American Journal of Epidemiology

Vol. 176, No. 6
DOI: 10.1093/aje/kws034
Advance Access publication:
August 8, 2012

Practice of Epidemiology

The "Case-Chaos Study" as an Adjunct or Alternative to Conventional Case-Control Study Methodology

Iain A. Gillespie*, Piers Mook, Goutam K. Adak, Sarah J. O'Brien, and Noel D. McCarthy

* Correspondence to Dr. Iain A. Gillespie, Health Protection Agency, 61 Colindale Avenue, London NW9 5EQ, United Kingdom (e-mail: iain.gillespie@hpa.org.uk).

Initially submitted December 10, 2010; accepted for publication January 20, 2012.

Case-control studies are important in infectious disease epidemiology for rapidly identifying and controlling risks, but challenges, including the need for speed, can place practical restrictions on control selection and recruitment. The biased comparisons that result can hamper or, worse, mislead investigators. Following a 2009 outbreak of Shiga-like toxin-producing *Escherichia coli* O157 infection associated with a petting farm in southeast England, it was hypothesized that case behavior alone could be used to identify risks. Case-patients' exposures were randomized on a case-by-case basis, and the resulting permuted data were compared with the actual events preceding illness by conditional logistic regression analysis. There was good agreement between the risks identified by using our new method and the risks elicited in the original outbreak case-control studies. This was also the case in analysis of 2 further historical outbreaks. These initial findings suggest that the technique, which we have called the "case-chaos" technique, appeared to be useful in this setting. Analysis of simulated data supports this view. Circumventing the need for traditional control data has the potential to reduce outbreak investigation lead times, leading to earlier interventions and reduced morbidity and mortality. However, further validation is necessary, coupled with an awareness of limitations of the method.

case-control studies; disease outbreaks; infectious disease; random allocation; research design; risk

Abbreviations: CI, confidence interval; OR, odds ratio; STEC, Shiga-like toxin-producing *Escherichia coli*.

Case-control studies are often undertaken as an emergency response to outbreaks of infectious disease. Appropriate control selection is critical for unbiased comparisons (1–3), but outbreaks can be particularly challenging, not least because the public seems increasingly reluctant to participate. This limits the results further (4) and, hence, outbreak management and control.

To overcome the need for speed in an emergency, as well as low participation among reluctant controls, we propose a new case-only design for rapid assessment of risks in outbreaks. We tested this method using data from an outbreak of Shiga-like toxin-producing *Escherichia coli* (STEC) O157 infection associated with visiting a large petting farm in southeast England (5) and from 2 nationally distributed foodborne disease outbreaks—STEC O157 (6) and *Salmonella enterica* serovar Newport (*S.* Newport) (7), respectively. We present here the levels of agreement between our proposed "case-chaos" method and the traditional case-control studies that were conducted at the time.

MATERIALS AND METHODS

Underlying principles

The main aim of our method is to rapidly identify risks associated with becoming a case in an outbreak. We are interested in binary decisions that would be reliable enough to pinpoint as quickly as possible a potential transmission route or food vehicle(s) for further investigation and control.

Brief description of outbreaks selected for this study

In August and September 2009, 93 cases of STEC O157 phage-type 21/28 Shiga toxin 2 infection occurred among

visitors to a petting farm in southeast England (henceforth the "petting farm outbreak") (5). Twenty-three laboratory-confirmed primary cases and 112 controls were included in a case-control study to identify risk factors for disease. Visiting the large barn increased the likelihood of being a case (odds ratio (OR) = 3.46×10^8, 95% confidence interval (CI): 2.24×10^7, 5.32×10^9; $P < 0.001$), with the risk increasing for those who petted sheep in the barn (OR = 1.13×10^9, 95% CI: 8.45×10^7, 1.51×10^{10}; $P < 0.001$).

In July 2007, 2 contemporaneous cases of STEC O157 infection in northwest England reported consuming the same brand of lemon and coriander chicken wrap from a national supermarket chain (henceforth the "chicken wrap outbreak") (6). Twelve people were affected in total. An unmatched case-control study of 8 cases and 39 controls revealed an association between the consumption of lemon and coriander chicken wraps and being a case (OR = 46.40, 95% CI: 5.39, infinity; $P = 0.0002$).

