

The use of a new virtual cohort study design to investigate an outbreak of *E. coli* O157 linked to a supermarket delicatessen

G. McCARTNEY¹*, J. COWDEN², S. MURRAY¹ AND S. AHMED¹

¹ Public Health, NHS Greater Glasgow and Clyde, Glasgow, Scotland, UK

² Health Protection Scotland, Glasgow, Scotland, UK

(Accepted 8 January 2010; first published online 10 February 2010)

SUMMARY

In outbreaks of foodborne disease associated with retail outlets, the outlet often closes as a precaution before the specific food vehicle has been identified. Suspect food vehicles may be named as part of general control measures. A conventional case-control study cannot be performed because both cases and potential controls are likely to be aware of the hypothesis and therefore potentially biased. Modern sales recording systems in many food retail outlets may provide a basis for constructing a virtual cohort and allow a statistical inference to be made about various possible vehicles of infection. In 2007, an outbreak of *E. coli* O157 infection in Paisley, Scotland, was linked to cooked meat from a supermarket delicatessen using descriptive epidemiology. Construction of a virtual cohort allowed a relative risk and confidence interval to be estimated which supported the hypothesis of cooked beef topside being the vehicle of infection. This novel method could be valuable in the investigation of future outbreaks.

Key words: Cohort analysis, epidemiology, *Escherichia coli*, outbreaks.

INTRODUCTION

Many food retail outlets in the UK have now installed sophisticated recording systems of their sales to facilitate ‘just in time’ delivery schedules. We used an example of such a recording system to test the hypothesis that cooked cold beef topside sold from the delicatessen in one branch of a national supermarket chain was responsible for an outbreak of *E. coli* O157 phage type 2 infection in Paisley (Scotland), during July/August 2007 [1].

The outbreak comprised 10 adult cases of *E. coli* O157 infection, with eight primary cases and two

secondary cases resident in two neighbouring post-code areas. The dates of onset of the primary cases ranged from 1 to 11 August. A case with the same uncommon phage type in Scotland in 2007 and an indistinguishable pulsed-field electrophoresis pattern was identified in a resident of a postcode area ~13 km distant, but this case could not be linked to the outbreak in person, place or time. The descriptive epidemiology suggested that the most likely vehicle of infection was cooked meat purchased from the delicatessen of a branch of a national supermarket chain between 31 July and 2 August 2007 (inclusive). A conventional case-control study would have been inappropriate since the naming of the supermarket chain, branch, and suspected vehicles of infection as part of the control measures put in place during the outbreak made the recruitment of blinded subjects (i.e. cases and controls whom we could be confident

* Author for correspondence: Dr G. McCartney, Specialist Registrar in Public Health, currently attached to the MRC Social and Public Health Sciences Unit, 4 Lilybank Gardens, Hillhead, Glasgow, G12 8RZ.
(Email: Gerry@sphsu.mrc.ac.uk)

were ignorant of the hypotheses being tested) impossible [2]. Neither was a traditional cohort study possible as there was no defined population to collect data from. These problems are common in outbreak investigation and can often result in conclusions being drawn on the basis of descriptive rather than analytical epidemiology [3].

The supermarket made the records of cooked meat delicatessen sales between 25 July and 13 August 2007 available to the outbreak control team. The use of some assumptions combined with the food histories for the eight confirmed primary cases allowed a novel study design. This aimed to test the hypothesis derived from descriptive epidemiology that cooked beef topside sold from a supermarket delicatessen was the most likely vehicle of infection [4].

METHODS

An outbreak case was defined as a resident of two postcode areas in the town of Paisley from whose stool *E. coli* O157 was isolated between 20 July and 31 August 2007. A case was primary if there was no documented contact with any other case in the week before the onset of symptoms. Only primary cases were included in this analytical study.

A 'case purchasing unit' (CPU) was defined as a group of people (e.g. a family or household) which included at least one primary case, and whose members therefore had the opportunity to eat cooked meat from the supermarket delicatessen. Non-CPUs were defined as those groups that had access to cooked meat from the implicated delicatessen between 31 July and 2 August but which did not include a primary case.

The period 31 July to 2 August (inclusive) corresponds to the range of dates that the implicated delicatessen produce was bought and was deemed the 'exposure period'. The case definition contains a longer time period as it was written during the hypothesis generation phase of the investigation.

