American Journal of Epidemiology Advance Access published July 2, 2014

American Journal of Epidemiology

DOI: 10.1093/aje/kwu123

Practice of Epidemiology

Is Case-Chaos Methodology an Appropriate Alternative to Conventional Case-Control Studies for Investigating Outbreaks?

Michael Edelstein*, Anders Wallensten, and Sharon Kühlmann-Berenzon

* Correspondence to Dr. Michael Edelstein, The Public Health Agency of Sweden, Department of Surveillance and Evaluation, Nobels väg 18, 171 82 Solna, Sweden (e-mail: Michael.edelstein@folkhalsomyndigheten.se).

Initially submitted October 17, 2013; accepted for publication March 21, 2014.

Case-chaos methodology is a proposed alternative to case-control studies that simulates controls by randomly reshuffling the exposures of cases. We evaluated the method using data on outbreaks in Sweden. We identified 5 case-control studies from foodborne illness outbreaks that occurred between 2005 and 2012. Using case-chaos methodology, we calculated odds ratios 1,000 times for each exposure. We used the median as the point estimate and the 2.5th and 97.5th percentiles as the confidence interval. We compared case-chaos matched odds ratios with their respective case-control odds ratios in terms of statistical significance. Using Spearman's correlation, we estimated the correlation between matched odds ratios and the proportion of cases exposed to each exposure and quantified the relationship between the 2 using a normal linear mixed model. Each case-control study identified an outbreak vehicle (odds ratios = 4.9–45). Case-chaos methodology identified the outbreak vehicle 3 out of 5 times. It identified significant associations in 22 of 113 exposures that were not associated with outcome and 5 of 18 exposures that were significantly associated with outcome. Log matched odds ratios correlated with their respective proportion of cases exposed (Spearman $\rho = 0.91$) and increased significantly with the proportion of cases exposed ($b = 0.054$). Case-chaos methodology missed the outbreak source 2 of 5 times and identified spurious associations between a number of exposures and outcome. Measures of association correlated with the proportion of cases exposed. We recommended against using case-chaos analysis during outbreak investigations.

case-control studies; disease outbreaks; epidemiology; methods

Abbreviations: CI, confidence interval; OR, odds ratio.

Editor's note: *An invited response appears on page 000.*

Case-control analysis is the method most commonly used to conduct analytical investigations during outbreaks that are not confined to an easily identified closed population group or that have a low incidence of disease (1). The method relies on an unbiased control selection. Achieving a high response rate in case-control studies is an increasingly difficult task (2), as participation of controls in general population-based, case-control studies has been declining since 1990 (3, 4). In 2012, Gillespie et al. (5) proposed a study design called the case-chaos method that bypasses the need to recruit controls. The case-chaos method simulates controls by reassigning cases as controls and randomly permuting their exposures (5). Avoiding recruitment of controls has many advantages, including lower cost, shorter duration, and a potential reduction in recruitment bias. Gillespie et al. tested the methodology on simulated data and data from 3 real-life outbreaks, calculating case-chaos odds ratios from single permutations. They concluded that the technique appeared to be useful as an adjunct or alternative to case-control methodology when investigating outbreaks (5) but also pointed out that further validation with real data was required. A response to the original article challenged the theoretical validity of the method, arguing that the simulated proportion of controls exposed to the studied exposure has no biological or epidemiologic plausibility (6) and that when the case-chaos methodology was applied, the exposure with the highest proportion of cases exposed would also have the highest odds ratio (6).

Another response argued that by assessing whether a factor is significantly more common in cases than are other factors that are considered as potential risk factors, case-chaos methodology does not enhance interpretation of the frequency of exposure (7). This response also evaluated the case-chaos methodology by introducing 2 hypothetical exposures to real-life data from a foodborne illness outbreak. The first, a common but randomly distributed factor, was significantly associated with the outcome, whereas the second, a rare but absolute risk factor for illness, was not (7).