In August 2004, 146 cases of *S*. Newport infections were reported across the United Kingdom (henceforth the "*S*. Newport outbreak"). Many cases had consumed salad items outside the home and, especially, from carryout/takeaway premises. In Northern Ireland, 23 cases were matched to 39 well controls who had eaten in the same takeaway establishment as the case (7). Eating lettuce was associated with being a case (OR = 23.7, 95% CI: 1.4, 404.3). Similarly in northeast England, lettuce was implicated in case-control studies focusing on 2 premises (matched OR = 11.43, 95% CI: 1.86, 70.27; $P = 0.009$, and OR = 12.8, 95% CI: 3.34, 49.12; $P < 0.001$, respectively) (8).

Data manipulation, generation, and analysis

All data sets were prepared for analysis by using Microsoft Excel 2007 (Microsoft Corporation, Redmond, Washington). First, control data were removed from the case-control data sets. Next, all exposure data were binary coded to represent "yes" and "no" responses, with missing data coded as such. Categorical exposures were dichotomized (any vs. no exposure) where possible or were removed.

For each outbreak, a copy of the case data set was created, and the outcome variable was recoded in the copy to represent the reference group (i.e., from 1 to 0 in Stata software; Stata Corporation, College Station, Texas) (Figure 1). The exposure data on each row were then randomized by using a macro (Web Appendix 1, the first of 2 Web appendixes, 3 Web tables, and 3 Web figures that appear on the *Journal*'s Web site (http://aje.oxfordjournals.org/)). Data were read into an array and returned to a random column position within the same row. The resulting "chaos" data set was combined with the original case data set, creating what is henceforth termed a "case-chaos" data set. The number of chaos data sets generated per outbreak depended on the number of cases available, with the process repeated until the traditional "one-to-many" ratio was obtained, until a study population of ≥40 was generated, or both of these criteria were fulfilled.

Statistical analysis, which was limited to univariate analysis (i.e., the study of the effect of 1 dependent variable on the outcome) for this initial study, was carried out in Stata versions 9 and 11. Exposure differences among "cases" (the original case-control cases) and "controls" (permuted case-chaos data sets) were assessed by using conditional logistic regression, with cases' unique identifiers used as the matching variable. The conditional odds ratios generated (with accompanying 95% confidence intervals and significance tests) were the odds of a case reporting exposure "A" compared with the odds of the case reporting the other exposures studied. Exact conditional logistic regression was used where conditional logistic regression could not provide estimates of effect. A full list of the covariates investigated is available in Web Tables 1–3.

To assess agreement between the 2 methods, arbitrary "risk factors" in each study were coded as 1, other exposures were coded as 0, and kappa statistics were calculated. Significance cutoffs of 95%, 99%, and 99.9% were investigated to test the robustness of the case-chaos technique to different choices of the arbitrary risk factor threshold. To assess agreement independent of dichotomization, the lower 95% confidence intervals generated by the 2 study types in each outbreak were compared by calculating intraclass correlation coefficients.

Simulation study

Simulations were undertaken to provide detailed information on how the case-chaos odds ratios relate to standard case-control odds ratios and to assess the performance of the case-chaos technique in a variety of scenarios. All simulations were programmed in Stata, versions 9 and 11, and the code is available (Web Appendix 2).

Cohorts of 5,000 people were simulated under a range of conditions, 100 people with the disease outcome and 4,900 without. Ten exposures were simulated each time. The binary exposure status of each individual was assigned randomly on the basis of the chosen probability of exposure for that simulation. The exposure levels simulated were 5%, 10%, and 25%. For the tenth, disease-associated exposure, exposure levels identified by using the Microsoft Excel Goal Seek function (Microsoft Corporation) were assigned to those with and without the disease with probabilities that would, in an infinitely large simulation, produce relative risks of 2, 5, and 10 and the overall level of exposure designated for the simulation.

Nine sets of simulations allowed assessment over the 3 fixed levels of exposure and 3 levels of relative risk. Three further simulations were undertaken with a relative risk of 5, with 3 variables each at exposure levels of 5%, 10%, and 25% and the tenth, disease-associated exposure, at either 5%, 10%, or 25%. For each set of values, 1,000 simulations were undertaken.

