

Online market research panel members as controls in case–control studies to investigate gastrointestinal disease outbreaks: early experiences and lessons learnt from the UK

Original Paper

Cite this article: Mook P *et al* (2018). Online market research panel members as controls in case–control studies to investigate gastrointestinal disease outbreaks: early experiences and lessons learnt from the UK. *Epidemiology and Infection* **146**, 458–464. <https://doi.org/10.1017/S0950268817002953>

Received: 21 June 2017

Accepted: 28 November 2017

First published online: 15 January 2018

Key words:

Case–control studies; epidemiologic research design; gastrointestinal infections; outbreaks

Author for correspondence: Piers Mook,
E-mail: piers.mook@phe.gov.uk

P. Mook^{1,7}, J. McCormick³, S. Kanagarajah³, G. K. Adak^{3,4}, P. Cleary^{1,4}, R. Elson^{3,4}, M. Gobin¹, J. Hawker^{1,3,4}, T. Inns^{1,4}, C. Sinclair^{1,5,6}, S. C. M Trienekens^{1,5,6}, R. Vivancos^{1,4,7} and N.D. McCarthy^{2,4}

¹Field Epidemiology Service, National Infection Service, Public Health England (PHE), London, UK; ²Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK; ³Gastrointestinal Infections, National Infection Service, PHE, London, UK; ⁴NIHR Health Protection Research Unit in Gastrointestinal Infections, UK; ⁵Field Epidemiology Training Programme, PHE, London, UK; ⁶European Programme for Intervention Epidemiology Training, European Centre for Disease Prevention and Control, Stockholm, Sweden and ⁷NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, UK

Abstract

Established methods of recruiting population controls for case–control studies to investigate gastrointestinal disease outbreaks can be time consuming, resulting in delays in identifying the source or vehicle of infection. After an initial evaluation of using online market research panel members as controls in a case–control study to investigate a *Salmonella* outbreak in 2013, this method was applied in four further studies in the UK between 2014 and 2016. We used data from all five studies and interviews with members of each outbreak control team and market research panel provider to review operational issues, evaluate risk of bias in this approach and consider methods to reduce confounding and bias. The investigators of each outbreak reported likely time and cost savings from using market research controls. There were systematic differences between case and control groups in some studies but no evidence that conclusions on the likely source or vehicle of infection were incorrect. Potential selection biases introduced by using this sampling frame and the low response rate are unclear. Methods that might reduce confounding and some bias should be balanced with concerns for overmatching. Further evaluation of this approach using comparisons with traditional methods and population-based exposure survey data is recommended.

Introduction

Established methods of recruiting population controls for case–control studies to investigate gastrointestinal disease outbreaks can be time consuming, resulting in delays in identifying the source or vehicle of infection. In 2013, we evaluated the use of online market research panel members as controls in a case–control study conducted in response to an outbreak of *Salmonella* Mikawasima gastroenteritis in the UK. We have previously described methods for recruiting ‘panel controls’ [1]. In brief, control recruitment and data collection proceeds from distribution of an online questionnaire by a market research company to randomly selected members of their panel (Internet users who have opted to complete questionnaires in return for rewards) who meet study-specific inclusion criteria. Responses are collected from those panel members who respond most rapidly until the predefined target quota is complete.

We collected exposure data from 123 controls frequency matched by age, location and sex to cases of *Salmonella* Mikawasima. Data collection was complete within 9 h of survey launch at a cost of £3.60 per control (study A; Table 1). We considered this approach to be efficient and cost-effective compared with recruitment of controls among Public Health England (PHE) staff members (as used by the outbreak control team and against which panel controls were evaluated). However, although the same associations with some food vehicles and eating behaviours were identified irrespective of the control set used, there were systematic differences between panel controls and PHE staff. Evidence that either group might differ systematically from the general population was markedly different rates of some exposures (including proton pump inhibitors, some foods and participating in outdoor activities) and some differing associations with illness. This suggested that neither control group represented a common population, even if this did not affect the main conclusion on the likely vehicle of infection. However, it was unclear if controls derived from PHE staff, panel members or both differed