The categories of cooked meat bought by CPUs were available from interviews conducted during investigation of the outbreak. The individual subtypes of beef topside, ham, turkey, chicken and lamb were grouped together into meat categories because the cases were unable to differentiate between them (e.g. the subtypes of beef topside on sale at the delicatessen were: 'Scottish', 'Best Aberdeen Angus', 'British' and 'Topside').

Since more than one individual can be exposed to a single purchase, it was necessary to use 'purchasing units' as a unit of analysis in this study rather than individual purchases.

We made a number of assumptions to define 'purchasing units' that bought cooked meat from the delicatessen between 31 July and 2 August (inclusive) and to ascertain whether or not they subsequently became 'CPUs' or 'non-CPUs'. These assumptions were:

- (1) Purchasing units only bought food from the delicatessen once between 31 July and 2 August (inclusive).
- (2) Non-CPUs bought the same number of items from the delicatessen as the mean of the CPUs.
- (3) All cooked meat sold by the delicatessen between 31 July and 2 August generated a sales record.
- (4) Non-CPUs did not have *E. coli* O157 infection.

The number of non-CPUs was calculated by subtracting the number of cooked meat sales to CPUs from the total number of cooked meat sales, and dividing this number by the mean number of items purchased by each CPU as per assumption (2) [see equation (1)]:

$$\text{no. of non-CPUs} = \frac{(\text{total no. of meat sales}) - (\text{no. of meat sales to CPUs})}{\text{mean number of items purchased in a CPU}} \quad (1)$$

The relative risk for each cooked meat subtype and category was then calculated using the formula:

$$\text{Relative risk (RR)} = (a/(a+b))/(c/(c+d)) \quad \text{with 95\% confidence intervals of } \text{RR}^{1 \pm (z/x)} \quad (2)$$

where a is the number of CPUs exposed to each cooked meat subtype or category; b is the number of non-CPUs exposed to each cooked meat subtype or category (this was estimated by subtracting the total number of known sales to the CPUs from the total number of sales during the exposure period, and then dividing this by the mean number of items purchased by the CPUs); c is the number of CPUs not exposed to each cooked meat subtype or category; d is the number of non-CPUs not exposed to each cooked meat subtype or category (this was calculated by subtracting b from the total number of non-CPUs); z is the value of the standard normal distribution (1.96); x is χ [calculated using the standard formula $\sqrt{\Sigma((\text{observed} - \text{expected})^2/\text{expected})}$] [5].

Table 1. Exposure history of case purchasing units (CPUs)

Primary case	CPU	CPU exposure						
		Beef silverside	Beef topside	Ham	Lamb	Pork lunch tongue	Roast pork	Turkey
1	A							
2	A		×					×
3	A							
4	B		×	×				
5	C		×	×		×	×	
6	D	×	×		×			
7	E		×		×			
8	F			×				
Total	6	1	5	3	2	1	1	1

Table 2. Relative risk (RR) and 95% confidence interval (CI) for each cooked meat category

	CPU exposed (a)	Non-CPU exposed (b)	CPU not exposed (c)	Non-CPU not exposed (d)	RR	95% CI
Beef silverside	1	28.5	5	385.9	2.7	0.3–24.0
Beef topside	5	27.8	1	386.7	59.2	7.9–617.2
Ham	3	367.5	3	46.9	0.1	0.0–0.7
Lamb	2	0.8	4	413.7	75.9	16.3–4676.4
Pork lunch tongue	1	22.5	5	391.9	3.4	0.4–31.1
Roast pork	1	21.8	5	392.7	3.5	0.4–32.3
Turkey	1	45.0	5	369.4	1.6	0.2–14.4

See equation (2) and following text for definition of columns (a), (b), (c), (d). CPU, Case purchasing unit.

RESULTS

The eight primary cases in the outbreak were part of six CPUs. The attack rate in these CPUs was 100%, and the mean number of individuals in each CPU was 1.3.

A total of 56 different cooked meat subtypes were on sale during the exposure period. These subtypes constituted 25 categories. Four subtypes were not sold to either cases or non-CPUs and only seven categories were bought by cases (beef topside, ham, lamb, turkey, beef silverside, pork lunch tongue, roast pork) (Table 1).

The total number of sales of cooked meat from the delicatessen over the 3 days was 981 of which 14 were to CPUs and 967 to non-CPUs. As the mean number of items purchased by the CPUs was 2.3, the number of non-CPUs was calculated as 414.4 (967/2.3). The relative risks for each meat category to which case-purchasing units were exposed are shown in Table 2. The other meat subtypes and categories are not shown as they do not yield meaningful

relative risks as they require division by zero. The only statistically significant relative risks (RR) are with exposure to beef topside [RR 59.2, 95% confidence interval (CI) 7.9–617.2] and lamb (RR 75.9, 95% CI 16.3–4676.4).

