted a population-based retrospective cohort study of all culture-confirmed *E. coli* O157:H7 cases reported to the Washington State Department of Health (DOH) from 2005 through 2014. Mandatory Shiga toxin-producing *E. coli* case reporting occurs primarily through diagnostic laboratories and healthcare providers. Local health jurisdiction personnel use a standardised DOH case report form to obtain demographic information, potential exposures and details of the course of illness.

We confirmed HUS status during a review of all reported, hospitalised, culture-confirmed *E. coli* O157:H7 cases from the study period. HUS was defined as a hematocrit <30%, platelet count <150 000/mm^3 and serum creatinine concentration above the normal for age. All criteria needed to be met on the same day. Non-hospitalised cases were considered to not have HUS because all patients with HUS would be hospitalised due to the severity of disease.

The Washington State Institutional Review Board designated this study as exempt.

Isolate typing

All *E. coli* O157:H7 isolates were sent to DOH for microbiological confirmation and pulsed-field gel electrophoresis (PFGE) analysis. We obtained these isolates from DOH and determined their lineage according to the phylogenetic tree developed by Bono *et al.* [18] and expanded by Jung *et al.* [19]. The 48-plex SNP assay developed by Jung *et al.* [19] was used to type 793 of the 1160 isolates. Isolates that did not undergo SNP-typing were then assigned the lineage of a SNP-typed isolate with the same PFGE profile (Supplementary Material), for a total of 1121 isolates with an assigned lineage. Of the 39 excluded isolates, six were biochemically atypical *E. coli* O157:H7 and 33 were not available for typing (Supplementary Material).

Based on Jung *et al.*'s categorisation as clinical, bovine-biased, or sparsely represented lineages [19], we retained the clinical lineages Ib, IIa and IIb as separate categories. The bovine-biased and remaining lineages were grouped into a clinically rare category, reflecting the low frequency of isolating these groups from human cases.

A subset of 480 isolates was also typed using a 32-plex SNP assay to determine Manning clade [10]. We used PFGE type to infer Manning clade for an additional 422 untyped isolates (Supplementary Material). Distribution of Manning clades by Bono-Jung lineages is shown in Supplementary Table S1.

A subset of 453 isolates underwent Stx-encoding bacteriophage insertion (SBI) typing. The SBI typing methods, which use PCR to detect 12 targets including *stx1*, *stx2a* and *stx2c*, have been described [19, 20].

Statistical analysis

Case data were merged with isolate typing results using a unique identifier and the dataset was de-identified before analysis. Age [21–24] and sex [25–27] were considered *a priori* confounders. Distributions of other potential confounders were summarised in contingency tables. Aside from age, no examined variable was significantly associated with both lineage and HUS.

Logistic regression with generalised estimating equations (GEE) was used to estimate the association between lineage and HUS. Lineage was modeled as a categorical, group-level variable, with the most common lineage (Ib) as the reference category and groups defined by PFGE types. An exchangeable working correlation matrix was used for all analyses. Robust standard errors calculated using the sandwich estimator accounted for any potential misspecification of the correlation structure. HUS is sufficiently rare that the odds ratio (OR) calculated from model coefficients could be interpreted as the risk of HUS associated with a lineage (e.g. IIa) relative to lineage Ib. A 95% confidence interval (CI) was estimated for each OR. Age, modelled as a continuous variable, and sex were added as covariates in the adjusted model. To examine effect modification, the sex-adjusted GEE model was stratified by age group and the lineage OR estimates were compared across strata. Unadjusted, adjusted and stratified analyses were also conducted for the association of Manning clade and HUS (Supplementary Material).

To understand the impact of our use of multiple isolates per PFGE type, we simulated the method used in previous studies [10, 16] for comparison with our results. One isolate per PFGE type was randomly drawn and used in a single-level adjusted logistic regression. For each of 10 000 repetitions of this process, the coefficient estimates for lineages IIa and IIb and the associated *P*-values were recorded. We examined the distribution of coefficient estimates by lineage and calculated the proportion that was statistically significant at $P < 0.05$.

The frequency of Shiga toxin subtypes was summarised by lineage and HUS status. Formal mediation analysis of the role of *stx* genotypes in the association between lineage and HUS was planned. However, most major *stx* genotypes were too highly correlated with lineage to differentiate the direct effect of lineage and that mediated by *stx* genotype.