© Cambridge University Press 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Table 1. Key features of case-control studies conducted in response to gastrointestinal outbreaks using market research panel controls, UK, 2013–2016

Study Identifier (ref)	A [1, 2]			B [3]			C [5]			D [6]			E [6]		
Year	2013			2014			2015			2015			2016		
Pathogen	<i>Salmonella</i> Mikawasima			STEC O157 PT8			<i>Salmonella</i> Enteritidis (5 SNP cluster)			<i>Cryptosporidium parvum</i> IIdA24G1			STEC O157 PT34		
Hypothesised exposure from trawl (noted excess of demographic groups)	A large proportion of cases reported taking PPIs prior to illness but no clear exposure associated with illness was identified. No excess of cases noted in defined age or gender groups			Handling or consumption of potatoes, root vegetables, Tomatoes, apples or bananas. No excess of cases noted in defined age or gender groups			Exposure to reptiles, particularly snakes. Excess of cases in children			Consumption of pre-prepared sandwiches with specific fillings, food bought from branches of two supermarkets and one coffee shop chain A, specific dairy products or consumption and/or handling of specified salad vegetables. Excess of cases in adult females			Consumption of salad vegetables, bagged salad, food purchased from a specific supermarket chain and salad items from catering premises. Excess of cases in adult females		
Method of collecting case data	Telephone interview			Web-survey			Telephone interview			Telephone interview			Telephone interview		
Target control criteria ^a	Age restricted to 18 years and over. Frequency-matched on sex and investigating region. Intended 2 : 1 control-to-case ratio			Age restricted to 18 years and over. Intended 2 : 1 control-to-case ratio			Frequency matched on age (including households with children) and sex. Intended 3 : 1 control-to-case ratio			Age restricted to 20 years and over. Frequency matched on country. Intended 2 : 1 control-to-case ratio			Age restricted to 18 years and over. Frequency matched on sex and geographical area. Intended 4 : 1 control-to-case ratio		
Panel (X or Y)	X			X			X			X			Y		
Number of controls (and cases) included in study	123 ^b (39 ^c)			96 (36 ^c)			180 ^b (28 ^c)			133 ^b (52 ^c)			91 ^b (21)		
Number of survey invitations distributed (% included in study)	1329 (9%)			Data not available			Data not available			Data not available			4772 (2%)		
Price per control (£)	£3.60			£3.20			£3.00			£3.30			£2.00		
Time to organise with company (to point of survey distribution)	18 days			5 days			7 days			2 days			2 days		
Time to recruit and collect data (from survey distribution)	<1 day (9 h)			7 days			14 days			2 days			<1 day (10 h)		
Means of distributing survey	By email			By email			By email			By email			Survey invitations delivered directly to targeted individuals' home page for the panel and notified of invite by email		
Distributions	Group	Cases (%)	Controls (%)	Group	Cases (%)	Controls (%)	Group	Cases (%)	Controls (%)	Group	Cases (%)	Controls (%)	Group	Cases (%)	Controls (%)
1. Age	18–25 years	6 (15)	5 (4)	18–30 years	5 (14)	11 (11)	<10 years	9 (32)	0 (–)	20–29 years	23 (44)	7 (5)	10–19 years	1 (5)	0 (0)

(Continued)

Table 1. (Continued.)