Relative risk was outputted and summarized to confirm that the simulation performed as intended. A matched case-control study was performed within each data set, randomly sampling 50 cases and 4 matched controls. A case-chaos analysis was performed by using the same 50 cases and 4 chaos data sets generated as described above.

To describe the transition from case-control to case-chaos odds ratios, logarithms of the conditional odds ratios generated by the 2 techniques were compared in terms of

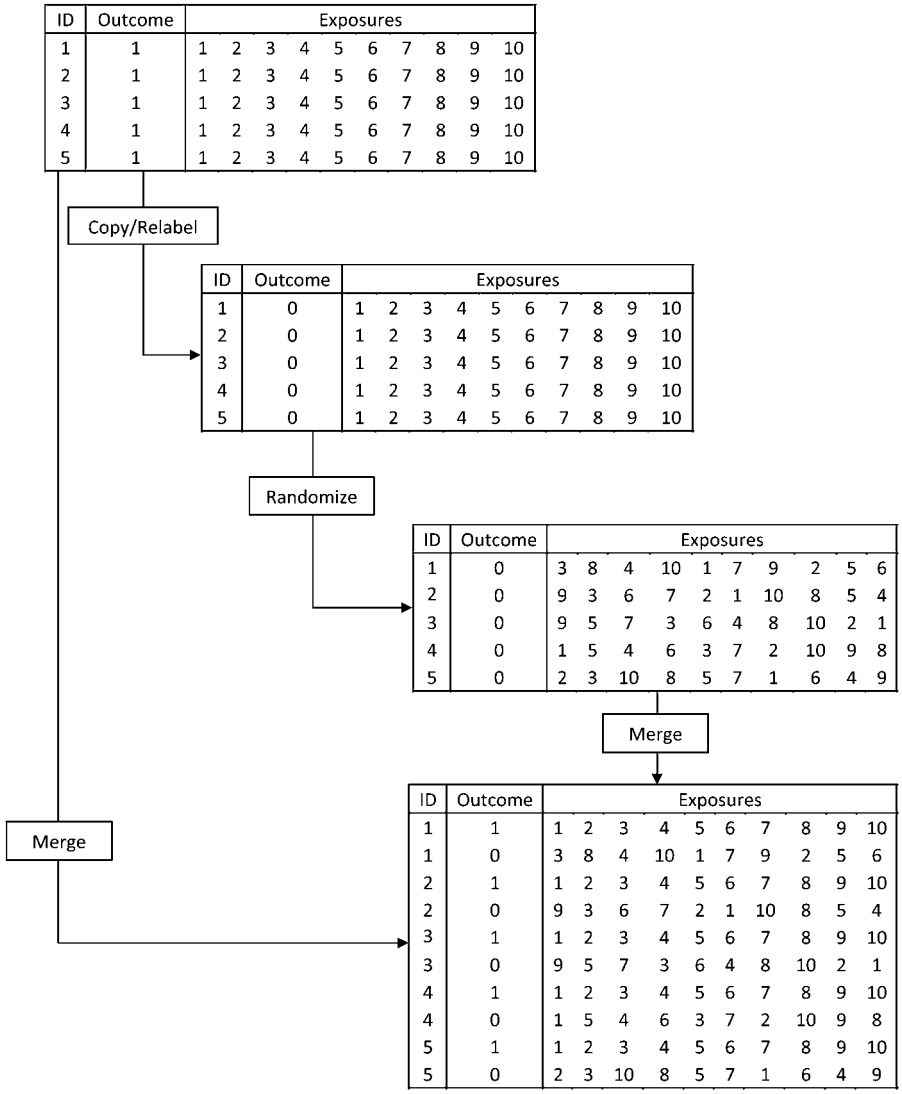

Figure 1. Derivation of the case-chaos data sets; numbers 1–10 for each individual's exposures indicate his/her response (yes/no/missing) for each exposure. ID, identifier.

their linear relation (graphically, with intraclass correlation) and in terms of the absolute difference between the 2 metrics (medians and interquartile ranges). To assess the performance of the case-chaos technique in relation to that of the case-control study, we calculated the associations with the outcome on the basis of arbitrary cutoffs of matched odds ratios > 1 and $P < 0.05$. The proportions of associations with the tenth, disease-associated variable were calculated for both study types. The number of false associations, defined as associations observed in the first 9 variables, was also calculated for the case-chaos studies.