The authors reiterated that the method should be tested using real-life data (8, 9). To determine whether the case-chaos method can be used as an adjunct or alternative to case-control methods in real outbreak investigations, we evaluated the validity of the case-chaos method using data from real outbreaks previously investigated by the Swedish Institute for Communicable Disease Control.

METHODS

Data collection

We identified community-acquired foodborne illness outbreaks among outbreaks investigated in Sweden by the Swedish Institute for Communicable Disease Control between 2005 and 2012. We selected the 5 most recent outbreaks that were initially analyzed using the case-control design for which data were available.

Statistical analysis

We applied the case-chaos methodology to each of the selected outbreaks, following the procedure described in the original article (5). We created 5 controls for each case through random permutation of the cases' exposures. We matched each case to its own simulated controls and calculated unadjusted matched odds ratios using conditional logistic regression. To generate more robust estimates than in the original articles, we generated 1,000 case-chaos data sets for each outbreak rather than using a single permutation. For each exposure, we used the median of the 1,000 generated matched odds ratios as the crude point estimate and the 2.5th and 97.5th percentiles as the 95% confidence interval. We considered a case-chaos matched odds ratio to be statistically significant if its 95% confidence interval excluded 1. We compared the association between exposure and outcome obtained from the case-chaos and case-control methods in terms of statistical significance.

We used Spearman's correlation to determine whether the case-chaos matched odds ratio was correlated to the proportion of cases exposed. Next, we fitted a normal linear mixed model with log matched odds ratios as the outcome variable. Explanatory variables included the proportion of cases exposed and the number of cases in the corresponding outbreak. We included the outbreak as a random effect to correct for the dependency within matched odds ratio from the same outbreak. We evaluated the validity of the model by checking the assumption of normal distribution of residuals with a quantile-quantile plot and graphically confirmed the random distribution of the residuals with a scatterplot of residuals against the proportion of cases exposed.

We generated the case-chaos data sets and analyzed them using Stata, version 12 (StataCorp LP, College Station, Texas). We applied a significance level of 5% throughout the study. The original investigations were exempt from ethical committee review because they constituted an emergency response, and the data sets used in the study contained no patient identifiable information. Hence, this study did not require clearance by an ethical committee.

RESULTS

The Swedish Institute for Communicable Disease Control conducted the 5 selected case-control studies between 2005 and 2012. The agent was *Salmonella* in 4 studies and *Escherichia coli* in one. Reports were published for 2 of the 5 outbreaks (10, 11). The case definitions used in the outbreaks were based on both clinical and laboratory criteria. In the 5 outbreaks, cases and controls had answered questionnaires that collected information regarding food items consumed in the previous 1–2 weeks. The 5 case-control studies had a matched design and used conditional logistic regression to estimate matched odds ratios. Two analyses, however, broke the match. Microbiological evidence was available in 2 studies (10, 11). The number of cases ranged from 12 to 93, with the proportion of cases exposed to specific exposures ranging from 3% to 97%. Table 1 summarizes the characteristics of the selected outbreaks, including number of cases, number of recorded exposures, and proportion of cases exposed to specific items.

We re-examined 131 exposures using case-chaos methodology (Web Table 1, available at http://aje.oxfordjournals.org/). In 3 of the 5 outbreaks, the case-chaos method identified a statistically significant association between the outcome and the outbreak vehicle identified by the case-control study (Table 1). However, the case-chaos analysis also identified significant associations for 22 of 113 (19%) exposures that were not associated with the outcome in the case-control analyses (Table 2). In outbreak A, the case-control and the case-chaos methods both identified the association between being a case and eating iceberg lettuce (odds ratio (OR) = 13, 95% confidence interval (CI): 2.9, 57, and matched OR = 7.5, 95% CI: 1.9, 29, respectively; Table 2). However, the case-chaos method also identified stronger significant associations between being a case and eating 2 food items that were not suspected as sources (for hard cheese, matched OR = 11, 95% CI: 3.4, 42; for tomatoes, matched OR = 38, 95% CI: 12, 110; Table 2). Neither hard cheese consumption nor tomato consumption was associated with being a case in the original case-control studies (Table 2). Overall, 95% and 97% of cases had consumed to these 2 items, respectively, whereas only 92% of cases had eaten iceberg lettuce.