Study Identifier (ref)	A [1, 2]			B [3]			C [5]			D [6]			E [6]		
Year	2013			2014			2015			2015			2016		
	26–35 years	8 (21)	19 (16)	31–64 years	16 (44)	63 (66)	10–19 years	6 (21)	0 (–)	30–39 years	15 (29)	23 (17)	20–29 years	5 (24)	2 (2)
	36–45 years	6 (15)	23 (19)	65+ years	15 (42)	22 (23)	20–29 years	7 (35)	14 (8)	40–49 years	13 (25)	29 (22)	30–39 years	1 (5)	7 (8)
	46–55 years	9 (23)	40 (33)	–	–	–	30–39 years	2 (7)	53 (29)	50–59 years	0 (–)	45 (34)	40–49 years	4 (19)	21 (23)
	56–65 years	10 (26)	36 (29)	–	–	–	40–49 years	3 (11)	66 (37)	60–69 years	1 (2)	29 (22)	50–59 years	4 (19)	22 (24)
	–	–	–	–	–	–	50–59 years	1 (4)	38 (21)	–	–	–	60–69 years	1 (5)	27 (30)
	–	–	–	–	–	–	60–69 years	0 (–)	7 (4)	–	–	–	70–79 years	3 (14)	12 (13)
	–	–	–	–	–	–	70–79 years	0 (–)	2 (1)	–	–	–	>80 years	2 (10)	0 (0)
P-value ^d	0.124			0.068			<0.001			<0.001			<0.001		
2. Sex	F	22 (56)	63 (51)	F	29 (81)	43 (45)	F	11 (39)	90 (50)	F	31 (60)	54 (41)	F	19 (90)	77 (85)
	M	17 (44)	60 (49)	M	7 (19)	53 (55)	M	17 (61)	90 (50)	M	21 (40)	79 (59)	M	2 (10)	14 (15)
P-value ^d	0.572			<0.001			0.293			0.020			0.768 ^e		
3. Geographical	Scotland	9 (23)	48 (39)	–	–	–	–	–	–	England	44 (84.6)	109 (82)	South East England	5 (24)	21 (23)
	North East England	13 (33)	32 (26)	–	–	–	–	–	–	Scotland	4 (7.7)	12 (9)	South West England	16 (76)	70 (77)
	West Midlands England	9 (23)	22 (18)	–	–	–	–	–	–	Wales	4 (7.7)	12 (9)	–	–	–
	London	8 (21)	21 (17)	–	–	–	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
P-value ^d	0.344			–			–			0.185			>0.999 ^e		
Study finding	Consumptions of a variety of chicken exposures, coriander, salad garnishes and eating out were significantly associated with illness			Consumption of pre-packed salad, and purchase of salad or potatoes from a specific retailer were independently associated with illness			Snake ownership was independently associated with illness			Coffee shop chain A was independently associated with illness			Mixed leaves and 'other' salad items associated independently with illness		

^aIntention to recruit controls based on number of cases expected to be included in the study.

^bThe target quota of recruited controls was exceeded before the web-survey was closed.

^cSome cases could not be contacted, were not eligible or did not agree to be interviewed.

^dEvidence of differences in distribution (at 5% level) between cases and controls.

^eFishers exact test (all others Chi-squared test).

in terms of exposure profile to the general population and how. Differences in the method of data collection (telephone vs. online questionnaire) complicated interpretation of the evaluation findings.

Market research panel control groups have subsequently been used in four investigations of national outbreaks of gastrointestinal pathogens in the UK between 2014 and 2016. Here we use data from all five studies between 2013 and 2016 to review operational issues, evaluate risk of bias in this approach and consider approaches to reducing confounding and bias.

Methods

Study design and information on costs, timeliness and operational considerations related to the use of panel controls was collated during interviews with members of each outbreak investigation team and by review of outbreak protocols, reports and published literature, as availability allowed. The main findings from each outbreak were reviewed to identify whether the main factors associated with illness in the corresponding case-control study were consistent with other available information. Data from each outbreak were analysed to identify differences in demographic characteristics between cases and controls. Representatives from the market research panel providers used in these studies were interviewed either in person or via email to better understand the process of identifying and collecting data from panel members and checks on data quality and to identify issues that could contribute to non-population representativeness of panel controls.

