Selection of population controls for a *Salmonella* case-control study in the UK using a market research panel and web-survey provides time and resource savings

P. MOOK[1]*, S. KANAGARAJAH[1], H. MAGUIRE[1,2], G. K. ADAK[3],
G. DABRERA[4,5,6], A. WALDRAM[1,5,6], R. FREEMAN[4,5,6], A. CHARLETT[7] AND
I. OLIVER[1,8]

[1] *Field Epidemiology Service, Public Health England (PHE), UK*
[2] *Department of Infection and Population Health, University College London, London, UK*
[3] *Gastrointestinal, Emerging and Zoonotic Infections Department, PHE, London, UK*
[4] *Centre for Infectious Disease Surveillance and Control, PHE, London, UK*
[5] *Field Epidemiology Training Programme, PHE, UK*
[6] *European Programme for Intervention Epidemiology Training (EPIET), European Centre for Disease Prevention and Control, Stockholm, Sweden*
[7] *Statistics, Modelling and Economics Department, PHE, London, UK*
[8] *School of Social and Community Medicine, University of Bristol, Bristol, UK*

SUMMARY

Timely recruitment of population controls in infectious disease outbreak investigations is challenging. We evaluated the timeliness and cost of using a market research panel as a sampling frame for recruiting controls in a case-control study during an outbreak of *Salmonella* Mikawasima in the UK in 2013. We deployed a web-survey by email to targeted members of a market research panel (panel controls) in parallel to the outbreak control team interviewing randomly selected public health staff by telephone and completing paper-based questionnaires (staff controls). Recruitment and completion of exposure history web-surveys for panel controls ($n = 123$) took 14 h compared to 15 days for staff controls ($n = 82$). The average staff-time cost per questionnaire for staff controls was £13·13 compared to an invoiced cost of £3·60 per panel control. Differences in the distribution of some exposures existed between these control groups but case-control studies using each group found that illness was associated with consumption of chicken outside of the home and chicken from local butchers. Recruiting market research panel controls offers time and resource savings. More rapid investigations would enable more prompt implementation of control measures. We recommend that this method of recruiting controls is considered in future investigations and assessed further to better understand strengths and limitations.

Key words: Case-control studies, epidemiologic research design, outbreaks, *Salmonella*.

INTRODUCTION

The recruitment of population controls in a timely fashion is an established limitation of case-control studies. Controls are selected independently of exposure from the population from which cases arise, do not have the health event of interest and should be

* Author for correspondence: Dr P. Mook, Field Epidemiology Services South East and London, Public Health England, 151 Buckingham Palace Road, London, SW1W 9SZ.   (Email: piers.mook@phe.gov.uk)

eligible as cases if they were to develop the health event of interest. Established methods of control recruitment – including case- or physician-nominated controls and random or sequential digit dialling – can be time consuming, resulting in delays in identifying the source or vehicle of infection and might also introduce selection biases. Alternative methods that have been proposed include case-crossover [1], case-case [2] and case-cohort [3, 4] study designs. As part of a review undertaken by Public Health England (PHE) to strengthen investigation of national outbreaks of infectious disease we identified the need to evaluate the potential use of commercial market research panels as a source of controls for epidemiological studies as part of outbreak investigations.

Commercial market research panels are pre-recruited groups of individuals who have opted in to receive online questionnaires in return for a reward. Review of the literature indicates that the use of market research panels as a sampling frame for prospectively identifying controls and collecting data using a web-survey has not yet been explored. Commercial market research databases that hold residential telephone numbers have previously been used as a sampling frame for recruiting controls by phone for epidemiological studies [5–7]. Pre-collected market research data have also been used as control data in a case-control study [8] and as denominator data [9]. However, the use of a market research panel for recruiting controls and collecting data with a web-survey has the potential for time and cost savings compared to traditional methods.

We evaluated the timeliness and cost of using a market research panel as a sampling frame for recruiting controls and capturing data against a control recruitment strategy employed by an outbreak control team (OCT) as part of a case-control study conducted in response to a national outbreak of *Salmonella* Mikawasima in the UK in 2013. The distribution of exposures among the two control groups and analytical findings of the case-control studies using each control group were also compared.

METHODS

Of those cases of *S.* Mikawasima identified as part of the outbreak, 61 were eligible for inclusion in a national case-control study. The OCT agreed that two controls per case, frequency-matched on sex and area of residence and restricted to those aged between 18 and 65 years, were required ($n = 122$) [10].

Recruitment of controls from public health staff

The OCT (including authors G.K.A., G.D., R.F., A. C.) recruited controls using public health staff as a sampling frame (henceforth referred to as 'staff controls'). Eligible controls were recruited from randomly sampled public health employees based in regions which corresponded to the residence of cases using random number generation in conjunction with lists of staff. These controls were interviewed by telephone by four investigating teams using a paper-based questionnaire. Double data entry from the paper questionnaire into a bespoke database was conducted to identify and address transcription errors prior to analyses; one team conducted all double data entry on behalf the four investigating teams.