RESULTS

Collectively and at the 95% significance level, there was moderate agreement between risk factors identified in the case-chaos studies and those in the original case-control studies (91.0% agreement; kappa = 0.60). Individually, agreement was moderate in the petting farm outbreak (Table 1), substantial in the chicken wrap outbreak (Table 2), and near-perfect in the *S.* Newport outbreak (Table 3). Results varied for significance levels of 99% and 99.9%, although smaller numbers were available for comparison in the latter 2 outbreaks (Tables 1–3). There was strong linear agreement between the lower 95% confidence intervals generated by both study designs for each outbreak (intraclass correlations: petting farm outbreak = 0.89; chicken wraps outbreak = 0.84; *S.* Newport outbreak = 0.998).

Simulation study results

Relative risks approximated to the predicted value in all simulations (Web Figure 1), and case-control odds ratios were well correlated to relative risks (Web Figure 2).

Table 1. Characteristics of, and Risk Factors Identified in, the Petting Farm Outbreak, 2009

Parameter	Results by Study Type						%	Kappa Value
	Case-Control[a]			Case-Chaos[b]				
	Odds Ratio	95% CI	*P* Value	Odds Ratio	95% CI	*P* Value		
Risk factors								
Prolonged visit (≥5 hours)	3.91	1.36, 11.25	0.012	4.37	1.17, 16.32	0.028		
<5 previous farm visits	30.85	5.96, 159.82	<0.001	23.72	3.11, 180.92	0.002		
Entered the farm via the top entrance	3.58	1.18, 10.86	0.024					
Exited the farm via the top entrance	3.94	1.13, 13.8	0.032					
Hand fed any animals	3.32	1.18, 9.35	0.023					
Visited the large barn	18.63	2.97, ∞	<0.001	33.76	5.68, ∞	<0.001		
Petted the adult goats in the large barn	4.39	1.56, 12.35	0.005	3.44	1.18, 10.05	0.024		
Petted the baby goats in the large barn	4.60	1.63, 12.95	0.004					
Petted the adult sheep in the large barn	6.24	2.13, 18.24	0.001	3.24	1.10, 9.55	0.033		
Visited the rabbit enclosure	25.41	3.12, 207.2	0.003	24.38	3.22, 184.44	0.002		
Visited the Shetland pony paddock				2.99	1.08, 8.33	0.036		
Visited the animal courtyard				5.42	1.51, 19.48	0.010		
Visited any other farm animals				12.74	1.58, 102.96	0.017		
After touching animals, washed hands with water				28.21	3.74, 212.85	0.001		
After touching animals, washed hands with soap and water				22.39	2.95, 170.22	0.003		
Ate any food				8.41	1.39, ∞	0.015		
Ate own food	13.27	1.63, 108.01	0.016	7.69	1.25, ∞	0.023		
Ate ice lolly[c]	3.68	1.13, 11.97	0.031					
Before eating food, washed hands with alcohol sanitizer	5.43	1.44, 20.55	0.013					
Had a picnic				8.41	1.39, ∞	0.015		
Agreement								
$P < 0.05$							85.4	0.43
$P < 0.01$							85.4	0.25
$P < 0.001$							93.3	0.22

Abbreviation: CI, confidence interval.
[a] Cases ($n = 23$); controls ($n = 112$).
[b] Cases ($n = 23$); controls ($n = 46$).
[c] Ice cream or water ice on a small wooden stick.

Median log odds ratios from case-control and case-chaos data sets were similar, with case-chaos studies generally producing marginally lower log odds ratios over a wider range (Web Figure 3). Within strata of relative risk, agreement between the 2 metrics increased slightly with increasing exposure prevalence. In the simulations with various exposure levels, differences between the 2 metrics were greater.

Generally, the case-chaos technique was less likely to identify the disease-associated exposure than the corresponding case-control study (Table 4). Where levels of population exposure were fixed, the case-chaos technique performed better when the underlying relative risk increased within each stratum of exposure. False associations decreased with increasing underlying relative risk and increasing underlying exposure prevalence. Where various levels of population exposure were simulated, the case-chaos performed well when the outcome-associated exposure was high but poorly when it was low. False associations were most prevalent where the background exposure prevalence was high.

