

## Original Paper

**Cite this article:** Perriat D, Mertens E, Dreesman J (2022). Use of a non-probabilistic online panel as a control group for case-control studies to investigate food and waterborne outbreaks in Lower Saxony, Germany. *Epidemiology and Infection* **150**, e53, 1–9. <https://doi.org/10.1017/S0950268821002594>

Received: 25 August 2021

Revised: 4 November 2021

Accepted: 22 November 2021

**Keywords:**

Epidemiological study design; case-control studies; outbreaks; gastrointestinal infection; online panel

**Author for correspondence:**

Delphine Perriat,

E-mail: [delphine.perriat@gmail.com](mailto:delphine.perriat@gmail.com)

# Use of a non-probabilistic online panel as a control group for case-control studies to investigate food and waterborne outbreaks in Lower Saxony, Germany

Delphine Perriat<sup>1,2</sup> , Elke Mertens<sup>2</sup> and Johannes Dreesman<sup>2</sup>

<sup>1</sup>European Programme for Intervention Epidemiology Training (EPIET), European Centre for Disease Control and Prevention (ECDC), Solna, Sweden and <sup>2</sup>Department for Infectious Disease Epidemiology, Public Health Agency of Lower Saxony, Hannover, Germany

**Abstract**

Established methods of recruiting population controls for case-control studies in infectious disease outbreak investigations are resource- and time-intensive, and are often subject to bias. The online panel have recently gained interest as an easy and timely method to select controls. We examined the feasibility, suitability and reliability of using an online panel to select controls for case-control studies as part of investigations of diffuse food and waterborne outbreaks. In January 2019, we deployed a web survey by email to the 277 members of a non-probabilistic online panel in Lower Saxony, Germany. We questioned them on basic sociodemographic characteristics and eating habits. They were frequency matched to cases on sex and age. Their food exposures were compared to those of traditionally recruited controls of four historical case-control studies, which successfully investigated food and waterborne outbreaks. We used logistic regressions to assess the association between the food exposures and the disease (odds ratios). The use of a control panel successfully led to the identification of the food items in three of the four historical outbreak investigations, and their recruitment benefitted from increased speed and limited costs. Timely outbreak investigations would enable rapidly implementing control measures. We recommend the further evaluation of using panellists as controls in parallel case-control studies and case-panel studies.

**Introduction**

In Germany, infectious disease outbreaks continue to pose a high burden on public health [1]. Control measures should be timely and rapidly implemented in order to prevent new infections and protect the health of the public. To decide on the best control measures, it is key to identify the outbreak source. Case-control methodologies are commonly used to achieve this [2]. Outbreak cases and controls are questioned about possible risk factors, and their answers are compared (e.g. food exposures during foodborne outbreaks).

Established methods of recruiting controls include case- or physician-nominated controls, random or sequential digit dialling and convenience sampling [2–4]. In addition of being resource-intensive, such methods are mostly slow, resulting in delays in identifying the source of infection, which can be vital for stopping severe outbreaks [5]. Additionally, those methods can be subject to selection bias, as controls are not representative of the population at risk.

New and innovative methods are required, which are easy and timely to implement and reduce bias [6]. Web surveys are increasingly used to collect data, including in the field of health and in epidemiological surveys as they are cost-effective while maintaining scientific rigour [7–10]. Moreover, methods that allow prior recruitment of controls are promising, because they offer the possibility to improve the timeliness and the representativeness of controls for the population that gave rise to the cases [11].

Recently, commercial online panels have shown encouraging results [12–14]. Those panels comprise individuals who have elected to receive internet-based questionnaires and can opt to complete them, usually in return of a reward. They are frequently used by marketing companies or polling organisations to obtain information about a target audience. Commercial online panels have successfully been used to collect epidemiological exposure data in retrospective [15] and prospective case-control studies [16–21]. One of the key benefits of this approach is that it allows efficiently and rapidly recruiting many controls, therefore enabling timely investigation of outbreaks.

According to German law, the primary responsibility of outbreak investigations lies with the local health authorities, and the coordination can be transferred to the state or national health authorities in complex situations (e.g. when the outbreak spread to several localities or regions) [22]. Analytic studies are a key component of outbreak investigations [23]. However, they are often not conducted as they require human and financial resources beyond

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

typical case investigation and case finding efforts [24]. Local health authorities would highly benefit from the availability of online panels to recruit controls. However, such a method remains beyond their reach because of their limited public health resources. As part of efforts to strengthen the investigation of food and waterborne outbreaks of infectious gastrointestinal diseases at the local level, we identified the need to explore the use of an online panel as a source of controls for case–control studies as part of outbreak investigations.

The ‘Hygiene and Health Online Survey (HuGO)’ panel consists of 277 adults, who live in the federal state of Lower Saxony, Germany. They accepted to regularly answer online to health- and hygiene-related questions. We examined the feasibility, suitability and reliability of using the non-probabilistic online HuGO panel as a control group to recruit controls for case–control studies to investigate food and waterborne outbreaks in Lower Saxony, Germany.