Recruitment of controls from the market research panel

In parallel, the evaluation team (authors P.M., S.K., H.M., I.O.) recruited a second control population from a market research panel (henceforth referred to as 'panel controls'). The evaluation team had no influence on the choice of controls by the OCT. The control definition was identical to that defined by the OCT with the substitution of public health staff with a market research panel member. The recruitment target, age restriction and the target gender and area of residence strata also mirrored that of staff controls.

The evaluation team developed a web-survey version of the staff control paper-based questionnaire using the web-survey software Select Survey (ClassApps, USA) and created a copy for each frequency-matching strata (i.e. each gender and geographical area combination). The web-survey, required number of responses and age, gender and geographical distribution of our target population was shared with market research company A, which maintains a volunteer panel with more than 600 000 respondents in the UK. A quote for the service of forwarding the web-survey link by email to a selection of panel members who met our criteria and rewarding those who participated was provided within 1 day. This quote was based on estimated time to complete the web-survey and the size and demographic distribution of the target population. The estimated time to complete the web-survey determines the point value received by participating panel members and collected points can be exchanged for rewards. Two links provided by company A were included on the completion and screen out pages so they might monitor completeness and screen out rates.

Company A stratified panel members into homogenous, mutually exclusive subgroups – based on frequency-matching criteria – and sent survey invitations by email to randomly selected panel members within each defined strata. The number of panel members randomly selected and to whom emails were sent was greater than the required number of responses and was determined by the number of complete responses required, the anticipated response rate of the panel (or strata) and expected proportion of panel members who meet the inclusion criteria (85% for this survey). Web-survey data from recruited individuals were captured via Select Survey and strata-specific web-surveys closed once strata-specific targets were achieved. It was the policy of company A that name, telephone number and postcode (items to be collected by telephone interview of staff controls) could not be collected from panel controls using the web-survey.

Comparison of exposures between control groups

The distribution of age, gender and place of residence among panel controls, staff controls and cases were summarized and compared. A 'control-control' study employing univariable analysis and multivariable logistic regression was conducted to compare the distribution of exposures among panel controls against that of staff controls. Odds ratios, 95% confidence intervals and P values were estimated.

Timeliness of control recruitment

The time required to recruit and capture data for panel and staff controls was summarized and compared. For panel controls, this period was calculated from the point at which company A launched the web-survey until the total required number of controls was obtained. For staff controls this period was calculated from when recruitment interviews by each investigating team began until the regional target for controls was obtained. The time required to develop the web-survey and paper-based questionnaire for panel and staff controls, respectively, was not considered.

Efficiency of control recruitment

Percentage of target panel controls recruited and details of how many panel members declined to participate by not opening the web-survey, opened but did not complete the web-survey or were not eligible were captured and summarized. These metrics were compared against percentage of target staff controls recruited and details of how many public health staff could not be contacted, declined to be interviewed or were not eligible.

Resource required for control recruitment

The direct resource required to recruit panel and staff controls was summarized and compared as an average cost per control recruited. The resource required to recruit panel controls was based on procurement costs for the services provided by company A. The resource required to recruit staff controls was estimated using information captured on a recruitment log: time per call to potential staff controls, outcome of the attempted recruitment and pay band of the staff member recruiting staff controls (assumed minimal level of progression into any given pay band unless otherwise stated). Additional cost of entering all hard copy questionnaires onto a bespoke database was also estimated.

Case-control analyses using each control group

Univariable analysis and multivariable logistic regression was conducted using the case data collected by the OCT and panel controls as the comparator to investigate the effect of exposures on outcome. Odds ratios, 95% confidence intervals and P values were estimated. The findings of these analyses were compared against those derived using staff controls by the OCT. The approach to refining models and other decision making in the analysis using panel controls was harmonized with that used by the OCT for analyses using staff controls [10] and was validated by a third party to the evaluation team and the OCT (author A.W.). Analyses and data manipulation were performed using Stata v. 12 (Stata Corporation, USA), OpenEpi (http://www.openepi.com) and Excel (Microsoft, USA).

RESULTS

Response rates

Of 61 cases initially identified as eligible for inclusion in the study as per a case definition developed by the OCT, 21 (34%) were lost to follow-up or could not be followed up due to limited resources in some investigating teams. In addition, on further review one case did not meet the inclusion criteria [10]. The number of responses from panel controls ($n = 123$) surpassed

Table 1. *Distribution of demographic characteristics among cases, panel controls and staff controls*