In addition, the case-chaos methodology failed to identify a significant association for 13 of the 18 (72%) exposures significantly associated with the outcome in case-control studies (Table 2). In outbreak C, 20% of cases reported eating baby spinach, and in outbreak D, 49% of cases reported eating alfalfa sprouts. Although both exposures were associated with outcome in the case-control studies, case-chaos methodology failed to detect an association with either of them (Table 1).

Table 1. Characteristics of Selected Outbreaks and Comparison of Matched Odds Ratios by Case-Control and Case-Chaos Methods of Analysis for the Suspected Outbreak Vehicles, Sweden, 2005–2012

Study	Outbreak	Year	Agent	No. of Exposures in the Case-Control Study	Number of Recruited Individuals in the Study		Suspected Outbreak Vehicle	Exposure to Suspected Outbreak Vehicle in Case Control Study[a]			Case-Control		Case-Chaos	
					Cases	Controls		No. Exposed	Total No.	%	mOR	95% CI	mOR	95% CI
Soderstrom, 2008 (10)[b]	A	2005	*Escherichia coli* O157 (VTEC)	10	67	100	Iceberg lettuce	59	64	92	13	2.9, 57	7.5	1.9, 29
Ledet Muller, 2007 (11)[c]	B	2006	*Salmonella enteritidis*	10	12	34	Almonds	10	12	83	45[d]	4.8, 420	17	4.1, ∞
G. Rydevik, Swedish Institute for Communicable Disease Control, unpublished data, 2007	C	2007	*Salmonella* Java	16	93	126	Baby spinach	10	50	20	4.9	1.7, 14	1.4	0.53, 3.6
G. Rydevik, Swedish Institute for Communicable Disease Control, unpublished data, 2007	D	2007	*Salmonella* Stanley	47	41	62	Alfalfa sprouts	19	39	49	18	4.5, 100	1.1	0.51, 2.8
M. Rehn, Swedish Institute for Communicable Disease Control, unpublished data, 2013[e]	E	2012	*Salmonella* Typhimurium	48	20	34	Salad	15	17	88	16[d]	2.7, 160	26	8.7, ∞

Abbreviations: CI, confidence interval; mOR, matched odds ratio.

[a] The denominator differs from the number of cases recruited because not all cases answered questions pertaining to each food item.

[b] The same *Escherichia coli* strain was identified in cases, lettuce irrigation water, and cattle at a farm upstream of the irrigation point.

[c] The *Salmonella* phage type was the same as the phage type in a previous almond-related *Salmonella* outbreak in the United States.

[d] Unmatched odd ratio.

[e] No salad specimens were tested as part of the outbreak investigation.

Table 2. Exposures Associated With Being a Case in Either a Case-Control or Case-Chaos Study From 5 Swedish Foodborne Illness Outbreaks, 2005–2012