Between 2013 and 2016, market research panel control groups were used in five investigations of diverse national outbreaks of gastrointestinal pathogens after observed increases in cases of: (study A) *Salmonella* Mikawasima with no clear excess among defined age or gender groups and no hypothesised exposures associated with illness identified from trawling questionnaires [2]; (study B) Shiga toxin-producing *Escherichia coli* (STEC) O157 with no clear excess among defined age or gender groups and handling or consumption of potatoes, root vegetables, tomatoes, apples or bananas identified as hypothesised exposures associated with illness [3]; (study C) *Salmonella* Enteritidis 5 single nucleotide polymorphism (SNP) single linkage cluster (methods for defining SNP profiles have been described previously [4]) with an excess observed among children and exposure to reptiles, particularly snakes, identified as hypothesised exposures associated with illness [5]; (study D) *Cryptosporidium parvum* IIdA24G1 with an excess observed among adult females and consumption of pre-prepared sandwiches with specific fillings, food bought from branches of two supermarkets and one coffee shop chain, specific dairy products or consumption and/or handling of specified salad vegetables identified as hypothesised exposures associated with illness [6]; and (study E) STEC O157 PT34 with an excess observed among adult females and consumption of salad vegetables, bagged salad, food purchased from a specific supermarket chain and salad items from catering premises identified as hypothesised exposures associated with illness [6] (Table 1).

Results

Cost and timeliness

The time required to organise with a market research company the content and distribution of a web-survey to a target quota

of a defined subset of panel members meeting study-specific criteria (up to the point of distribution to panel members) has decreased over time from 18 days (Table 1; study A) to 2 days (studies D and E). The time required to recruit at least the target number of controls after web-survey distribution ('campaign launch') ranged from 9 h (study A, $n = 123$) to 2 weeks (study C, $n = 180$). The cost per recruited control ranged from £2.00 (study E) to £3.60 (study A). These costs varied according to factors including the target number of controls, restriction and frequency-matching criteria and market research panel used. Each outbreak control team reported that the recruitment of controls using this method was timelier and required far less staff time than using other approaches, such as random or sequential digit dialling, based on prior experience.

Study findings

Each case-control study using panel controls found plausible associations with at least one exposure or premises that had been identified during hypothesis generation to be tested in an analytical study, where hypotheses to test had been identified (Table 1). For studies B, C and E, associations that had previously been associated with pathogen-specific gastrointestinal disease outbreaks in the UK were identified (study B, raw potatoes [7] and bagged salad [8]; study C, feeder mice for snakes [9]; study E, bagged salad [8]). For study C, the implicated strain was subsequently isolated from an epidemiologically linked exposure (mice fed to reptiles) during parallel microbiological investigations. For study E similar results were subsequently found from cohort, case-case and venue-based studies and a common supplier found for leaves consumed by cases. For study A, which was the original evaluation, a parallel analysis was conducted using PHE staff controls and this also identified the same chicken and eating out exposures (though each identified additional independently associated exposures).

Epidemiological approach

The studies used a control-to-case ratio of at least 2 to 1 (and up to 4 to 1) to ensure sufficient statistical power to test hypotheses and identify associations with a minimum odds ratio. In all but study B, which used a web-survey to collect case exposure data, paper surveys were administered to cases by telephone interview. Case questionnaires (web-survey or paper based) and control web-surveys for a given study used consistent question phrasing.

Four of the five studies frequency-matched controls to cases by at least one criterion (Table 1; studies A, C, D and E); one frequency matched on age (study C), three on sex (studies A, C and E) and three on geographical unit (studies A, D and E). Studies A-C and E restricted the sampling frame for recruitment to those aged 18 years and over while study D restricted to those aged 20 years and over. Study C included those aged less than 18 years in the target strata though information was collected from parents. The number of cases included in the study was less than expected when setting targets for panel controls, as per the protocol, in studies A, C and D and the number of controls was greater than the target number in studies A and C-E (Table 1). In study B, there was deliberate oversampling of controls with the intention of frequency matching by age and sex using a randomly selected subset but ultimately there were insufficient numbers in some strata to accommodate this approach.