Variable	Cases (N = 39) n (%)	Control targets* (N = 122) n (%)	Staff controls (N = 82) n (%)	Panel controls (N = 123) n (%)	P value†
Gender					
Male	17 (43·6)	60 (49·2)	35 (42·7)	60 (48·8)	0·658
Female	22 (56·4)	62 (50·8)	47 (57·3)	63 (51·2)	
Investigating region					
Scotland	9 (23·1)	48 (39·3)	18 (22·0)	48 (39·0)	0·218
North East	13 (33·3)	32 (26·2)	27 (32·9)	32 (26·0)	
West Midlands	9 (23·1)	22 (18·0)	21 (25·6)	22 (17·9)	
London	8 (20·5)	20 (16·4)	16 (19·5)	21 (17·1)	
Age (years)					
18–25	6 (15·4)	–	2 (2·4)	5 (4·1)	0·054
26–35	8 (20·5)	–	13 (15·6)	19 (15·5)	
36–45	6 (15·4)	–	23 (28·1)	23 (18·7)	
46–55	9 (23·1)	–	28 (34·2)	40 (32·5)	
56–65	10 (25·6)	–	16 (19·5)	36 (29·3)	
Ethnicity					
White	27 (77·1)	–	72 (87·8)	115 (95·0)	0·006
Other	8 (22·9)	–	10 (12·2)	6 (5·0)	
Not reported	4 (–)	–	0 (–)	2 (–)	

* Target only set for frequency-matching criteria (region and gender).
† χ^2 test, cases *vs.* staff controls *vs.* panel controls.

the initial target (*n* = 122) and met the target for each gender and investigating region strata. The number of responses from staff controls (*n* = 82) was 67% of the initial target.

Of 1329 market research panel members sent a link, 123 (9%) completed the web-survey, 262 (20%) started but did not complete, 20 (2%) were screened out based on eligibility criteria [10] and 924 (70%) did not follow the link while the web-survey was live. Where data on outcome of recruitment attempts were available for staff controls (*n* = 69), 64% of staff for whom an attempt to contact was made were ultimately interviewed, 26% could not be contacted, 9% did not meet the eligibility criteria and 1% declined.

Characteristics of control groups

The distribution of demographic characteristics among cases and both control sets were similar for gender, investigating region and age group but there was evidence of a difference in ethnicity ($P = 0.006$, Table 1). The proportion of staff controls collected by the Scotland investigating team compared to elsewhere was less than that for panel controls (Fisher's exact test, $P = 0.010$). The proportion of both staff controls and panel controls in the 18–25 years age group compared to other ages was less than that for

cases (Fisher's exact test, $P = 0.027$ and $P = 0.048$, respectively). In addition, the proportion of panel controls that were white was greater than that for cases (Fisher's exact test, $P = 0.007$).

The distribution of demographic characteristics (ethnicity, gender, age and investigating region) was similar between control sets (see Supplementary Appendix). When the distribution of clinical, travel and recreational, food consumption and food purchasing exposures were compared between control sets, panel controls more often reported having a headache, receipt of proton pump inhibitors (PPIs), eating pomegranate, chicken sandwiches, chicken pasties, chicken pies and chicken nuggets and eating in Chinese restaurants than staff controls (differences at the 5% level presented in Table 2; full details in the Supplementary Appendix). Compared to panel controls, staff controls more often reported recent travel in the UK, any outdoor activity, walks, other outdoor activities, using commercial bird food, buying raw chicken, eating or handling pumpkins, eating whole chicken or chicken portions, salad garnish, side dishes, herbs/spices, apples, dips, turmeric, cumin seed, mint, coriander, buying chicken from three national supermarket chains, eating out and eating in sandwich bars, fast-food restaurants, British restaurants or other locations (Table 2).

Table 2. *Comparison of distribution of exposures among staff controls and panel controls (P < 0·05)*