## Methods

The data sources and data analysis steps are summarised in Table 1.

### Data sources

We used three data sources:

- The Lower Saxony Microcensus 2018: The Microcensus contains basic sociodemographic information on a randomly selected sample of the population living in Lower Saxony, Germany ( $n = 6537$ ), including gender, nationality, age, household composition, school and postgraduate training and employment status. This sample is considered representative of Lower Saxony [25].
- Four historical case–control studies: The four historical studies successfully investigated food and waterborne outbreaks in Lower Saxony and in the neighbouring German federal States between 2001 and 2017, using case–control designs: *Campylobacter jejuni* infection via tap water (study A), *Salmonella enterica Bovismobificans* infection via raw pork (study B), *Salmonella enterica Goldcoast* infection via raw pork (study C) and *Salmonella enterica Oranienburg* via chocolate (study D). Details of the conduction of the studies are given in Table 2. Parallel microbiological investigations to the studies provided microbiological evidence that supported the epidemiological findings.
- The Hygiene and Health Online Survey (HuGO) panel: To form the HuGO panel, participants from the Hygiene and Behaviour Infectious Diseases Study (HaBIDS) panel were integrated. The HaBIDS panel was a population-based longitudinal panel, created by the Helmholtz Centre for Infection Research in 2014. A total of 2,379 participants aged 15 to 69 years were drawn by means of probability sampling from the regional population registries of Lower Saxony [26]. They were regularly consulted to answer online to questions on knowledge, attitudes and practices related to infections in Lower Saxony. In 2015, the HaBIDS panel included 1,037 participants aged more than 18 years old [27]. In 2018, they were offered to form the HuGO panel, led by the Public Health Agency of Lower Saxony (Niedersächsisches Landesgesundheitsamt, NLGA, <https://www.nlga.niedersachsen.de/>). The panel aims at supporting the activities of the Public Health Agency in infection prevention and control. This includes the investigation of outbreaks and the quick assessment of the population perceptions towards

infection control measures. A total of 277 members of the HaBIDS panel agreed to form the HuGO panel. On 17 January 2019, the HuGO panellists were asked by email to answer an online survey, in which they were questioned about their sociodemographic characteristics and eating habits of the past week. The sociodemographic questions were based on the Lower Saxony Microcensus 2018, and the eating habit questions were based on the four historical case–control studies. A reminder was sent on 22 February 2019. As controls of case–control study should be free of the outcome of interest, panellists with gastrointestinal symptoms in the week prior to answering the questionnaire were filtered out using a dedicated question at the beginning of the online questionnaire.

### Data analysis

We explored the use of the panel according to three components:

- **Feasibility of using the HuGO panel as a control group in outbreak investigations.** We reported the human and financial resources required to conduct the case–control study using panellists as controls, including the creation of the online questionnaire. We reported the response rate among HuGO panellists after receiving the questionnaire: on the same day, after one day, after one week, after receiving a reminder, and in total. When available, we provided information on the time, human and financial resources required to conduct the historical studies.
- **Suitability of the HuGO panel for Lower Saxony in terms of basic sociodemographic characteristics.** We compared the panellists to the sample of the Lower Saxony Microcensus 2018 on sex, age, nationality, household composition, education level and employment status, using  $\chi^2$  goodness of fit tests.
- **Reliability of using the HuGO panel as a control group in the investigation of four historical outbreaks.** We refer as ‘historical studies’ to case–control studies with historical cases and historical controls, and as ‘HuGO studies’ to case–control studies, with historical cases and HuGO panellists as controls. In the HuGO studies, we controlled for the possible confounding effects of sex and age by matching the frequencies between historical cases and panellists on both variables (Supplementary File 1). When the information on sex or age of historical cases was not available, the frequencies of panellists were matched to those of the Lower Saxony Microcensus 2018. Separately in historical and in HuGO studies, we used univariable logistic regression analyses to assess the association between the food and water exposures and the disease. When possible, we performed multivariable logistic regressions. We included in the regression models all exposure variables that were associated with the disease with a  $p$ -value less than 0.2 in univariable analyses. We also included in the models the categorical variables for age and sex (as possible confounders) when the controls were not frequency matched to cases on those variables. Odds ratios (OR) were used to determine whether a particular exposure was a risk factor for the occurrence of the disease. The 95% confidence interval (95% CI) was used to estimate the precision of the OR. An exposure was regarded as a risk factor if the OR was  $>1$  and the  $p$ -value  $<0.05$ , or as a protective factor if the OR was  $<1$  and the  $p$ -value  $<0.05$ . If an exposure was a risk factor or a protective factor in both historical and HuGO studies, we considered that both studies had similar results.

All analyses were performed using the statistical software R.