For studies that frequency-matched controls to cases on geography and sex, there was no evidence of a difference in the distribution of these matching variables between cases and controls at the 5% level (Table 1). For study C, in which there was frequency matching on age, and studies D and E, in which there was no frequency matching on age, there was evidence at the 5% level that controls were older than cases ($P < 0.001$). For those studies that did not frequency match on sex, there was evidence that controls contained fewer women than cases (study B, $P < 0.001$; study D, $P = 0.02$).

Data on response rate to surveys shared with panel members was available from the market research panel for only two studies: for study A, of 1329 panel members with whom a web-survey was shared, 9% provided a complete, eligible responses; and for study E, of 4772 panel members with whom a web-survey was shared, 2% provided complete and eligible responses. Ethical approval was not required prior to recruitment of market research panel members (as is typical in the context of outbreak investigation), no personal identifiable information was collected and no confidentiality issues or other data governance challenges were identified using this approach.

Market research panels

Two market research panels (panels X and Y) both with more than 200 000 panelists in the UK were used in these five studies; panel X was used in studies A–D and panel Y in study E. Both panels use a variety of online and offline methods of recruitment including referral programs, search engine optimisation, offline print trade marketing, location-based registration and radio advertising. The distribution of demographic factors among panel members is assessed against that of the general population and ‘while considered to be largely representative in terms of regional, social and age factors’ both panels X and Y had an over-representation of younger and female members. Panel X can target panel members to local authority level, while panel Y can target panel members to the postcode level.

Rather than distributing survey invitations to a random selection of targeted panel members by email as panel X did for studies A–D, for smaller target populations panel Y can (and did so for study E) deliver survey invitations to the profile page of a random selection of targeted panel members on their market research company website which triggers email notifications. Panel Y reported that this approach to survey delivery did not impact the number of panel members invited to participate. Speed of response of panel members to a survey invitation was considered by both market research companies to be influenced by factors including age (the younger being slower to respond) and time of year (slower response around holiday periods). Data quality of responses from panel members is monitored; unreliable respondents (identified either by clients, the panels when they are employed to do data analysis on client data or, for panel Y, through the use of intermittent quality check surveys) and those that rush through surveys (identified by monitoring the time taken to complete) are removed from the panel. Both monitor how active their members are, based on time since last survey, and only use engaged members.

Both companies maintained a policy – in line with industry standards in confidentiality – that name, telephone number and postcode could not be collected from panel members. While both panels could provide data on a metric for socioeconomic status (socio-economic group determined by the occupation of the

head of the household of the panel member), only panel Y offered to append indices of multiple deprivation (IMD) score (and potentially truncated postcode) to collected respondent data if they were supplied with appropriate postcode to IMD score look up tables. Panel members must be at least 18 years old but consenting panel members can report on their children’s exposures or allow their child to complete a survey under their supervision. Details of ethnicity and sexual orientation can be requested from panel members but such questions cannot be mandatory or be used to screen respondents out or to define frequency-matching strata.

Discussion

Our experience of five outbreak investigations using market research panel controls indicates that, in the view of the investigators, there were substantial time and cost savings compared with other approaches. More timely control recruitment and analysis should lead to timelier public health action, such as traceback investigations and recall of a short shelf-life product, which is important during an outbreak to limit additional cases. There were differences between case and control groups in measurable factors such as age and sex, which potentially complicates analysis and interpretation of the results. There was no evidence that conclusions regarding the likely vehicle or source of infection were incorrect due to these differences, though the quantitative estimates of association could have been affected. Parallel microbiological investigations to one study provided microbiological evidence to support the epidemiological findings and additional analytical studies produced similar findings in two other outbreaks.

Individuals who volunteer to join these panels might systematically differ to those that do not by demographic or behavioural factors (including shopping or dining out patterns, food exposures and level or type of physical activities) and other studies report likely bias when using such panels [10, 11]. Both panels X and Y used in these studies referred to their panels as representative of the general population but reported an over-representation of younger and female members [personal communication with panel Y representative]. The distributions of factors including age, sex, geography, ethnicity, measures of socio-economic status, behaviours and food exposure history might differ by market research company and potentially within a single market research company over time as a consequence of different or changing strategies for panel member recruitment from the general population, respectively. The low reported response rates indicate that these studies might be vulnerable to the introduction of selection bias as differing segments of the panel were known to respond with different speeds (e.g. younger panel members are slower to respond); the distribution of demographic factors between respondents and non-respondents should be compared in future studies to assess the potential for introduction of such bias.