DISCUSSION

We report a novel analytical approach to assessing risk in epidemiology. We have randomized patients' individual exposures, comparing the resulting randomized data with the actual events preceding their illness, and have switched the sampling frame for control selection from the population that gave rise to the cases to the cases themselves. When applied to data from 3 infectious disease outbreaks, the technique appeared to perform at least as well as conventional case-control studies in detecting those exposures

Table 2. Characteristics of, and Risk Factors Identified in, the Chicken Wrap Outbreak, 2007

Parameter	Results by Study Type						%	Kappa Value
	Case-Control[a]			Case-Chaos[b]				
	Odds Ratio	95% CI	*P* Value	Odds Ratio	95% CI	*P* Value		
Risk factors								
Food from supermarket chain P	32.51	2.52, 1,916.38	0.004					
Ate chicken cooked at home				14.35	1.83, 151.76	0.008		
Chicken sandwich eaten away from home	30.74	3.70, 445.73	<0.001	18.54	2.37, 250.5	0.003		
Yogurt eaten				13.4	1.43, 194.73	0.019		
Tomato eaten				14.61	1.95, 188.59	0.005		
Mains (municipal) water drunk				25.96	2.62, 1,364.44	0.002		
Normally shop at supermarket chain P	24.78	2.62, 1,260	0.001	78.55	6.43, 4,870.85	<0.001		
Supermarket chain P lemon and coriander chicken wrap	169.92	18.9, ∞	<0.001	54.75	4.92, 3,145.26	<0.001		
Agreement								
$P < 0.05$							96.3	0.75
$P < 0.01$							85.0	0.07
$P < 0.001$							97.5	0.49

Abbreviation: CI, confidence interval.
[a] Cases ($n = 8$); controls ($n = 39$).
[b] Cases ($n = 8$); controls ($n = 32$).

most plausibly contributing to infection. Analysis of simulated data supports this, as well as demonstrating those areas where the technique may perform less well.

Methodological considerations

The case-chaos approach allows the systematic and quantitative identification of those exposures among cases that are common compared with the average among the range of exposures evaluated and can be expected to identify those exposures that are common in the population of cases. Apart from chance findings, the 2 main reasons for this are either that the exposure is truly a risk for disease or that the exposure is common and reported to be so in the underlying population. In outbreaks of infectious diseases, where few exposures may explain a large number of cases,

Table 3. Characteristics of, and Risk Factors Identified in, the *Salmonella* Newport Outbreak, 2004

Parameter	Results by Study Type						%	Kappa Value
	Case-Control[a]			Case-Chaos[b]				
	Odds Ratio	95% CI	*P* Value	Odds Ratio	95% CI	*P* Value		
Risk factors								
Any chicken				8.74	1.92, 39.84	0.005		
Any salad	18.94	2.70, ∞	<0.01	9.31	2.07, 41.89	0.004		
Lettuce	15.76	1.89, 131.15	0.01	8.87	1.93, 40.74	0.005		
Tomato	16.95	2.47, ∞	<0.001	10.83	1.32, 88.7	0.026		
Mayonnaise	14.77	2.08, ∞	<0.001	3.84	1.16, 12.7	0.028		
Agreement								
$P < 0.05$							94.7	0.86
$P < 0.01$							79.0	0.21
$P < 0.001$							Not estimatable	

Abbreviation: CI, confidence interval.
[a] Cases ($n = 23$); controls ($n = 39$).
[b] Cases ($n = 23$); controls ($n = 46$).

Table 4. Identification of True Identified Associations (Odds Ratio > 1 and $P < 0.05$) Between Exposure and Outcome Within Simulated Case-Control and Case-Chaos Data Sets, as Well as False Associations Identified in Case-Chaos Analysis

Exposure Levels and Relative Risk	Proportion of Associations With Outcome					
	True Associations Identified				False Associations Identified	
	Case, no.	Control, no.	Case, no.	Chaos, no.	Case, no.	Chaos, no.
Fixed exposure						
5%						
RR = 2	370	1,000	217	1,000	177	9,000
RR = 5	998	1,000	867	1,000	70	9,000
RR = 10	1,000	1,000	995	1,000	25	9,000
10%						
RR = 2	603	1,000	321	1,000	143	9,000
RR = 5	1,000	1,000	966	1,000	63	9,000
RR = 10	997	1,000	996	1,000	30	9,000
25%						
RR = 2	846	1,000	511	1,000	140	9,000
RR = 5	1,000	1,000	994	1,000	68	9,000
RR = 10	1,000	1,000	1,000	1,000	34	9,000
Mixed exposure						
5%[a]						
RR = 5	998	1,000	247	1,000	0	3,000
10%						
RR = 5					14	3,000
25%						
RR = 5					1,394	3,000
5%						
RR = 5					0	3,000
10%[a]						
RR = 5	1,000	1,000	889	1,000	4	3,000
25%						
RR = 5					1,072	3,000
5%						
RR = 5					0	3,000
10%						
RR = 5					0	3,000
25%[a]						
RR = 5	1,000	1,000	1,000	1,000	577	3,000