**Table 1.** Overview of the study methods to investigate the feasibility, suitability and reliability of using the Hygiene and Health Online Survey (HuGO) panel as a control group for case-control studies to investigate food and waterborne outbreaks in Lower Saxony, Germany, 2019

Data sources	Data analysis
<b>Lower Saxony Microcensus 2018:</b> population sample, which is considered representative of Lower Saxony ( $n = 6537$ )	<b>Suitability:</b> Comparison of the sociodemographic characteristics of the HuGO panellists with participants to the Lower Saxony Microcensus 2018 (chi2 goodness of fit test).
<b>Hygiene and Health Online Survey (HuGO) panel:</b> non-probabilistic online panel comprised of 277 adults living in Lower Saxony, Germany, who accepted to answer hygiene- and health-related questions, 2019	<b>Feasibility:</b> Computation of response rates among HuGO panellists, and description of the human and financial resources required to conduct the HuGO and historical studies.
<b>Historical studies:</b> four case-control studies that successfully investigated food and waterborne outbreaks in Lower Saxony and neighbouring federal states in Germany, 2001–2017	<b>Reliability:</b> Comparison of the strength of the association (using odds ratios) between food and water exposures and disease, using the four 'historical studies' (case-control studies with historical cases and historical controls) vs. the four 'HuGO studies' (case-control studies with historical cases and HuGO panellists as controls).

## Results

### Feasibility

One scientist, working for the Public Health Agency of Lower Saxony conducted the study. A working day was required to create the online questionnaire using the Lamapoll® software. Apart from the software cost, no additional financial resource was required. A total of 203/277 (73%) HuGO panellists answered the survey: 76/277 (27%) answered the questionnaire on the day they received it, adding up to 107/277 (39%) on the following day and to 152/277 (54%) within a week. After receiving a reminder, a month later, 30/277 (11%) additional panellists answered.

### Suitability

There were statistically significant differences in the distribution of all measured sociodemographic characteristics between the panellists and the population of Lower Saxony (Table 3). For example, the proportion of women participating in the HuGO panel was higher than in the population of Lower Saxony (63% vs. 51%), the proportion of people younger than 30 was lower (7% vs. 17%) and a higher proportion attended university (32% vs. 9%).

### Reliability

No panellists experienced gastrointestinal symptoms in the week prior to the survey. Table 4 summarises the results of the univariable logistic regression analyses for the food items that were identified as outbreak sources in the historical studies. Results of univariable analyses where controls were frequency matched to cases on sex and age, and results of multivariable analyses for all food items are provided in the Supplementary File 2.

Using panellists as controls to investigate the four historical outbreaks gave consistent results in three instances. In the studies B, C and D, the ORs of the outbreak sources were significant and of similar magnitude to those of the historical studies. In Study A, drinking tap water at home was associated with the disease in the historical study (Odds Ratio in univariable analysis (ORu) = 9.6 (1.8–51)), but not in the HuGO study (ORu = 1.6 (0.33–7.5)). Results were consistent in univariable and multivariable analyses, with or without frequency matching.

## Discussion

We investigated the feasibility, suitability and reliability of using panellists as controls in case-control studies to investigate food and waterborne outbreaks. We found that using panellists would lead to similar results as using traditional controls, and the study would benefit from increased speed of recruitment as well as limited costs for public health action.

### Timely and low-cost outbreak investigation thanks to panel controls

The majority of panellists responded to the online questionnaire in less than a week, ensuring a higher response rate than traditional methods such as random digit dialling [9, 13]. The recruitment of panellists as controls also required far less staff and financial resources than other approaches, as it amounts to sending an email rather than conducting numerous face-to-face or telephone interviews [12, 14]. In the event of an actual food or waterborne outbreak, the HuGO panel offers a more timely and cost-effective control recruitment and analysis, which would lead to timelier public health actions to limit additional cases (e.g. trace-back investigations, recall of products). It is nevertheless important to consider that the constitution of the HaBIDS panel, from which the HuGO panel stems, did require substantial resources. Indeed, the 2,379 panellists were recruited thanks to the dedication of a 1–2 researchers, who drew 16,000 people from the registries of Lower Saxony and send them invitation letters per post therefore reaching an overall response rate of 8.9% [26].

### Suitability of panel controls for outbreak investigation

The study assessed the suitability of panel controls for case-control studies. Ideally, one would be confident using panellists as controls if these were not more biased than controls used in current best practices. Controls are used to estimate the prevalence of exposures in the population that gave rise to the cases. They are expected to be free of the outcome of interest, representative of the population at risk of the outcome and selected independently of the exposure of interest [28]. In the situation of an outbreak investigation, case-control studies aim to estimate a particular exposure-disease association, while appropriately controlling for confounding and avoiding other biases.