To assess in which scenarios or for which hypothesised food exposures or behaviours this method might be most appropriate, any selection biases introduced by using these panels need to be better understood. It might be appropriate to compare the reported food exposures and behaviours of market research panel members with that from probability sampled, population-based food exposure and behaviour surveys, where they exist [12], or other sources of such data for which selection biases are minimised or previously characterised [13]. Studies using

panel controls could also be rerun using additional control groups from traditional sources in analyses to validate findings until the biases are better understood. Evaluations of representativeness of panel populations should be specific to each company and should potentially be repeated as methods of recruitment might change over time and potentially introduce new selection biases. Once any selection biases associated with using a particular panel population are understood, then they can be corrected for in future studies.

Challenges with market research panel controls, including selection bias, are not unique to this approach of recruitment [14]. There are many possible approaches to reducing confounding as well as analytical approaches to dealing with it that are commonly used when other methods of recruiting controls are employed. In several of the outbreak studies, panel controls were frequency matched on one or more demographic factor to cases (ensuring that the distribution of controls matches that of cases) to address potential confounding. In addition, such matching should help address the potential selection bias that might be introduced as a result of the differing speed with which certain segments of the panel member population might respond to a survey and in turn the likelihood of these segments being included in the control group; differences are seen in age distribution between cases and controls in some studies where controls are not frequency matched to cases by age. While confounding can be assessed and controlled for in the analysis, capacity to do so might be limited if there are insufficient numbers of controls in strata of a certain factor.

It is not appropriate to frequency match on factors associated with putative causal exposure due to the risk of overmatching, which can obscure an association between an exposure and an outcome. However, if an exposure is still associated with illness in the presence of overmatching, the association may be an underestimate. Study D did not frequency match on age or sex to avoid overmatching and recruited controls were older and consisted of more males than cases. Frequency-matching variables should be selected to balance the need for representativeness with avoidance of overmatching. In some instances, frequency-matching controls to a standardised distribution of demographic factors in the general population might be appropriate to ensure there was no overmatching but still address to some extent any lack of representativeness on demographic factors among panel controls.

Study C frequency matched on age but there was strong evidence that controls were older than cases. This is because panel members from households with children in certain age groups were used rather than requesting that these panel members provide responses on behalf of their children or supervise their children when completing the web-survey (and record the age of the child instead). Adults were also matched on two broad age groups but the analysis conducted using further stratified groups. However, as the implicated exposures were rare – snake contact and contact with feeder mice – any bias in the control group is unlikely to have generated a spurious association with illness, a finding validated by the detection of the outbreak strain in feeder mice.

Additional approaches to address confounding and some of the bias introduced by the use of panel controls – a biased sample taken from the panel – might include substantial oversampling to allow selection of a subset of more valid controls with the desired composition. Such an approach would likely be feasible given the low cost per questionnaire by this method. Alternatively, the

feasibility of sampling far more controls than needed to ensure adequate statistical power would support the application of propensity score approaches, which utilise the probability of exposures of interest being conditional on other characteristics, and might improve the internal validity of studies using these controls. An approach which might highlight if observed associations are affected by selection bias would be to repeat analyses comparing the cases to differing subsets of controls. To assist in assessing selection bias in further studies using panel controls, comparing the distribution of demographic characteristics of those that were eligible and completed the web-survey while it was open with those who did not is recommended (the necessary data should be available from the market research company on request).

The market research panels used in these studies have standards of recruitment to try to ensure a representative panel population on some demographic characteristics and use methods to maintain high quality responses by panel members to surveys. Methods for validating respondents, maintaining quality and recruiting respondents are specific to each market research company and should be thoroughly reviewed before using a company. Each market research company should be able to provide responses to The European Society for Opinion and Market Research (ESOMAR) 28 questions [15] designed to inform interested parties with regards to a supplier's practices and samples.