Abbreviation: RR, relative risk.
[a] Disease-associated exposure at this level in the population.

these exposures may be identified by the case-chaos design. Our method is less likely to be successful for multifactorial diseases where individual risk factors make small contributions to outcome.

The matched odds ratios generated in these case-chaos studies thus provided a systematic approach to identifying exposures that were prominent among cases. However, the matched odds ratios generated in the case-chaos studies do not equate to causal parameters. This was not the point of the exercise. We were interested in binary decisions rather than estimating the magnitude of a causal effect.

In general, the risks identified by using the case-chaos method in the outbreaks that we examined showed good agreement with the risk factors identified in the case-control studies that had been performed. In the foodborne disease outbreaks, where a limited number of food exposures was linked to a large number of cases, the agreement between the methods was good. However, there was less agreement

between the 2 methods in the petting farm outbreak where contamination of the farm environment was widespread and diffuse. In this instance, cases might not have shared the same risk factors if infected at farm sites other than the main barn.

The results also demonstrate that some risks are particularly prone to, probably erroneous, identification using the case-case approach. The high rates of "ate chicken cooked at home" and "mains (municipal) water drunk" simply reflected very common exposures in the underlying population. These findings highlight the need for judgment in selecting exposures for analysis and interpreting the results. In some settings, a restricted range of exposures of interest will protect against this or at least guide interpretation. In others, this will rely more on the analyst. Studies assessing a range of rare and common exposures are problematic. This method appears to be less appropriate for the investigation of common exposures, but conventional case-control studies might also fail to identify such exposures correctly unless the sample size is sufficiently large, and this may be limited by the biology of the disease under investigation (e.g., when cases are exhausted during the hypothesis-generation phase of a point-source outbreak).

Outbreaks included in this study were chosen to test the robustness of the method in specific diverse scenarios. The petting barn outbreak was included because it was during this outbreak investigation that the case-chaos technique was conceived, when one of us (I. A. G.) noticed that, with the exception of the main barn, cases appeared to act randomly in terms of the attractions they visited. He hypothesized that this randomness could be used to investigate risk. The chicken wrap outbreak, although small, involved the investigations of a large number of exposures, increasing the potential for chance associations. Finally, the *S.* Newport outbreak was chosen as it presented methodological challenges at the time, because the population at risk (people consuming food from fast-food restaurants) was difficult to define accurately, making control selection problematic. It is possible, although unlikely, that these reasons, however valid, resulted in the selection of outbreaks more amenable to the case-chaos approach than others. Thus, the approach still needs to be evaluated on a wider range of outbreaks.

This initial study focused on binary variables. Categorical or continuous variables may represent a greater challenge, however, making the examination of dose dependence potentially problematic. In conventional analyses, the effects of categorical variables are assessed in relation to a reference group, and this reference group could be replicated across several binary variables prior to randomization. Thus, a categorical smoking variable could be split into a "no versus light smoker" variable and a "no versus heavy smoker" variable. Continuous variables are slightly more problematic, but in reality these are often categorized for analysis, in which case the above applies.

Finally, this approach could be affected by an unusual form of recall bias. Although it avoids recall bias between people and any systematic differential recall bias among cases and controls (9), it is susceptible to differential recall between exposures, with memorable exposures more easily recalled and, hence, identified than less memorable or even embarrassing exposures.

Advantages and disadvantages over alternative study designs

The most important potential benefits of the case-chaos approach for field epidemiologists are wider choice of study design, economy, and speed. Factors that drive decision-making should include the nature of the hypothesis to be tested and the availability of an appropriate sampling frame for controls. The "control" data generated in case-chaos studies eliminate the need for the latter, making it possible to assemble analytical epidemiologic evidence to inform action in situations where suitable controls, an alternative reference case series, or alternative estimates of population exposure are unavailable (10).