**Table 2.** Key features of four historical case-control studies that successfully investigated food and waterborne outbreaks in Germany between 2001 and 2017

Study identifier	A	B	C	D
Study period	July 2017	February 2005	June 2004	September 2001
Study area	City of Lamspringe, Lower Saxony	Lower Saxony	Saxony Anhalt (neighbouring state of Lower Saxony)	Several German states (including Lower Saxony)
Hypothesised exposure	Consumption of contaminated tap water	Consumption of contaminated raw pork	Consumption of contaminated raw pork	Consumption of contaminated chocolate
Pathogen (microbiological investigation)	<i>Campylobacter jejuni</i>	<i>Salmonella enterica Bovismobificans</i>	<i>Salmonella enterica Goldcoast</i>	<i>Salmonella enterica Oranienburg</i>
Control selection	Random digital dialling among inhabitants of Lamspringe	Random digital dialling among inhabitants of Lower Saxony	Random digital dialling among inhabitants of Saxony Anhalt	Random digital dialling among inhabitants of Germany, with 1:1 individual matching on sex and age
Data collection	Telephone interviews conducted by 4 staff members of the state health authority during 7 days. Overall 111 persons were called; 46 accepted to participate (41% response rate), among which 35 met the control selection criteria (e.g. presence in Lamspringe during the outbreak period)	Telephone interviews	Telephone interviews	Telephone interviews conducted by a large team of local, state and national health agency staff
Number of controls [Number of cases]	35 [12]	37 [38]	54 [14]	50 [48]
Studied exposures	<ul style="list-style-type: none"> <li>sex, age</li> <li>consumption in the last 7 days of unboiled tap water at home, unboiled tap water outside of home</li> </ul>	<ul style="list-style-type: none"> <li>sex, age</li> <li>consumption in the last 7 days of raw pork, raw sausage pork, cooked pork, raw beef, raw egg, salad, sprout</li> </ul>	<ul style="list-style-type: none"> <li>sex, age</li> <li>consumption in the last 3 days of raw pork, raw sausage pork, cooked pork, raw beef, raw egg</li> </ul>	<ul style="list-style-type: none"> <li>sex, age</li> <li>shopping in the last 7 days in the supermarket chain X</li> <li>consumption in the past 7 days of chocolate bought in the supermarket chain X</li> </ul>
Available data source	Individual-level data	Individual-level data	Aggregated data	Aggregated data
Outbreak details	A heavy rainfall led to a contamination of water supplies with surface water in the city of Lamspringe. Interviews of controls were conducted around 14 days after the first cases developed symptoms. At that time, information related to the likely contaminated tap water was communicated in the newspapers, with a recommendation to drink bottled water.		In 2004, Lower Saxony and Saxony Anhalt reported most of the <i>S. Goldcoast</i> cases in Germany [39]. A majority of cases included in the case-control study reported having eaten raw pork, which was found to have been supplied by a slaughterhouse in Lower Saxony. <i>S. Goldcoast</i> was found in raw pork as part of self-checks in this slaughterhouse. In Lower Saxony, over 90% of the surveyed cases stated that they had eaten raw pork products in the 3 days before disease onset.	A trace-back analysis of the implicated chocolates along supply chains allowed to identify that they were contaminated prior to their distribution in Germany by the supermarket chain X, and in other European countries [40]

**Table 3.** Comparison of sociodemographic characteristics between the Hygiene and Health Online Survey (HuGO) panellists ( $n = 203$ ) and the population of Lower Saxony (Lower Saxony Microcensus 2018)

	HuGO-Panel ( $n = 203$ )		Lower Saxony Microcensus ( $n = 6537$ )		$\chi^2$ goodness of fit test $p$ -value
	$n$	%	$n$	%	
Sex					
male	74	37	3232	49	
female	129	63	3304	51	<0.001
Nationality					
German	201	99	5956	91	
not German	2	1	581	9	<0.001
Age group					
$\geq 18$ and <30	14	7	1110	17	
$\geq 30$ and <40	24	12	938	14	
$\geq 40$ and <50	41	20	1054	16	
$\geq 50$ and <60	73	36	1286	20	
$\geq 60$	51	25	2148	33	<0.001
Household					
lives alone	29	14	1630	25	
$\geq 2$ persons	174	86	4907	75	<0.001
Education level					
no post-secondary education <sup>1</sup>	5	2	1612	25	
apprenticeship <sup>2</sup>	80	39	3447	53	
technical college <sup>3</sup>	17	8	481	7	
university of applied science <sup>4</sup>	36	18	400	6	
university <sup>5</sup>	65	32	595	9	<0.001
Employment status					
has a job	147	72	3916	60	
does not have a job or is retired	56	28	2621	40	<0.001

German translations: <sup>1</sup>kein Berufsabschluss, <sup>2</sup>Lehre/Berufsausbildung, <sup>3</sup>Fachschulabschluss, <sup>4</sup>Fachhochschule/Berufsakademie, <sup>5</sup>Universität / Promotion.

Selection bias is a particular problem inherent in case-control studies, where it gives rise to non-comparability between cases and controls, threatening the generalisability of the study results [2]. Our study is affected by a sampling bias, as HuGO panellists volunteered to participate in the study. First, HuGO panellists differ from the general population of Lower Saxony in terms of sociodemographic factors such as age, sex and education level. We reduced this bias by matching the frequencies between historical cases and panellists on sex and age, therefore controlling for the possible confounding effects of both variables. Additionally, controls recruited through traditional methods are also rarely representative of the general population [2]. It is largely accepted that a lack of representativeness on socio-demographic characteristics does not hamper scientific inference [29, 30]. Valid scientific inference is achieved if the confounders are controlled for, and there is no reason to believe that control of confounding can be more easily achieved in a randomly selected control group than in a panel group [29]. Second HuGO panellists are significantly more educated than the general population of Lower Saxony. As studies have shown that a higher education status

is associated with healthier eating habits [31], it is likely that HuGO panellists may be more health-conscious than the general population [7, 12].