Exposure by Outbreak	Cases			Case-Control Results		Case-Chaos Results	
	No. Exposed	Total No.	%	OR	95% CI	OR	95% CI
A							
Alfalfa sprouts	9	62	15	1.5	1.1, 16	0.11	0.03, 0.30
Ice cream	38	55	69	2.4	1.1, 5.2	1.5	0.34, 4.6
Iceberg lettuce	59	64	92	13	2.9, 57	7.5	1.9, 29
Hard cheese	54	57	95	2.5	0.70, 8.9	11	3.4, 42
Tomatoes	62	64	97	3.4	0.67, 17	38	12, 110
B							
Frankfurter sausages	8	12	67	1.6	0.30, 9.1	7.5	1.9, ∞
Almonds	10	12	83	45	4.8, 420	17	4.1, ∞
Hard cheese	11	12	92	0.80	0.10, 8.6	13	4.2, ∞
C							
Fresh baby spinach	10	50	20	4.9	1.7, 14	1.4	0.53, 3.6
Smoked fish	17	81	21	3.7	1.3, 11	0.36	0.16, 0.70
Omelette or fried egg	21	85	25	3.0	1.2, 7.6	0.46	0.19, 0.84
Spinach	18	68	26	4.4	1.6, 12	0.50	0.19, 0.97
Other salad	23	78	29	7.2	2.4, 21	0.57	0.23, 1.0
Hot dog	32	87	37	2.5	1.0, 6.1	0.84	0.37, 1.6
Dill	10	22	45	3.1	1.1, 9.2	1.2	0.45, 4.0
Any spinach	41	90	46	4.7	2.2, 10	1.3	0.49, 3.1
Berries	41	87	47	3.0	1.3, 6.8	1.3	0.55, 2.6
Restaurant	60	84	71	2.3	1.0, 5.2	4.0	1.8, 7.6
Any mentioned salad type	69	90	77	3.4	1.4, 8.0	8.5	2.7, ∞
D							
Other food outside home	14	31	45	4.6	1.5, 15	1.1	0.44, 2.5
Alfalfa sprouts	19	39	49	18	4.5, 100	1.1	0.51, 2.8
Ate in restaurant	25	38	66	1.3	0.53, 3.4	2.5	1.0, 6.4
Any chicken	29	40	73	1.5	0.59, 4.1	3.5	1.4, 9.7
Yogurt	27	37	73	2.2	0.85, 6.0	3.3	1.4, 8.2
Minced meat	30	41	73	1.0	0.39, 2.8	3.5	1.4, 9.3
Confectionery	29	37	78	0.96	0.32, 3.0	4.6	1.7, 13
Any sprouts	33	41	80	7.5	2.7, 22	5.9	2.2, 19
Shopping at a specific supermarket	36	41	88	1.5	0.45, 6.2	8.5	3.7, 23
Iceberg salad	36	40	90	3.4	0.98, 15	12	4.7, 34
Tomato	39	41	95	2.1	0.35, 22	24	10, ∞
E							
Juice	9	17	53	1.4	0.38, 5.6	2.9	1.1, 9.8
Chocolate	8	15	53	0.43	0.10, 1.9	2.8	1.0, 14
Other restaurant	11	19	58	1.9	0.52, 7.1	3.8	1.2, 14
Carrot	10	17	59	0.86	0.22, 3.4	4.0	1.3, 17
Mixed minced meat	12	18	67	3.6	0.84, 16	5.6	1.9, 32
White pepper	11	16	69	0.53	0.11, 2.7	5.3	1.7, 31
Peppers	13	18	72	1.7	0.42, 7.8	7.3	2.4, 36
Cucumber	17	20	85	1.4	0.25, 9.5	26	7.0, ∞
Black pepper	14	16	88	1.0	0.12, 12.3	20	7.0, ∞
Prepackaged mixed salad	15	17	88	16	2.7, 160	26	8.7, ∞

Abbreviations: CI, confidence interval; OR, odds ratio.

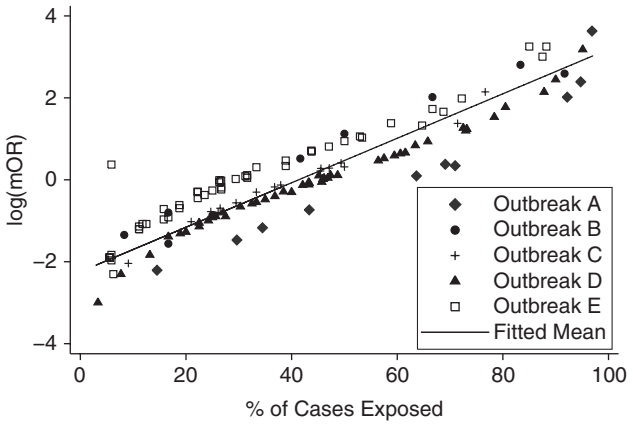

Figure 1. Relationship between percentage of cases exposed and log matched odds ratio (mOR) for each exposure in case-chaos studies from 5 Swedish foodborne illness outbreaks, 2005–2012. The solid line represents the mean log matched odds ratio given the percentage of cases exposed and mean ($n = 34$) number of cases.