R [28] was used for all analyses.

Results

There were 1160 culture-confirmed *E. coli* O157:H7 cases reported to DOH during the 10-year-period. Validated HUS status was available for 1082 cases of the 1121 cases with an assigned lineage; the HUS definition was met for 76 (7.0%). HUS status differed by age, with children <5-years-old constituting over half of HUS cases but less than one-fourth of non-HUS cases (Table 1). The case fatality was 3.9% among HUS cases and 0.4% among non-HUS cases.

Phylogenetic association with HUS

In the unadjusted GEE model, lineage IIb was associated with increased risk of HUS relative to lineage Ib (OR = 1.65, 95% CI 1.05–2.60) (Table 2). There was no elevation in HUS risk among lineage IIa cases, compared with lineage Ib cases. No HUS cases occurred in the group of rare lineages; effect estimates for this group are not presented because of statistical instability. After adjustment for age and sex, the association between IIb and HUS was attenuated and no longer distinguishable from the null (OR = 1.43, 95% CI 0.90–2.25).

The proportion of *E. coli* O157:H7 infections caused by lineage IIb strains decreased with age and those caused by Ib strains increased. An effect of lineage on HUS risk could not be established in 0–4 or 5–9 year-olds. In 10–19 and 20–59-year-olds, lineages IIa and IIb were associated with an increased risk of HUS, relative to lineage Ib, with effect estimates highest in the latter group (IIa OR = 12.7, 95% CI 1.57–103; IIb OR = 8.50, 95% CI

https://doi.org/10.1017/S0950268818001632

Table 1. Frequency of case characteristics of patients reported in Washington State with confirmed *E. coli* O157:H7 infection, 2005–2014

	No HUS ($n = 1042$)	HUS ($n = 76$)
Sex		
Female	591 (57.0%)	46 (60.5%)
Male	445 (43.0%)	30 (39.5%)
Age group (years)		
<5	227 (21.8%)	41 (53.9%)
5–9	145 (13.9%)	18 (23.7%)
10–19	184 (17.7%)	5 (6.6%)
20–59	358 (34.4%)	7 (9.2%)
⩾60	127 (12.2%)	5 (6.6%)
Ethnicity		
Hispanic or Latino	97 (12.5%)	8 (11.1%)
Not Hispanic or Latino	676 (87.5%)	64 (88.9%)
Race		
American Indian or Alaskan Native	10 (1.3%)	2 (2.8%)
Asian	55 (7.1%)	4 (5.6%)
Black	25 (3.2%)	0
Multiracial	11 (1.4%)	0
Native Hawaiian or Pacific Islander	5 (0.6%)	0
Other	20 (2.6%)	4 (5.6%)
White	651 (83.8%)	62 (86.1%)
Phylogenetic lineage		
Ib	531 (52.7%)	37 (49.3%)
IIa	235 (23.3%)	18 (24.0%)
IIb	173 (17.2%)	20 (26.7%)
Rare[a]	68 (6.8%)	0
Outbreak-associated[b]	105 (10.1%)	7 (9.2%)
Most likely source of infection		
Animal	70 (17.5%)	6 (18.8%)
Environment	43 (10.7%)	0
Food	201 (50.1%)	18 (56.2%)
Person	75 (18.7%)	7 (21.9%)
Water	12 (2.9%)	1 (3.1%)
Contact with lab-confirmed case		
Yes	127 (13.7%)	14 (20.3%)
No	799 (86.3%)	55 (79.7%)
Direct animal contact		
Yes	490 (55.9%)	39 (60.0%)
No	387 (44.1%)	26 (40.0%)
Underlying condition		
Yes	101 (10.7%)	9 (13.6%)
No	842 (89.3%)	57 (86.4%)
Bloody diarrhoea		
Yes	870 (87.1%)	73 (96.1%)
No	129 (12.9%)	3 (3.9%)
Received dialysis		
Yes	2 (0.2%)	40 (52.6%)
No	1036 (99.8%)	36 (47.4%)
Documented death[c]	4 (0.4%)	3 (3.9%)

HUS, hemolytic uremic syndrome.
[a]'Rare' lineages include 12 different lineages.
[b]Whether a case was associated with an outbreak was not reported for most cases, so only positive responses are shown.
[c]Death status was not reported in most cases. There were eight reported deaths. Only seven are shown in the table. The eighth was hospitalised, but the chart could not be abstracted to determine HUS status.