The use of case-nominated controls in case-control studies is well established. The method is invaluable in outbreaks where the population at risk is shown to be restricted, for example, to particular age or ethnic groups. It was applied successfully in an outbreak of *Salmonella anatum* infection affecting infants in 1996 linked to dried milk formula (11). However, in 2008 during a case-control study of *Salmonella* Typhimurium infection affecting infants (12), parents of affected infants were reluctant to provide contact details of suitable controls among the children of their family and friends. Their hesitancy stemmed from concerns over the disclosure of information on children to third parties. Data protection concerns among the public are having a negative impact on this method of control selection.

Random (or systematic) digit dialing telephone surveys can be used to select controls in outbreaks where the population at risk is not restricted to particular demographic groups. Advantageously, telephone numbering systems for land lines are geographically structured, making it possible to loosely align the geographic distribution of cases and controls. However, the validity of this approach is related to the representativeness of the population accessible by land lines compared with the community as a whole. Recent trends in telephone usage in the United States, which are likely to be mirrored elsewhere, highlight the rising number of households with wireless services only and identify demographic and health differences between households with and without access to land lines (13). If these trends continue, it will become increasingly difficult to find sampling frames to select suitable controls, necessitating alternative approaches. An advantage of the case-chaos design is in eliminating the inevitable delays in collecting and managing reference population data set(s).

Like case-case and case-crossover studies, an advantage of the case-chaos design is that all the "cases" and "controls" recruited will have been subject to similar selection procedures, reflecting the fact that cases in infectious disease surveillance systems are not randomly selected from the general population, so that comparisons with controls selected randomly from the population at risk require careful interpretation (14). The case-chaos method provides an advantage over the case-case design in that it does not require an external comparison group (e.g., a different subtype within a bacterial species). The method can be advantageous over the case-crossover design in

circumstances where contamination of a surrogate exposure is transient (9). This is because cases exposed to, for example, a contaminated foodstuff during their case exposure period will likely also be exposed to that product during their control exposure period—a time when the product did not actually pose a risk. A case-chaos study might be preferred to a case-crossover study for infections with long incubation periods, for example, as differential recall bias might increase for control exposure periods in the more distant past.

Overall, case-chaos studies should be less resource intensive and more rapid than alternative methods, as the delays accrued in collecting and managing reference population data sets are eliminated. For continuous source or propagated point-source outbreaks, the binary decisions afforded by the case-chaos design should expedite investigations, leading to earlier interventions and potential reductions in morbidity and mortality.

Future work

The potential benefits and limitations of the case-chaos technique described above are unlikely to be exhaustive so, while our initial findings are encouraging, further validation is necessary. For example, in a recent salmonellosis outbreak, a case-case study was performed alongside a case-control study and the results of each were reported (15). A similar approach for case-chaos studies could quickly generate information about the utility of the technique in various settings and for different outcomes, compared with classical analytical epidemiology approaches, or be tested against the confirmed risk factors from a combination of epidemiologic, microbiologic, and other investigations. Future work could also assess the appropriateness of the technique for assessing and controlling confounding through multivariate analysis (i.e., the simultaneous study of the effect of several dependent variables on the outcome).

Conclusions

Rapid execution of analytical epidemiology in infectious disease outbreaks is becoming increasingly challenging in technical and practical terms. We report a novel approach to assessing risk in epidemiology. Our initial findings suggest that the case-chaos technique appears to be as effective as conventional case-control studies in 3 diverse outbreak settings. By removing the need for traditional control data, the case-chaos technique has the potential to reduce outbreak investigation lead times, leading to earlier interventions and reduced morbidity and mortality. It should be considered as an adjunct to conventional methods because further validation is needed. Although the method has certain limitations, when used in conjunction with the results of supporting environmental and microbiologic investigations, it should provide a useful addition to the field epidemiologist's armamentarium.