The log matched odds ratios correlated with the respective proportion of cases exposed overall (Spearman $r = 0.92$, $P < 0.001$; Figure 1), as well as in each individual outbreak (r ranged from 0.93 to 1; $P < 0.001$ for all studies). In the model that was adjusted for outbreak, the log matched odds ratios increased significantly with the proportion of cases exposed ($b = 0.054$, $P < 0.001$) but was not associated with the number of cases ($b = -0.002$, $P = 0.35$). The variance between the outbreaks accounted for 24% of the total. The residuals from the model were normally distributed. No pattern could be identified when the residuals were plotted against the percentage of cases exposed.

DISCUSSION

Compared with case-control studies, case-chaos methodology failed to identify the outbreak vehicle in 2 out of 5 of outbreaks. Further, it identified associations that were not detected in the case-control studies. In fact, the matched odds ratios estimated from the case-chaos methodology strongly correlated with the corresponding proportion of cases exposed, regardless of sample size or outbreak. These empirical findings are consistent with proposed theoretical arguments (6, 7).

Although the original article on case-chaos methodology (5) included only a single iteration of controls, we used 1,000 iterations of Monte Carlo simulations to generate a more robust point estimate. Furthermore, pooling data from 5 outbreaks allowed us to document that the relation between the matched odds ratio and the proportion of cases exposed was systematic, predictable, and independent of the outbreak or the number of cases.

Using the case-chaos method in the analytical epidemiology stage of an outbreak investigation presents 2 inherent difficulties. First, because the proportion of cases exposed to one item is compared with the average proportion of cases exposed among all investigated items (6), items with high proportion of exposure will be associated with the outcome. Specific exposures will be strongly associated with the outcome because they are common exposures rather than because of a causal relationship. Second, case-chaos methodology will fail to detect associations between outcome and exposures for which a low proportion of cases was exposed. This has been shown both in our study using real-life data and when using hypothetical risk factors (7). Had case-chaos methodology been used in the 2011 *Escherichia coli* O104 outbreak that occurred in Germany in 2011 (12), it is unlikely that a source would have been identified because only 25% of cases reported eating sprouts (12). Furthermore, the limitations highlighted in the present study are inherent to the case-chaos method rather than specific to applying it to outbreak investigations. These findings are therefore likely to be similar regardless of the type of data used, as highlighted in an evaluation of the case-chaos method using data from an adverse drug reaction case-control study (13).

In the present study, we used case-control studies as a gold standard, assuming that the associations they identified were not caused by chance or bias. Case-control studies are susceptible to bias (14), and of the 5 outbreaks, only 1 had a source that was microbiologically confirmed. Although we could not verify whether the case-control studies correctly identified the sources in the other 4 outbreaks, we compared case-chaos methodology to the standard methodology when facing these types of open population outbreaks.

In conclusion, the case-chaos method is not an alternative to case-control investigation, as it cannot reliably detect associations between exposure and outcome. The odds ratios produced by the method are only correlated with the proportion of cases exposed. Hence, the case-chaos method identifies spurious associations when many cases have been exposed and fails to identify associations between exposure and outcome when few cases have been exposed.

We cannot recommend the use of case-chaos methodology as an adjunct or alternative to case-control studies. Other strategies are available to overcome the difficulty of recruiting controls, including testing whether the proportion of exposure differs from what was known before using binomial probabilities (15). This method has been successfully used during outbreak investigations (16, 17), has a stronger theoretical basis, and should be further investigated as an alternative or adjunct to case-control studies.