1.13–63.7). There were no lineage IIa or IIb HUS cases among ⩾60-year-olds (Table 2).

Results of unadjusted, adjusted and stratified analyses assessing the association between Manning clade and HUS were consistent with those seen for lineages (Supplementary Table S2).

In simulations of selecting and analysing only one isolate per PFGE type, the distribution of effect sizes for the association between lineage IIa and HUS was centered near 0 (OR = 1) and 0.1% of estimates had $P < 0.05$ (Supplementary Fig. S1). This is consistent with our adjusted effect estimate for lineage IIa (Table 2). The distribution of effect sizes for lineage IIb was centred near 2 (OR > 7) with $P < 0.05$ for 25% of simulations (Supplementary Fig. S1). This is not consistent with our results for lineage IIb using all isolates (OR = 1.43, $P = 0.13$) (Table 2).

Shiga toxin genotype

Shiga toxin genotype was determined for 469 cases, 453 of which also had a validated HUS status. Distribution of *stx* genotypes by lineage showed that 92% of isolates contained *stx2a*, whether alone or in combination with another *stx* gene (Table 3). Lineage Ib isolates were dominated by the *stx1-stx2a* genotype (90%). Lineage IIa isolates were predominantly the *stx2a-stx2c* genotype (84%). Most lineage IIb isolates (94%) had only the *stx2a* gene. Six isolates had none of the three probed *stx* genes at the time of typing. Relative to the frequency of HUS among cases infected with *stx1-stx2a* and *stx2a-stx2c* strains (11% and 12%, respectively), cases infected with *stx2a*-only strains had a higher frequency of HUS (21%).

Discussion

The results of this study do not support an association between phylogenetic lineage (or clade) and HUS for children <10, the age group with the greatest burden of *E. coli* O157:H7 and HUS. While lineage IIb was associated with increased risk of HUS in unadjusted analysis, stratifying by age indicated an increased risk of HUS associated with lineages IIa and IIb, relative to lineage Ib, only among 10–59-year-olds. In the eldest group, lineage Ib conferred greater HUS risk than either lineage IIa or IIb. Our analysis of the risk associated with Manning clade 8 was consistent with our lineage IIa/IIb results.

https://doi.org/10.1017/S0950268818001632

Table 2. Association of *E. coli* O157:H7 phylogenetic lineage and HUS

	N HUS/Total	OR	95% CI	*P*
Crude				
Lineage Ib	37/568	1	–	–
Lineage IIa	18/253	1.11	0.63–1.96	0.711
Lineage IIb	20/193	1.65	1.05–2.60	0.031
Adjusted[a]				
Lineage Ib	37/561	1	–	–
Lineage IIa	18/253	0.98	0.54–1.78	0.937
Lineage IIb	20/193	1.43	0.90–2.25	0.126
Age-stratified: 0–4-years-old[b]				
Lineage Ib	22/118	1	–	–
Lineage IIa	8/71	0.61	0.27–1.39	0.24
Lineage IIb	10/62	0.73	0.39–1.36	0.31
Age-stratified: 5–9-years-old[b]				
Lineage Ib	8/79	1	–	–
Lineage IIa	4/31	1.39	0.62–3.12	0.42
Lineage IIb	6/34	2.38	0.79–7.13	0.12
Age-stratified: 10–19-years-old[b]				
Lineage Ib	1/95	1	–	–
Lineage IIa	2/50	4.92	0.89–27.1	0.067
Lineage IIb	2/30	4.99	0.94–26.4	0.059
Age-stratified: 20–59-years-old[b]				
Lineage Ib	1/200	1	–	–
Lineage IIa	4/78	12.7	1.57–103	0.017
Lineage IIb	2/48	8.50	1.13–63.7	0.037
Age-stratified: ⩾60-years-old[b]				
Lineage Ib	5/75	1	–	–
Lineage IIa	0/23	0	0–0	<0.001
Lineage IIb	0/19	0	0–0	<0.001

Logistic regression, using GEE, of HUS status on phylogenetic lineage. No HUS occurred in the group of cases infected with rare lineages, so results are not shown for this group.
[a]Model adjusted for age as a continuous variable and sex.
[b]Model adjusted for sex. CI, confidence interval; GEE, generalised estimating equations; HUS, hemolytic uremic syndrome; OR, odds ratio.