Using the HuGO panellists as controls decreased the likelihood of information bias as compared to using traditional controls, resulting in a greater validity of the food exposure information. First, there is little recall bias as panellists responded within a few days to the questionnaire. As traditional controls are often recruited with days or weeks of delay after the outbreak occurs, they have more difficulties to remember their food exposures at a particular time point. Panellists would then be particularly useful when investigating outbreaks caused by uncommon food exposures, as they are oftentimes more forgotten than common food exposures [32].

Additionally, online surveys are less likely to suffer from social desirability bias as other data collection methods, such as phone or face-to-face interviews [26]. Panellists are less likely to under-report 'bad' food items, and over report 'good' food items than traditional controls. They would therefore also be particularly adequate to investigate sensitive exposures, benefitting from the survey anonymity [12, 26].

**Table 4.** Comparison of the strength of association between food and water exposures and disease (odds ratios) in univariable univariable logistic regression analyses between historical and HuGO case-control studies to investigate food and waterborne outbreaks, Germany 2019

Study*	Outbreak source	Historical cases		Historical controls		HuGO panel controls		Historical study	HuGO study	Historical study (black) and HuGO study (blue)		Similar results in both studies	
		N	%	n	%	n	%	OR (95%CI) p	OR (95%CI) p	● OR	— (95%CI)		
A	Tap water												
	Total	12	100	35	100	203	100						
	No	2	17	23	66	49	24	9.6 (1.8–51) p < 0.05	1.6 (0.30–7.5) p = 0.70			No	
	Yes	10	83	12	34	154	76						
B	Raw pork												
	Total	38	100	37	100	203	100						
	No	21	55	34	92	171	85	9.2 (2.4–35) p < 0.001	4.3 (2.1–9.1) p < 0.001			Yes	
	Yes	17	45	3	8	32	16						
C	Raw pork												
	Total	14	100	54	100	203	100						
	No	5	36	47	87	171	84	12 (3.1–47) p < 0.001	9.6 (3.0–30) p < 0.001			Yes	
	Yes	9	64	7	13	32	16						
D	Chocolate												
	Total	48	100	50	100	203	100						
	No	33	75	43	96	200	98	5.0 (1.1–47) p < 0.05	22 (5.8–83) p < 0.001			Yes	
	Yes	11	25	2	4	3	2						

Historical study: case-control study with historical cases and historical controls, HuGO study: case-control study with historical cases and HuGO panel controls (2019), OR: Odds Ratio, 95% CI: 95% Confidence Interval, p: two-sided p-value from Fisher's exact test, \*The four historical studies successfully investigated food and waterborne outbreaks in Lower Saxony and in the neighbouring German federal states, using case-control designs: *Campylobacter jejuni* infection via tap water in 2017 (study A), *Salmonella enterica Bovismobificans* infection via raw pork in 2005 (study B), *Salmonella enterica Goldcoast* infection via raw pork in 2004 (study C) and *Salmonella enterica Oranienburg* via chocolate in 2001 (study D).



### Reliability of control panels to identify outbreak sources

The study explored the circumstances in which using panellists would be reliable to investigate food and waterborne outbreaks. Indeed, investigators expect consistent conclusions regarding the likely outbreak source, regardless of the study design they use for the investigation.

In the present study, three of four outbreak sources could be identified using panellists as controls (studies B, C and D). The differences in the magnitude of the effect estimates (odds ratio) between historical and HuGO studies did not affect the ability of panellists to successfully identify the outbreak sources. As odds ratios measuring associations between contaminated food items and the disease are usually very high during food and waterborne outbreaks [32], we argue that they are not significantly affected by differences in exposure proportions between panellists and traditional controls.

The study shows that in outbreak A, the association between the consumption of tap water and gastrointestinal disease could not be identified with a statistical significance by using panellists as controls. We hypothesise that, in this particular situation, panellists did not have the chance of being exposed to the very localised outbreak source (contamination of water supplies with surface water due to heavy rainfall), and were thus not representative of the population at risk of the outcome. They did not live in the same area as the cases, and were not questioned at the particular time during which the waterborne outbreak occurred. This result furthers the argument that the reliability of panellists is likely to be higher when used as controls in the investigation of region- or nationwide outbreaks (like outbreaks B, C and D) as compared to very local outbreaks (like outbreak A). On the other hand, historical controls were likely to have changed their drinking habits, and drank significantly less tap water than usual, as they were interviewed around 14 days after the first cases developed symptoms. At that time, information related to the contaminated tap water was already communicated in the news, with a recommendation to rather drink bottled water.

