HHS Public Access

Author manuscript

Epidemiology. Author manuscript; available in PMC 2023 May 22.

Published in final edited form as:

Epidemiology. 2022 September 01; 33(5): 633–641. doi:10.1097/EDE.0000000000001504.

Nearest-Neighbors Matching for Case–Control Study Analyses:

Better Risk Factor Identification From a Study of Sporadic Campylobacteriosis in the United States

Zhaohui Cui,

Ellyn P. Marder,

Eleanor S. Click,

Robert M. Hoekstra,

Beau B. Bruce

Enteric Diseases Epidemiology Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract

Background: Case–control studies are commonly used to explore factors associated with enteric bacterial diseases. Control of confounding is challenging due to a large number of exposures of interest and the low frequencies of many of them.

Methods: We evaluated nearest-neighbors matching in a case–control study (originally 1:1 matched, published in 2004) of sporadic *Campylobacter* infections that included information on 433 exposures in 2632 subjects during 1998–1999. We performed multiple imputations of missing data (m = 100) and calculated Gower distances between cases and controls using all possible confounders for each exposure in each dataset. We matched each case with ≤20 controls within a data-determined distance. We calculated odds ratios and population attributable fractions (PAFs).

Results: Examination of pairwise correlation between exposures found very strong associations for 1046 pairs of exposures. More than 100 exposures were associated with campylobacteriosis, including nearly all risk factors identified using the previously published approach that included only 16 exposures and some less studied, rare exposures such as consumption of chicken liver and raw clams. Consumption of chicken and nonpoultry meat had the highest PAFs (62% and 59%, respectively).

Conclusions: Nearest-neighbors matching appear to provide an improved ability to examine rare exposures and better control for numerous highly associated confounders.

Correspondence: Beau B. Bruce, MD, PhD, Enteric Diseases Epidemiology Branch, Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road NE, MS H24-9, Atlanta, GA 30329-4018. lue7@cdc.gov.

Disclosure: The authors report no conflicts of interest.

Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Keywords

Campylobacter; Case–control studies; Nearest-neighbors; Risk factors

Campylobacter is estimated to be the most common cause of bacterial gastroenteritis in the United States; the vast majority of cases are sporadic.[1] Case–control studies are commonly used to explore factors associated with campylobacteriosis and to estimate population attributable fractions (PAFs) for exposures.[2] *Campylobacter* colonizes the intestines of a wide range of wild and domesticated warm-blooded animals. The feces of those animals can contaminate foods and the environment. Furthermore, contamination of foods can occur at multiple points from production to consumption. Consequently, studies often evaluate many exposures, including rare ones.[3,4] The need to control for many exposures while evaluating rare ones requires a sizable sample to achieve stable regression models and sufficient statistical power. However, a recent systematic review found that most studies have relatively small sample sizes,[2] and so investigators have typically reduced the number of exposures included in regression models by screening exposures through univariate analysis,[3,5–7] combining multiple exposures into a single one,[3,6] or selecting one exposure among correlated ones.[6] Even so, the number of exposures included in regression models can exceed 100.[3] Propensity score matching has been used as an alternative approach but does not perform well as a distance measure in case–control matching.[8]

The nearest-neighbors matching approach has been used in a variety of fields, including ecology[9] and biology.[10] In epidemiology, this approach has been used in cross-sectional and quasi-experimental studies,[11,12] but not commonly for case–control studies. We hypothesized that nearest-neighbors matching could improve case–control matching and covariate adjustment compared with other methods. Our approach calculates a selected distance metric between cases and controls using all possible exposures except the one under active consideration, and then matches each case with its nearest neighbors to achieve better comparability across all measured covariates between controls and cases. We reanalyzed a case–control study (originally 1:1 matched) of sporadic *Campylobacter* infections in the United States conducted during 1998–1999, to permit the comparison of the nearest-neighbors matching to a traditional approach and to investigate factors associated with campylobacteriosis.[3]

METHODS

Study Enrollment

The Foodborne Diseases Active Surveillance Network (FoodNet) is a collaboration among the Centers for Disease Control and Prevention, 10 state health departments, the US Department of Agriculture's Food Safety and Inspection Service (USDA-FSIS), and the Food and Drug Administration. FoodNet has conducted active, population-based surveillance for laboratory-diagnosed infections of *Campylobacter* since 1996.[13] In 1998, FoodNet included seven sites; its catchment accounted for 8% of the US population. We used data from a 12-month, population-based case–control study conducted during 1998–1999 in the seven sites: Connecticut, Georgia, Minnesota, Oregon, and selected counties in

California, Maryland, and New York. The study design was reported previously.[3] Briefly, we defined a case as a culture-confirmed *Campylobacter* infection in a patient who was not a part of a recognized outbreak, and where diarrhea occurred <10 days before a clinical specimen that yielded *Campylobacter* was obtained. If more than one member of a household had a culture-confirmed illness, only the one with the earliest onset of diarrhea was enrolled in the study. We matched control to a case based on age group and telephone exchange. We enrolled a total of 2632 subjects: 1316 cases and 1316 matched controls. The age groups were 0 to <6 months, 6 to <24 months, 2 to <6 years, 6 to <12 years, 12 to <18 years, 18 to <40 years, 40 to <60 years, and ≥60 years old. The Centers for Disease Control and Prevention and FoodNet site Institutional Review Boards approved the protocol and obtained informed consent.