Age has long been considered the strongest predictor of progression to HUS among those with *E. coli* O157:H7 infection and our results are consistent with that: 15.9% of children <10-years-old progressed to HUS, compared to 2.5% of individuals ⩾10. Similarly, the incidence of HUS in our study was 6.79 per 100 000 <10-year-olds, compared with 0.29 per 100 000 ⩾10-year-olds. The lack of association we found between lineage and HUS among those aged <10 years suggests that differential infection by high virulence lineages does not explain why young children are more likely to progress to HUS. However, our findings show that lineage IIb strains disproportionately establish disease in young children, driving the observed unadjusted association between lineage IIb and HUS and suggesting that there is a difference in either exposure or early disease manifestation that leads to more IIb-infected cases being reported in this age group than cases infected with other lineages.

Among those aged 10 and over, we observed substantially more reported cases infected by lineage Ib than lineages IIa and IIb, which may indicate less exposure to lineage IIa and IIb strains or greater difficulty for these strains in establishing disease. However, if IIa or IIb strains are successful in establishing disease, they appear more likely to cause HUS than the more common lineage Ib strains. The eldest group, ⩾60-year-olds, appears to be an exception, with higher risk associated with lineage Ib strains. Individuals ⩾60-years-old have a slightly higher incidence of HUS (0.58 per 100 000) than 10–59-year-olds (0.26 per 100 000), and *E. coli* O157:H7 outbreaks have occurred in nursing homes [24], making this age group of particular interest. The reversal of the association in the eldest group is curious and with a low number of cases among older children and adults, urges caution in interpreting our results in 10–59 and ⩾60-year-olds.

In a 2008 study of 333 Michigan cases with unique PFGE fingerprints, Manning *et al.* [10] identified a sevenfold increased odds of HUS among patients infected with *E. coli* O157:H7 clade 8 strains after adjustment for age (0–18 *vs.* 19–64), sex and symptoms. Subsequent studies have also suggested an association of varying magnitudes between clade 8 and HUS [15, 16]. Lineages IIa and IIb in the present study overlap with clade 8 and show an elevation in risk of HUS only among 10–59-year-olds. There are multiple reasons why our results may have differed from those of others. First, some previous studies have either not adjusted for age [15] or adjusted by large age groups [10], increasing the potential for residual confounding. Only one previous study stratified by age [16]. In our analysis, sensitive age groups defined based on the epidemiology of the disease were critical in better understanding the association.

Second, both Manning *et al.* [10] and Iyoda *et al.* [16] used one representative isolate from each outbreak or PFGE-defined strain. We demonstrated through simulation that studies using only one isolate per strain had an average effect estimate higher than that obtained using all isolates for lineage IIb and that for lineage IIb, 25% of analyses would appear statistically significant merely by chance. This finding emphasises the importance of incorporating the complete data.

Third, previous studies relying on logistic regression [10, 16] appear to have modelled each clade as an independent variable, interpreting estimates as the odds of HUS in one clade *vs.* all other clades. This method introduces perfect multicollinearity, which can induce large, unpredictable biases in point estimates and standard errors [29, 30]. Perfect multicollinearity also gives the OR dubious interpretability, because, by definition, you cannot hold the other clades constant (at 0 or 1) and change the clade of interest from 0 to 1. To avoid this pitfall, we modelled lineage (and clade) as a categorical value in which the most common lineage Ib (clade 2/3) was used as the reference category.