### Strengths and limitations

The main strength of this study is that it is a proof of concept for a promising method to recruit controls for case-control studies. Panellists can be asked specific questions about particular products they ate, ways of cooking, places they shopped at etc., as soon as an outbreak occurs, enabling a quick investigation. The present study participates in the efforts to provide local public health professionals with innovative methods to empower them in the conduction of outbreak investigations [33]. The study also assesses likely bias of using panellists as controls in the context of outbreak investigations. It supports the findings of other studies that the risk of bias must be assessed anew when a study is conducted to investigate an outbreak [11]. Yet overall, the risk of bias when using the HuGO panel should be smaller in situations where the cases seem to resemble the general population (with a tendency of higher education because this is what the panel is composed of). Another strength of the study is the oversampling of the control group. This allows the selection of a subset of more valid controls to match the frequencies of cases in terms of sex and age, thereby addressing confounding and some of the bias introduced by using panellists as controls. Such an approach is here possible given that there is no additional cost per additional questionnaire.

The main limitation of the study concerns the historical studies. Panellists could only be used when population-based controls

were used in the historical studies, but not in specific settings (e.g. outbreaks during a party or at a restaurant). This strongly reduced the choice of historical studies for this analysis. The available studies had limited information on some key exposures and multivariable analyses could therefore not always be conducted. Additionally, the panel is affected by a selection bias. As this cannot be controlled via study design, it is therefore important to consider whether the measure of effect for a particular exposure may have been due to such an inherent sampling issue. In the situation of a prospective outbreak investigation, socio-demographic data could easily be collected among cases and panellists, and a weighting method could minimise this bias (e.g. frequency matching, propensity scores, quotas) [13].

### Perspectives

The study provides encouraging results and warrant further exploration to prove the validity of using panellists as controls in case-control studies. First, we will recruit more panellists (including children), in order to increase the sociodemographic diversity of the panel and its suitability to investigate the upcoming outbreaks. Then, as soon as a likely food or waterborne outbreak occurs in Lower Saxony, we will conduct two parallel case-control studies in which controls would be selected either through a traditional method or from the HuGO panel. The results of both studies will be compared. Propensity score matching will be used to reduce selection bias [34, 35].

Finally, in order to assess in which scenarios or for which hypothesised food and water exposures or behaviours using panellists as controls might be more appropriate than traditional controls, any selection bias introduced by using the panel need to be better understood. A prospect is to compare the reported food and water exposures and behaviours of panellists with that of probability samples from population-based food exposure surveys, such as with the German food exposure survey [36, 37] or other sources of such data for which selection biases are minimised or previously characterised [38].

### Conclusion

The study shows that using panellists as controls in case-control studies is feasible and suitable to investigate diffuse outbreaks within an adequate time frame, and researchers can benefit from increased speed of recruitment and limited costs. Nevertheless, the circumstances in which panellists are reliable to investigate food and waterborne outbreaks should be further investigated. We therefore recommend the further evaluation of this approach in parallel case-control studies and case-panel studies, especially in the context of food and waterborne outbreak investigations.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268821002594>.

**Acknowledgements.** We thank the HuGO panellists who contributed to the study. We thank the outbreak investigation team leaders for their insights on the studies they conducted: Hanna Oppermann for the Study C and Dirk Werber for the Study D. We thank Nicole Rübbsamen for her insights on the HaBIDS panel. We thank Frantiska Hruha for her insightful comments in the conduction of the study, and Sooria Balasegaram for her careful reading of the manuscript.

**Financial support.** This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Ethical standards.** The recruitment of participants for a panel was approved by the Ethics Committee of Hannover Medical School (No. 2021-2013) and by the Federal Commissioner for Data Protection and Freedom of Information in Germany. All participants provided informed consent before entering the HuGO panel.

**Conflicts of Interest.** None.

**Data availability statement.** Data supporting the findings of this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.5243412>, reference number [5243412].