Exposures

We interviewed cases (or caregivers for cases aged <12 years) by phone within 21 days after specimen collection using a standardized questionnaire about exposures in the 7 days before diarrhea began. Controls were interviewed with the same questionnaire about exposures during the corresponding 7 days. To reduce recall bias, we interviewed controls no later than 7 days after the matched patient was interviewed. We collected data on food, water, and animal exposures; demographics; international and domestic travel; dining locations; and food preparation and kitchen practices. We defined nonpoultry meat as beef, veal, lamb, pork, or venison. We defined farm animals as chickens, turkeys, cattle, goats, sheep, lambs, horses, or pigs. We defined house animals as kittens, cats, puppies, dogs, or reptiles. We defined a rare exposure as having a prevalence of ≤1% in the control group.

Statistical Analysis

To be consistent with the previously published analysis, the same exclusions were applied in the present study. Children under 24 months old (75 cases and 75 controls) were excluded due to a large number of missing values in that age group. Cases (162) and controls (18) with recent international travel were excluded because the study aimed to identify domestic risks. Our final analytic sample included 1079 cases and 1223 controls.

Unlike the previously published analysis, we addressed missing exposure data using multiple imputations (m = 100). Gower distance is the weighted mean of the standardized difference in each variable between two observations. It ranges between 0 and 1, with 0 indicating the two observations are identical, and 1 indicating maximally dissimilar. In each of the 100 imputed datasets, iterating through each exposure, we matched up to 20 nearest controls to each case using Gower distance calculated from all other exposures except possible intermediates between the exposure and illness status (see eAppendix; http://links.lww.com/EDE/B929).[14] For example, when examining ‘fried chicken that was pink inside’, we considered ‘ate fried chicken’ as an intermediate and excluded it from calculating Gower distance. A control could match a case only if its distance was within a threshold established by logistic regression comparing Gower distances between the closest case–control pairs (each case uniquely matched to one control by the maximum bipartite graph algorithm) vs. a randomly selected control. The threshold corresponded to a distance in which there was a logistic regression-estimated probability that the pair was the closest

one vs. a random selection. We tested probabilities from 0.50 to 1 with an interval of 0.05. Each case and its matched controls formed a stratum. Our approach also ensured each control was in only one stratum. If two cases shared the same controls, the cases were merged into the same stratum. To favor matching as many cases as possible and to make the strata more balanced in size, a control was prioritized to be matched with a case from the stratum with the smallest size. If the sizes of strata were tied, priority was given to the closer case.

The previously published conditional logistic regression analysis identified 16 exposures that were associated with campylobacteriosis.[3] To compare the nearest-neighbors matching approach to the published analysis, we first applied the previously published conditional logistic regression model to each of the 100 imputed datasets and combined the regression coefficients from the 100 imputed datasets using Rubin's rules ("original analysis").[15] We then applied conditional logistic regression analysis, including the exposure of interest and strata to each of the 100 datasets re-matched using nearest-neighbors matching for each of the 16 exposures. Again, we combined regression coefficients for each exposure using Rubin's rules. The previously published regression model examined contact with farm animals stratified by age; we did not conduct stratified analysis because we focused on a side-by-side comparison of directions of associations found in the original analysis vs. the nearest-neighbors matching approach.

For our comprehensive analysis of exposures associated with campylobacteriosis using nearest-neighbors matching, we only excluded exposures from analysis if neither cases nor controls were exposed in any of the 100 re-matched datasets or no case could be matched to control in any of the 100 imputed datasets. Since some exposures were rare among participants and conditional logistic regression cannot produce reasonable regression coefficients for these exposures, we applied Firth's bias-reduced penalized-likelihood logistic regression analysis to each of the 100 re-matched datasets for each exposure to examine associations between an exposure and campylobacteriosis after controlling for strata.[16] We combined the results from the 100 re-matched datasets using penalized-likelihood profiles because of the non-normal distribution of regression coefficients.[17] We calculated PAFs for individual binary risk factors using the method described by Bruzzi et al.[18] We calculated the 95% confidence interval (CI) for PAFs from the 95% confidence limits of odds ratios (OR). We examined the association between each pair of exposures using Cramer's V.[19] Association was considered very strong if Cramer's V was greater than 0.25. We selected associations for the presentation that were significant at the 0.05 level and that had OR <0.5 or >2. All analyses were conducted using R version 3.6.1.[20] We provide an R package called nncc that can perform the analyses through GitHub.

RESULTS

Cases were more often male and had higher income than controls but were otherwise similar (Table 1). Examination of pairwise correlation between exposures found very strong associations for 1046 pairs of exposures among 93,528 pairs of exposures examined.