Exposure	Staff controls			Panel controls			OR (95% CI)	P value
	Total	Exposed	%	Total	Exposed	%		
Symptoms and medication								
Headache	82	10	12·2	123	49	39·8	0·2 (0·1–0·5)	<0·001
Proton pump inhibitors	81	4	4·9	123	21	17·1	0·3 (0·1–0·8)	0·010
Travel and recreation								
Commercial bird food*	19	14	73·7	105	12	11·4	21·7 (5·9–87·7)	<0·001
Other outdoor area	81	43	53·1	123	30	24·4	3·5 (1·8–6·7)	<0·001
Travel in UK	82	22	26·8	123	8	6·5	5·3 (2·1–14·4)	<0·001
Walks	81	60	74·1	123	59	48·0	3·1 (1·6–6·0)	<0·001
Any outdoor activity	82	60	73·2	123	64	52·0	2·5 (1·3–4·8)	0·002
Chicken items								
Buy raw chicken	82	62	75·6	123	58	47·2	3·5 (1·8–6·8)	<0·001
Whole chicken or chicken pieces	82	40	48·8	123	35	28·5	2·4 (1·3–4·5)	0·003
Sandwiches	82	7	8·5	123	29	23·6	0·3 (0·1–0·8)	0·006
Pasties	82	1	1·2	123	11	8·9	0·1 (0·0–0·9)	0·021
Pies	82	4	4·9	123	18	14·6	0·3 (0·1–0·96)	0·027
Nuggets	82	1	1·2	123	10	8·1	0·1 (0·0–1·02)	0·031
Non-chicken food items								
Pumpkins	82	31	37·8	123	12	9·8	5·6 (2·5–12·9)	<0·001
Salad garnishes	78	12	15·4	123	2	1·6	11·0 (2·3–102·0)	<0·001
Side dishes	78	15	19·2	123	5	4·1	5·6 (1·8–20·5)	<0·001
Herbs/spices	82	17	20·7	123	7	5·7	4·3 (1·6–12·9)	0·001
Apples	82	52	63·4	123	50	40·7	2·5 (1·4–4·7)	0·001
Dips	78	8	10·3	123	1	0·8	13·9 (1·8–623·0)	0·002
Turmeric	81	11	13·6	123	4	3·3	4·7 (1·3–20·7)	0·006
Cumin seeds	81	8	9·9	123	2	1·6	6·6 (1·3–65·2)	0·008
Mint	81	7	8·6	123	2	1·6	5·7 (1·0–57·4)	0·017
Coriander	81	6	7·4	123	2	1·6	4·8 (0·8–49·9)	0·037
Pomegranate	82	1	1·2	123	9	7·3	0·2 (0·00–1·2)	0·047
Location of chicken purchase								
Supermarket chain G	81	18	22·2	123	13	10·6	2·4 (1·04–5·7)	0·023
Supermarket chain F	81	17	21·0	123	12	9·8	2·5 (1·03–6·0)	0·025
Supermarket chain H	81	3	3·7	123	0	0·0	∞ (1·2-∞)	0·032
Eating out venues								
Sandwich bar	82	14	17·1	123	3	2·4	8·2 (2·2–45·8)	<0·001
Other eating out	81	13	16·1	123	3	2·4	7·6 (2·0–42·9)	<0·001
Fast food	82	31	37·8	123	22	17·9	2·8 (1·4–5·6)	0·001
Any eating out	82	41	50·0	123	39	31·7	2·2 (1·2–4·0)	0·009
British	82	11	13·4	123	5	4·1	3·7 (1·1–13·9)	0·014
Chinese	82	2	2·4	123	12	9·8	0·2 (0·0–1·1)	0·042

OR, Odds ratio; CI, confidence interval.
* In the panel control web-survey the question on commercial bird food was not conditional on feeding birds.

In total, there were differences at the 5% level for 29% of the 115 exposure and demographic items compared between control sets. Multivariable analysis identified eight exposures (7% of all items compared) with independent associations with either set of controls. Buying raw chicken, recent travel in the UK, eating or handling pumpkins, eating apples and participating in outdoor activities were positively associated with staff controls. Experiencing a headache in the last 7 days, receiving PPIs and eating any other chicken products were positively associated with being a panel control (Table 3).

Timeliness of control recruitment

Panel controls (n = 123) were recruited within 14 h of deployment of emails (Fig. 1) and staff controls (n = 82) were recruited over a 15-day period

Table 3. *Final multivariable logistic regression model showing significant associations with either set of controls*

Variable	aOR (95% CI)	*P* value	Control set with greater odds
Age*	1·0 (0·9–1·0)	0·016	–
Female*	1·4 (0·7–3·1)	0·363	–
Buy raw chicken	5·7 (2·5–13·1)	<0·001	Staff
Headache	0·2 (0·1–0·5)	0·001	Panel
Travel in UK	7·5 (2·4–23·2)	0·001	Staff
Pumpkins	4·2 (1·6–10·8)	0·003	Staff
Proton pump inhibitors	0·2 (0·0–0·6)	0·007	Panel
Apples	2·7 (1·3–5·9)	0·009	Staff
Outdoor activities	2·4 (1·1–5·5)	0·035	Staff
Any other chicken product	0·5 (0·2–1·0)	0·048	Panel

aOR, Adjusted odds ratio; CI, confidence interval.
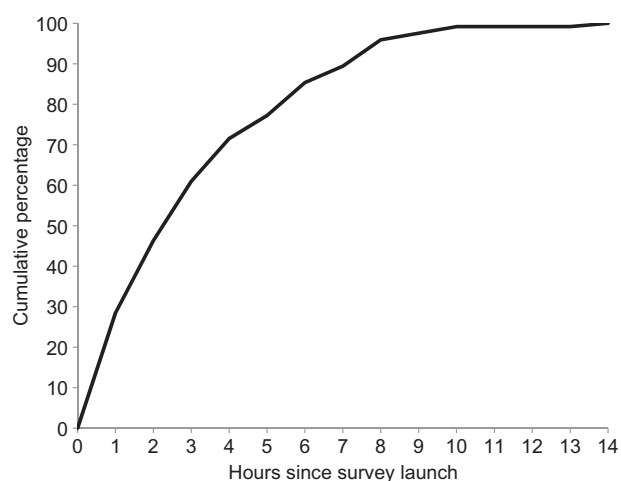* Retained as *a priori* confounders.