## References

1. **Robert Koch Institute** (2018) 2018 Epidemiological Yearbook of nationally notifiable infectious diseases in Germany. Language: German, p. 236. Available at [https://www.rki.de/DE/Content/Infekt/Jahrbuch/Jahrbuch\\_2018.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/Infekt/Jahrbuch/Jahrbuch_2018.pdf?__blob=publicationFile).
2. **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. Available at <https://pubmed.ncbi.nlm.nih.gov/25762171/>.
3. **Rothman KJ, Greenland S and Lash TL** (2008) *Modern Epidemiology*, 3rd Edn. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, p. 758.
4. **Wacholder S et al.** (1992) Selection of controls in case-control studies. *American Journal of Epidemiology* **135**, 1029–1041. Available at <https://pubmed.ncbi.nlm.nih.gov/1595689/>.
5. **Buchholz U et al.** (2011) German Outbreak of *Escherichia coli* O104:H4 associated with sprouts. *New England Journal of Medicine* **365**, 1763–1770. Available at <http://www.nejm.org/doi/abs/10.1056/NEJMoa1106482>.
6. **Bandera EV et al.** (2013) Rethinking sources of representative controls for the conduct of case-control studies in minority populations. *BMC Medical Research Methodology* **13**, 71. Available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3681587/>.
7. **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. Available at <http://www.jmir.org/2014/12/e276/>.
8. **Rhodes SD** (2003) Collecting behavioural data using the world wide web: considerations for researchers. *Journal of Epidemiology & Community Health* **57**, 68–73. Available at <http://jech.bmj.com/cgi/doi/10.1136/jech.57.1.68>.
9. **Schonlau M** (2004) Will web surveys ever become part of mainstream research? *Journal of Medical Internet Research* **6**, e31–e31. Available at <https://pubmed.ncbi.nlm.nih.gov/15471757>.
10. **van Gelder MMHJ, Bretveld RW and Roeleveld N** (2010) Web-based questionnaires: the future in epidemiology? *American Journal of Epidemiology* **172**, 1292–1298. Available at <https://pubmed.ncbi.nlm.nih.gov/20880962/>.
11. **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 & Infection* **146**, 458–464. Available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5848756/>.
12. **Baker R et al.** (2010) Research synthesis: report on online panels. *Public Opinion Quarterly* **74**, 711–781. Available at <https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfq048>.
13. **Craig BM et al.** (2013) Comparison of US panel vendors for online surveys. *Journal of Medical Internet Research* **15**, e260. Available at <http://www.jmir.org/2013/11/e260/>.
14. **Hays RD, Liu H and Kapteyn A** (2015) Use of internet panels to conduct surveys. *Behavior Research Methods* **47**, 685–690. Available at <https://pubmed.ncbi.nlm.nih.gov/26170052/>.
15. **Gillespie IA et al.** (2010) *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* **7**, 1373–1379. Available at <http://www.liebertpub.com/doi/10.1089/fpd.2010.0568>.
16. **Gobin M et al.** (2018) National outbreak of shiga toxin-producing *Escherichia coli* O157:h7 linked to mixed salad leaves, United Kingdom, 2016. *Eurosurveillance* **23**. Available at <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2018.23.18.17-00197>.
17. **Kanagarajah S et al.** (2018) Whole genome sequencing reveals an outbreak of *Salmonella* Enteritidis associated with reptile feeder mice in the United Kingdom, 2012–2015. *Food Microbiology* **71**, 32–38. Available at <https://linkinghub.elsevier.com/retrieve/pii/S0740002016308796>.
18. **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 & Infection* **144**, 1220–1230. Available at <https://pubmed.ncbi.nlm.nih.gov/26493476/>.
19. **Scheil W et al.** (1998) A south Australian *Salmonella* Mbandaka outbreak investigation using a database to select controls. *Australian and New Zealand Journal of Public Health* **22**, 536–539. Available at <https://online-library.wiley.com/doi/abs/10.1111/j.1467-842X.1998.tb01434.x>.
20. **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 & Infection* **145**, 864–871. Available at <https://pubmed.ncbi.nlm.nih.gov/27964764/>.
21. **Yahata Y et al.** (2018) Web survey-based selection of controls for epidemiological analyses of a multi-prefectural outbreak of enterohaemorrhagic *Escherichia coli* O157 in Japan associated with consumption of self-grilled beef hanging tender. *Epidemiology & Infection* **146**, 450–457. Available at <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/web-surveybased-selection-of-controls-for-epidemiological-analyses-of-a-multiprefectural-outbreak-of-enterohaemorrhagic-escherichia-coli-o157-in-japan-associated-with-consumption-of-selfgrilled-beef-hanging-tender/8B0B212F14F72885C2BAE06822CFAAF3>.
22. **German Infectious Disease Prevention and Control Act (Infektionsschutzgesetz, IfSG)**. Language: German. Available at <https://www.gesetze-im-internet.de/ifsg/BJNR104510000.html>.
23. **Centers for disease control and prevention (CDC)** (2019) Rasmussen SA, Goodman RA (eds). *The CDC Field Epidemiology Manual*. Oxford, New York: Oxford University Press, 528 p. Available at <https://www.cdc.gov/eis/field-epi-manual/index.html>.
24. **Robert Koch Institute** (2021) Investigation of Outbreaks. Language: German. Available at [https://www.rki.de/DE/Content/Infekt/Ausbrueche/Ausbruchsuntersuchungen/Ausbruchsuntersuchungen\\_inhalt.html](https://www.rki.de/DE/Content/Infekt/Ausbrueche/Ausbruchsuntersuchungen/Ausbruchsuntersuchungen_inhalt.html).
25. **State Statistical Office of Lower Saxony** (2018) (Landesamt für Statistik Niedersachsen, LSN). Data of the 2018 Lower Saxony Population Microcensus. Language: German. Available at <https://www.statistik.niedersachsen.de/themenbereiche/haushalte/themenbereich-haushalte-und-familienmikrozensus-und-freiwillige-haushaltsbefragungen---tabellen-mikrozensus-150088.html>.
26. **Rübsamen N et al.** (2017) Comparison of response patterns in different survey designs: a longitudinal panel with mixed-mode and online-only design. *Emerging Themes in Epidemiology* **14**, 4. Available at <http://ete-online.biomedcentral.com/articles/10.1186/s12982-017-0058-2>.
27. **Rübsamen N et al.** (2017) Factors associated with attrition in a longitudinal online study: results from the HaBIDS panel. *BMC Medical Research Methodology* **17**, 132–132. Available at <https://www.ncbi.nlm.nih.gov/pubmed/28859617>.
28. **Grimes DA and Schulz KF** (2005) Compared to what? Finding controls for case-control studies. *The Lancet* **365**, 1429–1433. Available at <https://linkinghub.elsevier.com/retrieve/pii/S0140673605663799>.
29. **Richiardi L, Pizzi C and Pearce N** (2013) Commentary: representativeness is usually not necessary and often should be avoided. *International Journal of Epidemiology* **42**, 1018–1022. Available at <https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyt103>.
30. **Rothman KJ, Gallacher JE and Hatch EE** (2013) Why representativeness should be avoided. *International Journal of Epidemiology* **42**, 1012–1014. Available at <https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dys223>.
31. **Canada Government** (2017) Healthy Eating Assessment. Available at <https://www.hss.gov.nt.ca/sites/hss/files/resources/healthy-eating-assessment.pdf>.