The time required to organise the distribution of an online survey by a market research company has decreased over time and this might be as a result of better familiarisation with and documentation of the process of recruiting controls using this method by PHE. The response rate to the distributed survey was low for the two studies where data were available (<10%; similar to that reported elsewhere for volunteer controls [11, 16] but low compared with other methods) but more panel members might have responded had the survey not been closed when the target was reached. In addition, the number of panel members to whom web-surveys were distributed depends on the target quota, complexity of the target population and their expected response rate (to ensure a sufficiently rapid target quota completion).

The panel control target quotas given to market research companies reflected an intention to recruit based on total number of cases and the target control-to-case ratio but ultimately the ratio for all studies reviewed here differed because either not all cases could be contacted, were not eligible or did not agree to be participate and/or there was accidental overrecruitment due to a web-survey not being shut when the target quota was full. For each study the control-to-case ratio was greater than intended and the power of the study to detect a minimum odds ratio should not have been diminished. Ensuring a market research panel provider was to distribute the web-survey links only once recruited case data is collected could prevent accidental overrecruiting and associated costs though this might cause some delay.

While all studies to date have investigated widely distributed exposures, both panels have a substantial number of members in the UK (over 200 000) and can refine target populations to at least local authority level. However, only panel Y can offer frequency matching (given the limitations of current PHE web-survey software), supplement collected data with IMD and potentially frequency match geographically to postcode level; while the number of panel members in a single postcode might be limited it means that any higher geographical level can be aggregated (including local authority and PHE sub-national operational

areas). As these companies often maintain panels in a number of countries or can collaborate with other companies, it is also plausible that this approach could be used to recruit controls for an international outbreak investigation.

Additional information – previous associations with pathogen-specific gastrointestinal disease outbreaks in the UK, microbiological evidence, findings from other parallel epidemiological studies or traceback activities – provided further validation of the findings of four studies (all except study D which was a coffee shop chain). For studies where hypothesised exposures or premises were identified, a plausible exposure with one of these was found. This method of control recruitment might be considered suitable for a variety of exposure types given that identified exposures linked to infection and validated by other sources of evidence were varied, including rare (contact with feeder mice for snakes) and more widespread exposures (specific retailers and raw food items).

Other developments which might provide less biased control exposure data in some scenarios include: a national exposure survey of individuals randomly selected from and therefore more representative of the population, as conducted in the Netherlands [12] but these would need to be repeated to account for changing exposure distribution among the general population by season and over time; harnessing existing systems of primary care record providers [17, 18] for delivering ad hoc questionnaires to patients via participating general practitioners but associated costs, governance and timeliness might make this approach prohibitory for outbreak investigation; a nationally pre-agreed approach for efficiently accessing controls directly through general practices; or use of aggregated routine sources of population data from pre-collected surveys, including shopping patterns, pet ownership [19] and food consumption [13, 20], which might provide efficient and valid information to complement control data but survey-specific methodologies, associated selection biases and length of validity given changing consumption patterns and behaviours over time should be considered.

To date, case-control studies using panel controls to investigate gastrointestinal disease outbreaks have demonstrated time and cost savings and have not obviously been influenced by bias in terms of the conclusion on the source or vehicle of infection. However, selection biases that are potentially introduced by this method, as with some other methods for recruiting controls, are not fully understood and may change as recruitment strategies of panel members from the general population differ between and, over time, within market research panel providers. Without validation, results from studies using panel controls could potentially be undermined technically if challenged during any related legal prosecutions. Further evaluation of the inherent biases associated with the use of market research panel members as controls is recommended so that they might be addressed in future studies.