Finally, the only other study to consider effect modification by age, Iyoda *et al.* [16] reported an OR for clade 8 of 6.1 for 0–9-year-olds and 3.1 for children and adults ⩾10 years. Their results are in contrast to those we report here, potentially because of their use of asymptomatic controls. We estimated the odds of HUS for ill *E. coli* O157:H7 cases, thus estimating virulence, the probability of progressing from non-severe to severe disease.

https://doi.org/10.1017/S0950268818001632

Table 3. Distribution of Shiga toxin genotypes by phylogenetic lineages

	All lineages *n* (%)	Lineage Ib *n* (%)	Lineage IIa *n* (%)	Lineage IIb *n* (%)	Rare lineages *n* (%)
No *stx*	6 (1.3)	2 (0.9)	1 (0.8)	3 (3.3)	0
HUS	0	0	0	0	–
stx1	6 (1.3)	4 (1.9)	0	0	2 (7.4)
HUS	1 (1.7)	1 (3.7)	–	–	0
stx1-stx2a	192 (42.4)	192 (90.1)	0	0	0
HUS	22 (36.7)	22 (81.5)	–	–	–
*stx1-stx2**c***	15 (3.3)	0	0	0	15 (55.6)
HUS	0	–	–	–	0
stx2a	117 (25.8)	13 (6.1)	19 (15.4)	85 (94.4)	0
HUS	24 (40.0)	4 (14.8)	2 (13.3)	18 (100)	–
*stx2a-stx2**c***	106 (23.4)	1 (0.5)	103 (83.7)	2 (2.2)	0
HUS	13 (21.7)	0	13 (86.7)	0	–
stx2c	11 (2.4)	1 (0.5)	0	0	10 (37.0)
HUS	0	0	–	–	0

HUS, hemolytic uremic syndrome; *stx*, Shiga toxin gene.
No isolates were observed with the *stx1-stx2a-stx2c* genotype.

Comparing HUS cases with asymptomatic carriers mixes virulence with pathogenicity, the probability of becoming ill if infected.

We observed a very close correlation of lineage and *stx* genotype, which is similar to previous studies [19, 31]. This may be suggestive of a major role of *stx* genotype in the association between lineage and HUS. Other cohorts, including one of <10-year-olds, have shown *stx2a*-only and, to a lesser degree, *stx2a-stx2c* genotypes associated with progression to HUS [6, 7, 32]. These are also the most common genotypes among clade 8 isolates [10, 16, 17] and studies of clade 8 isolates have described the potential for high Stx2 production [12, 33, 34]. Our analysis, which shows that most lineage IIa strains carry *stx2a-stx2c* and most IIb strains carry only *stx2a*, is consistent with these studies.

Our study was limited to reported cases, which are likely more severe than unreported cases. Our results can therefore not be extended to unreported cases. However, it is unlikely that any HUS cases went unreported due to the severity of the condition. Our study also included cases from only Washington State, potentially limiting its generalisability to areas with differing *E. coli* O157:H7 populations. Indeed, previous work has suggested local *E. coli* O157:H7 circulation in Washington [35], emphasising the importance of small geographic areas in the bacteria's population dynamics. The strains composing each lineage may differ in other geographic regions and those within-lineage differences could alter the association observed with HUS. However, it is reassuring that a large number of isolates in our study are from the most commonly isolated PFGE types in the USA.

We were also not able to assign phylogenetic lineage to 39 isolates, one of which was identified as a HUS case. These isolates tended to be from earlier in the study period, indicating that they are not missing completely at random. The composition of the bacterial population shifted slightly during the study period [35], with lineage Ib more dominant early in the period. However, the small number of untyped isolates relative to the whole sample likely did not alter our results.

Over 75% of our HUS cases were in children <10-years-old, giving us limited precision to estimate the effect of lineage on HUS in older children and adults. This is reflected in the large confidence intervals around estimates for the 10–19 and 20–59 age groups. A larger sample of cases ⩾10-years-old would provide a better estimate of the true effect of lineage on HUS in this age group. However, we are confident in our estimates for the effect in young children, the age group with the highest incidence of both *E. coli* O157:H7 and HUS.

The lack of sufficient variability of most *stx* genotypes in a single lineage precluded formal mediation analysis. It is possible that with a much larger sample one could differentiate the direct effect of lineage on progression to HUS from the effect mediated by *stx* genotype. Ideally, mediation should be examined stratified by age group, to reflect the apparent effect modification of the overall association.

This study benefited from over 1100 *E. coli* O157:H7 cases, including 76 HUS cases. HUS outcomes were validated with hospital records using a standardised definition to ensure the comparability of our outcome. By employing correlated data methods, we were able to incorporate data from the entire cohort instead of limiting the study to representative isolates from each PFGE type, which our simulation study showed is an important step in accurately estimating the association. By using a consistent reference group (lineage Ib), we were also able to avoid the perfect multicollinearity of previous studies, reducing bias and allowing meaningful interpretation of our effect estimates. Applying these methods to the Bono-Jung lineages and Manning clades produced consistent results.