Fig. 1. Cumulative percentage of completed responses by panel controls in hours since web-survey launch (09:00 hours, 23 December 2013) (*n* = 123).

(Table 4) – this included the date of panel control recruitment. Double data entry of staff controls and case paper-based questionnaires onto a database took 76·5 h and was conducted intermittently over a 37-day period, after which staff control data were available for analysis (Table 4). Transfer of paper-based questionnaires from other investigating teams to the investigating team responsible for double data entry was a rate-limiting step.

Resource required for control recruitment

Where data on time required to recruit and interview staff controls was available (*n* = 56, 68% of staff

controls recruited), the average staff cost per collected questionnaire was estimated to be £3·88. These data were available for three of the four investigating teams; the total number of staff involved in follow-up in these three teams was 13 (range 2–6) and data was available for 11 of these staff members. The average staff cost to double-enter case or staff control questionnaires was £9·25; the total cost to recruit, interview and enter data for each staff control was £13·13. The average direct, invoiced cost per completed web-survey for panel controls was £3·60 (Table 5).

Findings of case-control studies

The findings of the OCT analyses, which used staff controls, indicated that receipt of PPIs, having visited a fast-food outlet regardless of chicken consumption history, having visited a restaurant and eaten chicken and having bought chicken from a local butcher were factors independently associated with illness [10] (Table 6*a*). Analyses using panel controls, indicated that having visited a fast-food outlet regardless of chicken consumption history, visited a restaurant and eaten chicken, bought chicken from a local butcher and consumed coriander or garnishes were factors independently associated with illness (Table 6*b*). When the model derived using staff controls was applied to the analytical study using panel control data, those exposures independently associated with illness remained so with the exception of receipt of PPIs, which was not selected to be included in the panel control model (Table 6*c*).

DISCUSSION

We have demonstrated that the use of market research panel members as a sampling frame for recruiting population controls and data collection through a web-survey offers value for money and substantial time savings compared to telephone interviews of public health staff. At £3·60, the average cost per control using market research panel members was over three times cheaper than the estimated average staff cost to interview and double-enter questionnaire data per control using public health staff (£13·13). The initial target of frequency-matched controls was surpassed using panel controls with 123 controls collected in 14 h at which point data were available to be cleaned and analysed. Eighty-two staff controls were recruited in 15 days using staff controls and subsequent double data entry was required before data were available to

Table 4. *Time taken to recruit, collect data and double enter data, as necessary, for staff and panel controls*

Control type	No. of control questionnaires	Date started	Date finished	Duration (days)
Staff controls				
Recruit and interview controls by phone	82	9 Dec. 2013	24 Dec. 2013	15
Double data entry*	126	23 Dec. 2013	29 Jan. 2014	37
Panel controls				
Recruit and collect data through a web-survey	123	23 Dec. 2013	23 Dec. 2013	1

* Double data entry of cases and staff controls. On review of the data one case and four staff controls met the exclusion criteria; the remaining 121 questionnaires comprised of 39 cases and 82 staff controls.

Table 5. *Direct staff costs required to recruit, collect data and double enter data, as necessary, for staff and panel controls*

Control type	Cost (£)	No. of controls	Average cost per control (£)
Staff controls			
Recruitment and interview of controls by phone*	217·61	56	3·88
Double data entry†	1165·01	126	9·25
Total recruitment and entry cost	n.a.	n.a.	13·13
Panel controls			
Recruitment and data collection through web-survey	442·8	123	3·60

n.a., Not available.
* Based on available staff cost data.
† Double data entry of cases and staff controls; five controls were entered which were later excluded.

be cleaned and analysed. While there were differences in the distribution of exposures in the two control groups, both analytical studies found that illness was associated with consumption of chicken outside of the home, and purchasing chicken from local butchers. There was no microbiological evidence to support or refute these findings.

An ideal evaluation of a novel approach for recruiting controls and collecting data would be made against best current practice but all existing methods of control selection have limitations. In this instance, we have compared the use of a market research panel as a sampling frame for recruiting controls against a quasi-novel method using public health staff – a mechanism that has been used once before (PHE, unpublished data). It is not possible to assess which control set is most representative of the at-risk population but it is likely that both have inherent biases. It is feasible that in addition to a 'healthy worker' effect, public health staff might have healthier 'lifestyle' behaviours and more disposable income than the general population. Members of the public who volunteer to participate as a panel member might have systematic differences to those that do

not, e.g. access to the internet, age, gender or lifestyle behaviours, and previous studies would suggest that volunteer panels do provide biased estimates [11, 12]. It is therefore difficult to interpret the findings of the comparison of the distribution of exposures between recruited public health staff and market research panel members reported here. Testing panel controls against other methods of control section may help understand any inherent biases with such controls.