32. **Inns T et al.** (2019) Are food exposures obtained through commercial market panels representative of the general population? Implications for Outbreak Investigations. *Epidemiology & Infection* **147**. Available at [https://www.cambridge.org/core/product/identifier/S0950268819000219/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0950268819000219/type/journal_article).
33. **Robert Koch Institute** (2020) Investigating Foodborne Outbreaks. Language: German. Available at [https://www.rki.de/DE/Content/Infekt/Ausbrueche/LM/lebensmittelbedingte\\_Ausbrueche\\_node.html](https://www.rki.de/DE/Content/Infekt/Ausbrueche/LM/lebensmittelbedingte_Ausbrueche_node.html).
34. **Natilli M, Romano MF and Leone L** (2008) Overcome Selection Bias in Web Surveys. An Application of Propensity Score Method When Target Differs From General Population. Available at [http://old.sis-statistica.org/files/pdf/atti/rs08\\_spontanee\\_3\\_5.pdf](http://old.sis-statistica.org/files/pdf/atti/rs08_spontanee_3_5.pdf).
35. **Randolph JJ et al.** (2014) A step-by-step guide to propensity score matching in R. *Practical Assessment, Research and Evaluation* **19**, 6. Available at <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.665.635&rep=rep1&type=pdf>.
36. **Heuer T et al.** (2015) Food consumption of adults in Germany: results of the German national nutrition survey II based on diet history interviews. *British Journal of Nutrition* **113**, 1603–1614. Available at <https://pubmed.ncbi.nlm.nih.gov/25866161/>.
37. **Max Rubner Institute** (2008) Results of the German National Consumption Study II. 196. 2008;196. Language: German. Available at [https://www.mri.bund.de/fileadmin/MRI/Institute/EV/NVS\\_II\\_Abschlussbericht\\_Teil\\_1\\_mit\\_Ergaenzungsbericht.pdf](https://www.mri.bund.de/fileadmin/MRI/Institute/EV/NVS_II_Abschlussbericht_Teil_1_mit_Ergaenzungsbericht.pdf).
38. **Rosner BM et al.** (2020) Population-based food consumption survey as an additional tool for foodborne outbreak investigations, Germany, 2017. *Epidemiology & Infection* **148**, e66. Available at [https://www.cambridge.org/core/product/identifier/S0950268820000564/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0950268820000564/type/journal_article).
39. **Robert Koch Institute** (2005) Epidemiological Bulletin 33/2005, Bacterial Gastroenteritis – Focus on Salmonellosis and Pork-Associated Outbreaks. Language: German, p. 33. Available at [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi99P-D7tLuAhVPr6QKHRP-CPUQFjAAegQIBRAC&url=https%3A%2F%2Fwww.rki.de%2FDE%2FCContent%2FInfekt%2FEpidBull%2FArchiv%2F2005%2FAusgabenlinks%2F33\\_05.pdf%3F\\_\\_blob%3DpublicationFile&usq=AOvVaw1NIPneAjdqGU6HTMeS7UGF](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi99P-D7tLuAhVPr6QKHRP-CPUQFjAAegQIBRAC&url=https%3A%2F%2Fwww.rki.de%2FDE%2FCContent%2FInfekt%2FEpidBull%2FArchiv%2F2005%2FAusgabenlinks%2F33_05.pdf%3F__blob%3DpublicationFile&usq=AOvVaw1NIPneAjdqGU6HTMeS7UGF).
40. **Werber D et al.** (2005) International outbreak of Salmonella Oranienburg due to German chocolate. *BMC Infectious Diseases* **5**, 7. Available at <https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-5-7>.