This study demonstrates that *E. coli* O157:H7 phylogenetic lineage likely only contributes to HUS risk among older children and adults. Further studies are needed to confirm this association, given the rarity of the disease among adults. In young children, the proportion of infections caused by lineage IIb strains was higher than in older groups. It will be important to determine whether this is driven by differences between age groups in

https://doi.org/10.1017/S0950268818001632

exposure, transmission and/or early disease development. Additionally, given the close correspondence of lineage and *stx* genotype, learning how exposure and early illness differ across lineages may translate to prevention opportunities for the strains that tend to carry more virulent *stx* genotypes.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0950268818001632

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Acknowledgements. This work was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health (G.A.M.T., grant number T32ES015459); the National Institute of Allergy and Infectious Disease of the National Institutes of Health (G.A.M.T., grant number F31AI126834); and the US Department of Agriculture National Institute of Food and Agriculture (T.E.B., grant numbers 2009-04248, 2010-04487). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or US Department of Agriculture.

Conflict of interest. None.

References

1. **Banatvala N *et al.*** (2001) The United States national prospective hemolytic uremic syndrome study: microbiologic, serologic, clinical, and epidemiologic findings. *The Journal of Infectious Diseases* **183**, 1063–1070.
2. **Centers for Disease Control and Prevention** (2016) Summary of notifiable diseases, 2014. *MMWR Morbidity and Mortality Weekly Report* **63**, 1–154.
3. **Marder EP *et al.*** (2017) Incidence and trends of infections with pathogens transmitted commonly through food and the effect of increasing use of culture-independent diagnostic tests on surveillance — foodborne diseases active surveillance network, 10 U.S. sites, 2013–2016. *MMWR Morbidity and Mortality Weekly Report* **66**, 397–403.
4. **Crim SM *et al.*** (2015) Preliminary incidence and trends of infection with pathogens transmitted commonly through food – foodborne diseases active surveillance network, 10 U.S. sites, 2006–2014. *MMWR Morbidity and Mortality Weekly Report* **64**, 495–498.
5. **Crim SM *et al.*** (2014) Incidence and trends of infection with pathogens transmitted commonly through food--foodborne diseases active surveillance network, 10 U.S. Sites, 2006–2013. *MMWR Morbidity and Mortality Weekly Report* **63**, 328–332.
6. **Friedrich AW *et al.*** (2002) *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. *The Journal of Infectious Diseases* **185**, 74–84.
7. **Persson S *et al.*** (2007) Subtyping method for *Escherichia coli* Shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *Journal of Clinical Microbiology* **45**, 2020–2024.
8. **Eklund M, Leino K and Siitonen A** (2002) Clinical *Escherichia coli* strains carrying stx genes: stx variants and stx-positive virulence profiles. *Journal of Clinical Microbiology* **40**, 4585–4593.
9. **Luna-Gierke RE *et al.*** (2014) Outbreaks of non-O157 Shiga toxin-producing *Escherichia coli* infection: USA. *Epidemiology and Infection* **142**, 2270–2280.
10. **Manning SD *et al.*** (2008) Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4868–4873.
11. **Abu-Ali GS *et al.*** (2010) Increased adherence and expression of virulence genes in a lineage of *Escherichia coli* O157:H7 commonly associated with human infections. *PLoS ONE* **5**, e10167.
12. **Neupane M *et al.*** (2011) Shiga toxin 2 overexpression in *Escherichia coli* O157:H7 strains associated with severe human disease. *Microbial Pathogenesis* **51**, 466–470.
13. **Pianciola L *et al.*** (2014) Genotypic characterization of *Escherichia coli* O157:H7 strains that cause diarrhea and hemolytic uremic syndrome in Neuquen, Argentina. *International Journal of Medical Microbiology: IJMM* **304**, 499–504.