Systematic differences between the distribution of demographic characteristics among the panel and that of the general population should be addressed using frequency matching where possible (e.g. for age and sex) to ensure that data are collected from controls that most closely resemble the at-risk population. Age of the panel member is known to affect promptness to respond and therefore might impact on the age distribution of responders if age strata targets are not used. Company A reported that younger age groups in its panel are slower to respond to surveys than older age groups and this may have resulted in the difference in age distributions observed between panel controls and cases even though the overall age distribution of panel

Table 6. *Final multivariable logistic regression models*

Exposure	aOR (95% CI)	P value
(a) With staff controls (observations in model = 118)		
Age*	1·0 (0·9–1·0)	0·310
Female*	1·7 (0·6–4·6)	0·301
Proton pump inhibitors	8·8 (2·0–38·3)	0·004
Visit to fast-food restaurant (*vs.* no visit)		
Visit to fast-food restaurant, no chicken	4·3 (1·4–13·4)	0·012
Visit to fast-food restaurant, ate chicken	5·0 (1·2–21·2)	0·028
Visit to restaurant (*vs.* no visit)		
Visit to restaurant, no chicken	1·2 (0·3–4·6)	0·817
Visit to restaurant, ate chicken	10·3 (2·6–40·4)	0·001
Bought chicken from local butcher	7·2 (1·5–34·4)	0·013
(b) With panel controls (observations in model = 160)		
Age*	1·0 (1·0–1·1)	0·477
Female*	4·1 (1·3–12·8)	0·017
Visit to fast-food restaurant (*vs.* no visit)		
Visit to fast-food restaurant, no chicken	21·4 (5·5–83·7)	<0·001
Visit to fast-food restaurant, ate chicken	10·7 (2·1–56·2)	0·005
Visit to restaurant (*vs.* no visit)		
Visit to restaurant, no chicken	1·3 (0·3–6·1)	0·724
Visit to restaurant, ate chicken	11·3 (2·4–52·8)	0·002
Bought chicken from local butcher	28·9 (3·9–217·4)	0·001
Coriander	13·9 (1·8–107·7)	0·012
Salad garnishes	10·6 (1·5–76·5)	0·019
(c) Staff control model applied to panel control data (observations in model = 160)		
Age*	1·0 (0·9–1·0)	0·593
Female*	3·3 (1·2–9·0)	0·023
Proton pump inhibitors	2·0 (0·6–6·4)	0·263
Visit to fast-food restaurant (*vs.* no visit)		
Visit to fast-food restaurant, no chicken	20·8 (5·8–74·1)	<0·001
Visit to fast-food restaurant, ate chicken	12·9 (3·0–55·7)	0·001
Visit to restaurant (*vs.* no visit)		
Visit to restaurant, no chicken	1·1 (0·3–4·3)	0·898
Visit to restaurant, ate chicken	10·2 (2·5–41·7)	0·001
Bought chicken from local butcher	45·2 (5·5–369·0)	<0·001

aOR, Adjusted odds ratio; CI, confidence interval.
* Retained in model as *a priori* confounder.

members is younger than that for the general population (personal communication with company A). In addition, time of year might also affect how quickly panel members respond to a survey.

The process by which panel controls were recruited will, by design, lead to only a small proportion of those that receive an invitation to participate being recruited. Approximately 9% of panel controls invited to participate were recruited. A low response rate has been reported elsewhere for volunteer panels [12, 13], but more panel members might have responded had the survey remained open once the target was achieved. To better assess ascertainment bias in future studies using panel controls, a comparison of the breakdown by demographic characteristics of those that did respond with those that did not (at least while the survey was open) is recommended and these data should be available from market research companies. We were unable to conduct this assessment due to 12 survey completers being misclassified as non-completers having not selected the completion link provided by company A. However, we were able to address this for the totals presented in the results by reviewing the collected data.

We observed a difference in the ethnicity distribution between cases and panel controls with the latter containing fewer individuals from a non-white ethnic group. Such a distribution has been observed previously in volunteer panels compared to the general population [11] but it cannot be addressed by

frequency matching on ethnicity because this is not a data item collected when a panel member first volunteers. Ethnicity was not found to improve the fit of the presented multivariable models or be independently associated with outcome in this study.

The inherent differences in the distribution of exposures between control sets resulted in differences in the final multivariable models using each control set. PPIs (5% exposure in staff controls *vs.* 17% exposure in panel controls) were independently associated with illness only when staff controls were used, most likely as a predisposing factor [14–16] rather than a causal exposure, while coriander (7% *vs.* 2%) and salad garnishes (15% *vs.* 2%) were independently associated with illness only when panel controls were used. However, it is not possible to assess which control set provides the most accurate measures of association. It is possible that for a given exposure biases exist in opposite directions for the two control sets and that the distributions of these exposures in the population from which cases arose might be between that for the two control sets. In addition, the ultimate interpretation of findings from any outbreak analysis would take into account not just the significance test but also the size of the effect, the proportion of cases that can be accounted for by a given exposure and biological plausibility.