Acknowledgements. We thank staff in the public health investigating teams within Health Protection Scotland, Public Health England and Public Health Wales who contributed to the studies considered in this review and Sam Bracebridge for her assistance in the preparation of the manuscript. The research was in part funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of East Anglia, University of Oxford and the Institute of Food Research. G. K. Adak, P. Cleary, R. Elson, J. Hawker, T. Inns and R. Vivancos are based at Public Health England and N. McCarthy at the University of Oxford. The views expressed

are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Declaration of Interest. The authors declare that they have no conflict of interest.

References

1. Mook P, *et al.* (2016) 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. *Epidemiology and Infection* **144**, 1220–1230.
2. Freeman R, *et al.* (2016) 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* **144**, 968–975.
3. Sinclair C, *et al.* (2017) Investigation of a national outbreak of STEC *Escherichia coli* O157 using online consumer panel control methods: Great Britain, October 2014. *Epidemiology and Infection* **145**, 864–871.
4. Ashton P, *et al.* (2015) Revolutionising Public Health Reference Microbiology using Whole Genome Sequencing: *Salmonella* as an exemplar. *bioRxiv* 33225. doi: 10.1101/033225.
5. Kanagarajah S, *et al.* (2017) Whole genome sequencing reveals an outbreak of Salmonella Enteritidis associated with reptile feeder mice in the United Kingdom, 2012–2015. *Food Microbiology* Published online: doi: 10.1016/j.fm.2017.04.005.
6. Gobin M, *et al.* (in press) National outbreak of Shiga toxin producing *E. coli* O157:H7 linked to mixed salad leaves, United Kingdom, 2016. *Eurosurveillance*.
7. Launders N, *et al.* (2016) A large Great Britain-wide outbreak of STEC O157 phage type 8 linked to handling of raw leeks and potatoes. *Epidemiology and Infection* **144**, 171–181.
8. Launders N, *et al.* (2013) Outbreak of Shiga toxin-producing *E. coli* O157 associated with consumption of watercress, United Kingdom, August to September 2013. *Euro Surveillance* **18**(44).
9. Harker KS, *et al.* (2011) An outbreak of Salmonella Typhimurium DT191a associated with reptile feeder mice. *Epidemiology and Infection* **139**, 1254–1261.
10. Erens B, *et al.* (2014) 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* **16**, e276.
11. Baker R, *et al.* (2010) Research synthesis AAPOR report on online panels. *Public Opinion Quarterly* **74**, 711–781.
12. Whelan J, *et al.* (2010) National outbreak of Salmonella Typhimurium (Dutch) phage-type 132 in the Netherlands, October to December 2009. Available at <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19705> (Accessed 8 January 2017).
13. **National Diet and Nutrition Survey:** results from Years 1 to 4 (combined) of the rolling programme for 2008 and 2009 to 2011 and 2012 – Publications - GOV.UK. Available at <https://www.gov.uk/government/statistics/national-diet-and-nutrition-survey-results-from-years-1-to-4-combined-of-the-rolling-programme-for-2008-and-2009-to-2011-and-2012> (Accessed 17 March 2016).
14. Waldram A, *et al.* (2015) Control selection methods in recent case-control studies conducted as part of infectious disease outbreaks. *European Journal of Epidemiology* **30**, 465–471.
15. **ESOMAR.** 28 Questions to Help Buyers of Online Samples – ESOMAR. Available at <https://www.esomar.org/knowledge-and-standards/research-resources/28-questions-on-online-sampling.php> (Accessed 2 May 2017).
16. Craig BM, *et al.* (2013) Comparison of US panel vendors for online surveys. *Journal of Medical Internet Research* **15**, e260.
17. **Clinical Practice Research Datalink - CPRD.** Available at <https://www.cprd.com/intro.asp> (Accessed 13 October 2016).
18. **IMS Health.** Available at <http://www.epic-uk.org/index.html>. (Accessed 13 October 2016).
19. **Pet Population** (2015) Available at <http://www.pfma.org.uk/pet-population-2015> (Accessed 19 September 2016).
20. Available at <http://www.ukbiobank.ac.uk/2011/03/detailed-diet-questionnaire-underway/> (Accessed 17 March 2016).