14. **Amigo N *et al.*** (2016) Overexpressed proteins in hypervirulent clade 8 and clade 6 strains of *Escherichia coli* O157:H7 compared to *E. coli* O157:H7 EDL933 clade 3 strain. *PLoS ONE* **11**, e0166883.
15. **Soderlund R *et al.*** (2014) Molecular typing of *Escherichia coli* O157:H7 isolates from Swedish cattle and human cases: population dynamics and virulence. *Journal of Clinical Microbiology* **52**, 3906–3912.
16. **Iyoda S *et al.*** (2014) Phylogenetic clades 6 and 8 of enterohemorrhagic *Escherichia coli* O157:H7 with particular stx subtypes are more frequently found in isolates from hemolytic uremic syndrome patients than from asymptomatic carriers. *Open Forum Infectious Diseases* **1**, ofu061.
17. **Haugum K *et al.*** (2011) Detection of virulent *Escherichia coli* O157 strains using multiplex PCR and single base sequencing for SNP characterization. *Journal of Applied Microbiology* **110**, 1592–1600.
18. **Bono JL *et al.*** (2012) Phylogeny of Shiga toxin-producing *Escherichia coli* O157 isolated from cattle and clinically ill humans. *Molecular Biology and Evolution* **29**, 2047–2062.
19. **Jung WK *et al.*** (2013) Lineage and genogroup-defining single nucleotide polymorphisms of *Escherichia coli* O157:H7. *Applied and Environmental Microbiology* **79**, 7036–7041.
20. **Shringi S *et al.*** (2012) Carriage of stx2a differentiates clinical and bovine-biased strains of *Escherichia coli* O157. *PLoS ONE* **7**, e51572.
21. **Rogers MF *et al.*** (1986) A population-based study of hemolytic-uremic syndrome in Oregon, 1979–1982. *American Journal of Epidemiology* **123**, 137–142.
22. **Taylor DN *et al.*** (1988) Clinical and microbiologic features of Shigella and enteroinvasive *Escherichia coli* infections detected by DNA hybridization. *Journal of Clinical Microbiology* **26**, 1362–1366.
23. **Ostroff SM, Kobayashi JM and Lewis JH** (1989) Infections with *Escherichia coli* O157:H7 in Washington State. *The First Year of Statewide Disease Surveillance. JAMA* **262**, 355–359.
24. **Reiss G *et al.*** (2006) *Escherichia coli* O157:H7 infection in nursing homes: review of literature and report of recent outbreak. *Journal of the American Geriatrics Society* **54**, 680–684.
25. **Rowe PC *et al.*** (1991) Epidemiology of hemolytic-uremic syndrome in Canadian children from 1986 to 1988. *The Canadian Pediatric Kidney Disease Reference Centre. The Journal of Pediatrics* **119**, 218–224.
26. **Rivas M *et al.*** (2008) Risk factors for sporadic Shiga toxin-producing *Escherichia coli* infections in children, Argentina. *Emerging Infectious Diseases* **14**, 763–771.
27. **Al-Jader L *et al.*** (1999) Outbreak of *Escherichia coli* O157 in a nursery: lessons for prevention. *Archives of Disease in Childhood* **81**, 60–63.
28. **R Core Team** (2015) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
29. **Tu YK, Clerehugh V and Gilthorpe MS** (2004) Collinearity in linear regression is a serious problem in oral health research. *European Journal of Oral Sciences* **112**, 389–397.
30. **Vatcheva KP *et al.*** (2016) Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology* (*Sunnyvale*) **6**, 227.
31. **Mellor GE *et al.*** (2015) Geographically distinct *Escherichia coli* O157 isolates differ by lineage, Shiga toxin genotype, and total Shiga toxin production. *Journal of Clinical Microbiology* **53**, 579–586.
32. **Jelacic S *et al.*** (2002) ABO and P1 blood group antigen expression and stx genotype and outcome of childhood *Escherichia coli* O157:H7 infections. *The Journal of Infectious Diseases* **185**, 214–219.
33. **Ogura Y *et al.*** (2015) The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. *Scientific Reports* **5**, 16663.
34. **Amigo N *et al.*** (2015) Clade 8 and clade 6 strains of *Escherichia coli* O157:H7 from cattle in Argentina have Hypervirulent-like phenotypes. *PLoS ONE* **10**, e0127710.
35. **Tarr GAM *et al.*** (2018) Geogenomic segregation and temporal trends of human pathogenic *Escherichia coli* O157:H7, Washington, USA, 2005–2014(1). *Emerging Infectious Diseases* **24**, 32–39.