Both investigations identified the consumption of chicken outside of the home and the purchase of chicken from local butchers as being independently associated with illness. This demonstrates that the two control sets were similar enough to identify the same risk exposures when there is a strong causal association as appears to be the situation here.

There was incomplete data on timeliness and resource required for staff controls. We have assumed that there was no systematic difference in the time and resource required to recruit staff controls where data were not available but ascertainment bias could not be assessed. In addition, the opportunity cost for those staff recruiting and interviewing staff controls and the staff controls themselves while being interviewed was not estimated and therefore the total cost to PHE might be underestimated.

The exclusion of cases resulted in a change of gender and regional distributions which staff controls mirrored, maintaining a 2:1 ratio of controls to cases. Panel controls ultimately had more than a 3:1 ratio because they were recruited on the presumption of no case drop-out and ensured frequency matching as per the protocol. This increased ratio might have

resulted in additional power to detect effects. As a result of the revised target, the investigating teams needed to only collect two thirds of the number of staff controls compared to panel controls which makes the difference in timeliness more apparent. The pricing policy of company A for panel controls was dependent on the required number of completed surveys, subject and the expected exclusion rate. Had the target number and geography been revised as per that for staff controls, the overall cost for collecting panel control data would have been less.

As with all outbreak investigations, the scale and severity of the outbreak and competing priorities probably influenced availability of staff resource which in turn might have impacted on the timeliness of recruiting staff controls and dispatch of questionnaires to the investigating team responsible for double data entry. The Christmas break might also have had some impact on how quickly staff control questionnaires were sent to the investigating team responsible for double data entry and emphasizes the impact of contextual considerations on any outbreak investigation. Similarly, deployment of the web-survey to panel members on 23 December might have had some effect on recruitment of panel controls.

Public health staff are likely to be more available during working hours and, as a more engaged population, more likely to respond to such a survey than members of the general public. Recruiting staff controls was therefore likely to represent the 'best case' in terms of accessibility, compliance and cost compared to other methods of population control recruitment through telephone interviews, e.g. random digit dialling. As a result, the use of panel controls might be even more cost-effective and timely when compared to other such methods. Panel controls receive a real time payment for a response and are therefore likely to be more compliant than the general public.

With the exception of name, telephone number and postcode, all questions in the paper-based questionnaire used to interview staff controls were replicated in the web-survey. It was important that those designing the web-survey were clear how questions were applied to staff controls and the two data collection instruments were as harmonized as possible. Responses to most questions in the web-survey were mandatory unless they were conditional on a previous response, ensuring high levels of data completeness.

There was a systematic difference in the mechanism of data collection for cases and staff controls

(telephone interview) *vs.* that for panel controls (self-reporting by web-survey). Telephone interviews allow for further explanation of questions where the respondent is unclear or has misunderstood and is likely to improve quality of responses. This might have resulted in the introduction of information bias, specifically differential misclassification between panel controls and staff controls or cases. In future cases should ideally be interviewed by the same mechanism, i.e. by web-survey, to minimize potential differential misclassification.

A more direct comparison of the speed and resources required to recruit panel controls might have been made against contacting public health staff or members of the public by phone (e.g. by random digit dialling) and requesting an email address to which an online survey could be sent. This would have helped to assess whether the shorter times for recruitment and data collection for panel controls was as a result of recruiting from a market research panel or the use of a web-survey.

Company A audit data provided by panel members and remove individuals as necessary to maintain a high-quality panel. This should minimize instances where, given the incentive to respond, panel members provide quick but low-quality responses to surveys. They assert that given the size of the panel, panel members do not suffer from survey fatigue and this also benefits data quality. As per market research industry standards in confidentiality, full postcode details were not available and consequently it was not possible to determine an index of multiple deprivation (IMD) score for each recruited panel member although it might have been useful to consider as an independent variable in the analyses. Alternative suppliers of panels have indicated that an IMD score could be provided for each completed web-survey respondent if they were provided with an appropriate postcode to IMD score look-up table and the email address with which the respondent signed up to the panel was captured on the web-survey.

The use of a market research company to recruit controls for a case-control study using random digit dialling as part of the response to an outbreak of hepatitis A in Ireland was described as effective and efficient and required on average 1 h of calls per recruited control ($n = 42$), although the process was reported to have been delayed as a result of having to match to county [17]. The cost of using a commercial database of telephone numbers as a sampling frame to recruit controls by phone interview for a Lyme disease case-control study in the United States was not insignificant and the median time required to reach the person who ultimately completed the study questionnaire was approximately 19 min ($n = 353$) [5]. In both instances, these were lengthier processes per recruited control than the method we present using panel controls and a web-survey. However, it is difficult to fairly compare time and resource required with our findings as previous prospective studies using market research companies had different inclusion criteria, survey lengths and controls targets as well as differences in recruitment approach. Using retrospective market research panel data in a listeriosis case-control study in England was not considered to be inexpensive, was limited to those items that had been pre-collected and could not examine confounding because only grouped data were available (individual-level data was prohibitively expensive) [8].