DISCUSSION

This new virtual cohort design supports the hypothesis that beef topside was the vehicle for an outbreak of *E. coli* O157 in Paisley, Scotland, in 2007. Beef topside and lamb are the only meat categories identified as having a statistically significant relative risk. All of the confidence intervals for the relative risks are wide because of the small number of CPUs in the study. The study supports the hypothesis that beef topside was the vehicle for the infection: the association of CPUs with the purchase of beef topside was highly statistically significant and it explains five of the six CPUs. Although purchasing lamb was also significantly associated with being a CPU and had a

higher relative risk, it only explains two of the six CPUs. Both of these CPUs were also exposed to beef topside, therefore the lamb hypothesis explains no more cases than the beef topside hypothesis alone. One CPU was exposed to neither of the meat categories significantly associated with becoming a CPU (having purchased only ham). This case can be explained either by an error in the exposure history, cross-contamination in the delicatessen (but in insufficient quantities to cause sufficient cases to result in statistical significance) or that it was actually a secondary case and the link to a primary case was missed.

Traditional analytical epidemiological techniques were not carried out in this outbreak because local publicity that the delicatessen may have been involved precluded the recruitment of blinded controls (for a case-control study). It was not possible to perform a cohort study because a full list of people who bought food from the delicatessen was not available. This situation is common in outbreak investigations. This design has not been published previously, and builds upon the more common use of sales data for syndromic surveillance [6]. Ethelberg *et al.* [7] performed a similar study using sales data derived from credit card transactions in Denmark to provide exposure information for an outbreak investigation. This is a more robust design than ours as it reduces the possibility of recall bias and did not require the same number of assumptions to generate a comparison group. However, this is not a design that could have been used for the Paisley outbreak because of the low use of electronic money in the community. Such an investigation is also likely to have required ethics approval in the Scottish context.

Each of the assumptions outlined in the methodology may have affected the results. If the first assumption (that all purchasing units – people or families – only bought cooked meat once during the exposure period) was wrong and there was a difference between the CPUS and non-CPUS the results could be biased. The second assumption (that the mean number of sales was the same for all purchasing units), could bias the study if there was a difference between CPUs and non-CPU units with more members were more likely to contain at least one case of *E. coli* O157. This could have contributed to an overestimation of the significance of the relative risks, although the potential for this bias to be a factor

during an outbreak investigation could be calculated using a sensitivity analysis of the number of individuals in a ‘purchasing unit’. The last assumption (that none of the non-CPU had *E. coli* O157 infection) may have led to an overestimation of relative risks. This potential bias is particularly relevant in our study because the analysis has been performed on the same cases that generated the hypothesis. This bias will have been minimized because the outbreak was widely publicized (thereby encouraging people to submit samples). This was reflected in a 40% rise in the number of requests for stool sample analysis in the local microbiology laboratory.

Our experience suggests that the use of sales data to estimate relative risks should be considered in future outbreak investigations where traditional cohort and case-control studies are not practicable.

DECLARATION OF INTEREST

None.

ACKNOWLEDGEMENTS

All the authors are funded by the NHS in Scotland. We thank Geoff Der for his statistical advice on the revised manuscript.

REFERENCES

1. Stirling A, *et al.* An outbreak of *Escherichia coli* O157 phage type 2 infection in Paisley, Scotland. *Eurosurveillance* 2007; **12**: 3253.
2. Anderson D. Woman dies in *E. coli* outbreak linked to Paisley superstores. *Evening Times*, 14 August 2007.
3. Chouinard E, Walter S. Recall bias in case-control studies: an empirical analysis and theoretical framework. *Journal of Clinical Epidemiology* 1995; **48**: 245–254.
4. Public Health Department. Report of the *E. coli* O157 outbreak in Paisley during August 2007. Glasgow, NHS Greater Glasgow and Clyde (Public Health Department), 2007.
5. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston: Little, Brown and Company, 1987, pp. 256.
6. Pavlin JA, Pavlin JA. Investigation of disease outbreaks detected by ‘syndromic’ surveillance systems. *Journal of Urban Health* 2003; **80**: i107–114.
7. Ethelberg S, *et al.* An outbreak of Verocytotoxin-producing *Escherichia coli* O26:H11 caused by beef sausage, Denmark 2007. *Eurosurveillance* 2007; **12**(22): pii = 3208 (<http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=3208>).