ACKNOWLEDGMENTS

Author affiliations: Department of Surveillance and Evaluation, Public Health Agency of Sweden, Solna, Sweden (Michael Edelstein, Anders Wallensten, Sharon Kühlmann-Berenzon); and European Program in Intervention Epidemiology Training (EPIET), European Centre for Disease Prevention and Control, Stockholm, Sweden (Michael Edelstein).

Conflict of interest: none declared.

REFERENCES

1. *World Health Organization*. Foodborne disease outbreaks: guidelines for investigation and control. http://whqlibdoc.who.

int/publications/2008/9789241547222_eng.pdf. Published 2008. Accessed August 9, 2013.
2. Kogevinas M, Chatzi L. Cross Sectional Studies. In: Detels R, Beaglehole R, Lansang MA, et al eds. *Oxford Textbook of Public Health*. Oxford, UK: Oxford University Press; 2009:509–528.
3. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol*. 2006;163(3):197–203.
4. Woodward A, Cook A. Low response rates among controls in case-control studies: possible causes and potential remedies. *Australas Epidemiol*. 2003;10(2):29–33.
5. Gillespie IA, Mook P, Adak GK, et al. The "case-chaos study" as an adjunct or alternative to conventional case-control study methodology. *Am J Epidemiol*. 2012;176(6):497–505.
6. Hohle M. Re: "the 'case-chaos study' as an adjunct or alternative to conventional case-control study methodology" [letter]. *Am J Epidemiol*. 2013;177(9):1021–1022.
7. Pulliam JR, Dushoff J. Re: "the 'case-chaos study' as an adjunct or alternative to conventional case-control study methodology" [letter]. *Am J Epidemiol*. 2014;179(2):261–262.
8. McCarthy ND, Gillespie IA, O'Brien SJ. Three authors reply [letter]. *Am J Epidemiol*. 2013;177(9):1022.
9. Gillespie IA, Mook P, Adak GK, et al. The authors reply [letter]. *Am J Epidemiol*. 2014;179(2):262–263.
10. Söderström A, Osterberg P, Lindqvist A, et al. A large *Escherichia coli* O157 outbreak in Sweden associated with locally produced lettuce. *Foodborne Pathog Dis*. 2008;5(3): 339–349.
11. Ledet Müller L, Hjertqvist M, Payne L, et al. Cluster of *Salmonella enteritidis* in Sweden 2005–2006—suspected source: almonds. *Euro Surveill*. 2007;12(6):E9–E10.
12. Buchholz U, Bernard H, Werber D, et al. German outbreak of *Escherichia coli* O104:H4 associated with sprouts. *N Engl J Med*. 2011;365(19):1763–1770.
13. Doerken S, Mockenhaupt M, Schumacher M, et al. Assessment of the "case-chaos" design as an adjunct to the case-control design. *Am J Epidemiol*. 2014;179(6):775–780.
14. Kopec JA, Esdaile JM. Bias in case-control studies. A review. *J Epidemiol Community Health*. 1990;44(3):179–186.
15. Keene B. The use of binomial probabilities in outbreak investigations. Presented at Association of Public Health Laboratories 7th Annual OutbreakNet Conference, Long Beach, California, September 19–22, 2011. http://www.aphl.org/conferences/proceedings/Documents/2011/2011_PulseNetOutbreakNet/017-Kenne.pdf. Accessed June 30, 2013.
16. Keene WE, Hedberg K, Herriott DE, et al. A prolonged outbreak of Escherichia coli O157:H7 infections caused by commercially distributed raw milk. *J Infect Dis*. 1997;176(3): 815–818.
17. Soborg B, Lassen SG, Müller L, et al. A verocytotoxin-producing *E. coli* outbreak with a surprisingly high risk of haemolytic uraemic syndrome, Denmark, September-October 2012. *Euro Surveill*. 2013;18(2).