A review of published outbreaks of a similar impact in Europe between 2003 and 2013 and for which a case-control study was conducted determined a median time interval of 16 days between hypothesis generation and availability of analytical results and a range of 7–59 days ($n = 16$) [18]. Control recruitment and data collection within a matter of a few days if not within a day using panel controls might contribute to reducing this interval in the future. The potential for substantial time savings has important consequences for public health benefit as more rapid investigations would enable more prompt implementation of control measures and might mitigate the burden of disease associated with an outbreak. Applying this method does not restrict subsequent use of a traditional recruitment mechanism if a clear answer is not apparent, or the exposure distribution of the panel controls appears unusual.

This approach could be used to recruit controls from further refined geographical granularity – at least to local authority level (personal communication with company A). Such panels are available in many high- and middle-income countries and by approaching companies that hold or collaborate with panels elsewhere this approach could be used to recruit population controls as part of an international outbreak investigation. Data are not directly available from such panels for individuals aged <18 years but consenting panel members might report on their children's exposures if an outbreak included children or consent for their child to complete a survey under their supervision. Rather than using self-selected volunteers, some

panels in the United States and Europe have employed probability-based sampling methods, which might address any sampling bias [11].

While all methods of recruiting population controls are likely to have some inherent bias, the additional benefits of this method, particularly in terms of timeliness, make this a potentially valuable addition to the range of existing methods for use in outbreak investigations where self-reporting via a web-survey is acceptable. Panel controls should be considered as a possible method of control selection in future outbreaks even though further evaluation work, particularly against a range of control selection methods, would be useful to better understand the value and limitations of this approach.

SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0950268815002290.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Maclure M.** The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* 1991; **133**: 144–153.
2. **McCarthy N, Giesecke J.** Case-case comparisons to study causation of common infectious diseases. *International Journal of Epidemiology* 1999; **28**: 764–768.
3. **Prentice R.** A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**: 1–11.
4. **Le Polain de Waroux O, Maguire H, Moren A.** The case-cohort design in outbreak investigations. *Eurosurveillance* 2012; **17**.
5. **Connally NP, Yousey-Hindes K, Meek J.** Selection of neighborhood controls for a population-based Lyme disease case-control study by using a commercial marketing database. *American Journal of Epidemiology* 2013; **178**: 276–279.
6. **Olson SH, Mignone L, Harlap S.** Selection of control groups by using a commercial database and random digit dialing. *American Journal of Epidemiology* 2000; **152**: 585–592.
7. **Valery PC, et al.** Electronic telephone directory listings: preferred sampling frame for a population-based study of childhood cancer in Australia. *Annals of Epidemiology* 2000; **10**: 504–508.
8. **Gillespie IA, et al.** *Listeria monocytogenes* infection in the over-60s in England between 2005 and 2008: a retrospective case-control study utilizing market research panel data. *Foodborne Pathogens and Disease* 2010; **7**: 1373–1379.
9. **Gillespie IA, et al.** Human listeriosis in England, 2001–2007: association with neighbourhood deprivation. *Eurourveillance* 2010; **15**: 7–16.
10. **Freeman R, et al.** Association between use of proton pump inhibitors and non-typhoidal salmonellosis identified following investigation into an outbreak of *Salmonella* Mikawasima in the UK, 2013. *Epidemiology and Infection* (in press).
11. **Erens B, et al.** Nonprobability web surveys to measure sexual behaviors and attitudes in the general population: a comparison with a probability sample interview survey. *Journal of Medical Internet Research* 2014; **16**:e276.
12. **AAPOR.** Report on online panels, 2010 (http://www.webcitation.org/6Nnji0q7q).
13. **Craig BM, et al.** Comparison of US panel vendors for online surveys. *Journal of Medical Internet Research* 2013; **15**:e260.
14. **Bavishi C, DuPont HL.** Systematic review: the use of proton pump inhibitors and increased susceptibility to enteric infection. *Alimentary Pharmacology & Therapeutics* 2011; **34**:1269–1281.
15. **Wu H-H, et al.** Association between recent use of proton pump inhibitors and nontyphoid salmonellosis: a nested case-control study. *Clinical Infectious Diseases* 2014; **59**:1554–1558.
16. **Bowen A, et al.** Role of acid-suppressing medications during a sustained outbreak of *Salmonella enteritidis* infection in a long-term care facility. *Infection Control & Hospital Epidemiology* 2007; **28**:1202–1205.
17. **Fitzgerald M, et al.** Outbreak of hepatitis A infection associated with the consumption of frozen berries, Ireland, 2013 – linked to an international outbreak. *Eurourveillance* 2014; **19**.
18. **van de Venter E, Oliver I, Stuart JM.** Timeliness of epidemiological outbreak investigations in peer-reviewed European publications, January 2003 to August 2013. *Eurosurveillance* 2015; **20**.