ACKNOWLEDGMENTS

Author affiliations: Health Protection Services: Colindale, Health Protection Agency, London, United Kingdom (Iain A. Gillespie, Piers Mook, Goutam K. Adak); Health Protection Services: Local and Regional Services, Health Protection Agency, London, United Kingdom (Noel D. McCarthy); Department of Zoology, University of Oxford, Oxford, United Kingdom (Noel D. McCarthy); School of Translational Medicine, University of Manchester and Manchester Academic Health Science Centre, Manchester, United Kingdom (Iain A. Gillespie, Sarah J. O'Brien); and University of Liverpool Institute of Infection and Global Health and National Centre for Zoonosis Research, Liverpool, United Kingdom (Sarah J. O'Brien).

The authors would like to thank the investigators of the outbreaks included in this study. They are also grateful to an anonymous member of an Internet Excel forum for writing the macros used to randomize the data.

These findings were presented in part at the Health Protection Agency symposium, "Outbreak Investigation: New Frontiers," on February 25, 2011.

Conflict of interest: none declared.

REFERENCES

1. Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol*. 1992;135(9):1019–1028.
2. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol*. 1992;135(9):1029–1041.
3. Robins J, Pike M. The validity of case-control studies with nonrandom selection of controls. *Epidemiology*. 1990;1(4):273–284.
4. Tam CC, Higgins CD, Neal KR, et al. Chicken consumption and use of acid-suppressing medications as risk factors for *Campylobacter enteritis*, England. *Campylobacter* Case-Control Study Group. *Emerg Infect Dis*. 2009;15(9):1402–1408.
5. Ihekweazu C, Carroll K, Adak B, et al. Large outbreak of verocytotoxin-producing *Escherichia coli* O157 infection in visitors to a petting farm in South East England, 2009. *Epidemiol Infect*. 2012;140(8):1400–1413.
6. Whittaker PJ, Sopwith W, Quigley C, et al. A national outbreak of verotoxin-producing *Escherichia coli* O157 associated with consumption of lemon-and-coriander chicken wraps from a supermarket chain. *Epidemiol Infect*. 2009;137(3):375–382.
7. Irvine WN, Gillespie IA, Smyth FB, et al. Investigation of an outbreak of *Salmonella enterica* serovar Newport infection. Outbreak Control Team. *Epidemiol Infect*. 2009;137(10):1449–1456.
8. Health Protection Agency. Update—outbreak of *Salmonella* Newport infection in England, Scotland, and Northern Ireland: association with the consumption of lettuce. *Commun Dis Rep CDR Wkly*. 2004;14(41). London, United Kingdom: Public Health Laboratory Service Communicable Disease Surveillance Center; 2004. (http://www.hpa.org.uk/cdr/archives/2004/cdr4104.pdf). (Accessed December 10, 2010).
9. McCarthy N, Giesecke J. Case-case comparisons to study causation of common infectious diseases. *Int J Epidemiol*. 1999;28(4):764–768.
10. Ethelberg S, Smith B, Torpdahl M, et al. Outbreak of non-O157 Shiga toxin-producing *Escherichia coli* infection from

consumption of beef sausage. *Clin Infect Dis*. 2009;48(8):e78–e81.
11. *Salmonella anatum* infection in infants linked to dried milk. *Commun Dis Rep CDR Wkly*. 1997;7(5):33, 36. London, United Kingdom: Public Health Laboratory Service Communicable Disease Surveillance Center; 1997. (http://www.hpa.org.uk/cdr/archives/1997/cdr0597.pdf). (Accessed December 10, 2010).
12. Harker KS, Lane C, De Pinna E, et al. An outbreak of *Salmonella* Typhimurium DT191a associated with reptile feeder mice. *Epidemiol Infect*. 2011;139(8):1254–1261.
13. Blumberg SJ, Luke JV. Wireless substitution: early release of estimates from the National Health Interview Survey, July-December 2007. Hyattsville, MD: National Center for Health Statistics; 2008. (http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200805.pdf). (Accessed December 10, 2010).
14. Tam CC, Rodrigues LC, O'Brien SJ. The study of infectious intestinal disease in England: what risk factors for presentation to general practice tell us about potential for selection bias in case-control studies of reported cases of diarrhoea. *Int J Epidemiol*. 2003;32(1):99–105.
15. Krumkamp R, Reintjes R, Dirksen-Fischer M. Case-case study of a *Salmonella* outbreak: an epidemiologic method to analyse surveillance data. *Int J Hyg Environ Health*. 2008;211(1-2):163–167.