As logistic regression-estimated probabilities increased from 0.50 to 0.90, the Gower distance threshold decreased gradually; from 0.90 to 1.00, it decreased rapidly (Table 2). We found the same pattern for average sample size of re-matched datasets and number of exposures with 100 re-matched datasets. In contrast, statistical power, indicated by the number of exposures associated with disease, gradually increased up to a logistic regression-estimated probability of 0.90. Therefore, we used that probability, corresponding to a Gower distance threshold of 0.81, for the analysis.

On average, 97% of 1:1 matched pairs in the 100 imputed datasets had Gower distances larger than the threshold established by our logistic regression approach. We applied the nearest-neighbors approach to each of the 100 imputed data and created separate re-matched datasets for every exposure. We excluded 57 (13%) of 433 exposures from analysis because we obtained fewer than 100 valid re-matched datasets for each. Among the 376 exposures retained, 101 were rare binary exposures (prevalence ≤1% in controls in the imputed datasets). Almost all cases (>98%) were included in at least one re-matched dataset for an exposure. On average, a re-matched dataset included 431 cases (interquartile range width = 13) and 901 controls (interquartile range width = 12). Each stratum included no more than five cases and 15 controls for any exposure. Only 0.4% of strata included more than one case, whereas 52% of strata included more than one control.

We compared ORs from the original analysis with the nearest-neighbors matching approach for the 16 exposures that were associated with campylobacteriosis in the previously published analysis (Table 3). For the 10 exposures with odds ratios greater than 1 in the original analysis, the nearest-neighbors matching approach produced associations in the same direction and nearly all CIs overlapped. For the six exposures with ORs less than 1 in the original analysis, the nearest-neighbors matching approach produced consistent direction of associations for two exposures (i.e. eating fresh berries bought at a store and female sex) and positive associations for the other four exposures.

Applying conditional logistic regression analysis (nearest-neighbors matching in Table 3) and Firth's bias-reduced penalized-likelihood logistic regression analysis (Table 4) to the re-matched datasets produced consistent associations for all 16 exposures. The latter method (Table 4) also identified 91 more exposures associated with campylobacteriosis compared to the original analysis.

Among the 91 exposures, eating five or more types of meat (poultry and nonpoultry) was associated with increased risk compared with eating fewer. Eating meals prepared on an outdoor grill was a risk factor.

Consumption of chicken was associated with elevated risk (OR = 3.3, 95% CI = 2.0, 5.6) and had the largest PAF (62%, 95% CI = 45%, 73%). Eating rotisserie chicken, microwaved chicken, outdoor-grilled chicken (especially when pink inside or prepared at a large gathering), and several specific types of chicken (e.g., wings, fingers, nuggets, patties, stir-fry, or rotisserie) prepared at a restaurant were risk factors. Consumption of chicken liver was a risk factor (OR = 7.1, 95% CI = 1.4, 41) although the PAF was low (1.2%).

Purchasing, storing, handling, and cooking raw chicken were associated with increased risk. PAFs were 49% for preparation in the home and 28% for purchase of raw chicken. Using thigh joint looseness to determine if whole chicken is cooked and placing grilled chicken on the same plate used to hold raw chicken without washing between uses were associated with increased risk.

Eating nonpoultry meat was another risk factor (OR = 2.7, 95% CI = 1.6, 5.0) and had a PAF of 59% (95% CI = 34%, 74%). Eating nonpoultry meat that was prepared at a large gathering, pork roast, spareribs, bacon, or venison was also associated with elevated risk. Consumption of certain types of nonpoultry meat (e.g., bacon, steak, roast beef, lamb, sausage) prepared at a restaurant was associated with increased risk.

Eating seafood, such as clams and raw oysters, was associated with increased risk. Eating raw clams had the greatest OR (39, 95% CI = 4.6, 2500) among exposures examined; however, the PAF was low (1.3%).

Contact with animals, particularly animals with diarrhea, was associated with increased risk. Residing or visiting a farm and contact with animal feces were risk factors.

Exposures associated with decreased risk included not handling raw meat or cleaning hands after handling raw meat.

DISCUSSION

We developed a novel nearest-neighbors matching approach in a case–control study that maximizes power and confounding control without evidence of overmatching and enables examination of many granular and rare exposures. We found that more than 100 exposures were associated with campylobacteriosis, whereas only 16 were identified in the previously published conditional logistic regression model. We confirmed that eating chicken and nonpoultry meat were major modifiable exposures because each had a PAF of approximately 60%.

The improved confounding control is a major strength of our method. In the original analysis, we matched only age and telephone exchange between cases and controls and included only a limited number of independent variables in the multivariate regression analysis. The inability of the method used in the original analysis to control for many confounders may be responsible for the protective associations found in that analysis. These include eating nonpoultry meat prepared at home, eating chicken prepared at home, eating fresh berries bought at a store, eating fried chicken, and eating turkey prepared at home, and in some previous studies, such as consumption of raw eggs,[2] poultry,[3,21] meats,[6,22] and fish,[2,22] contact with pets or animals,[5,6,21–23] handling raw poultry,[21] poor kitchen hygiene habits,[22] and swimming in pools or natural water.[6] Our method permitted better assessment of possible confounders. The threshold for Gower distance ensured that the matching was effective. The large proportion of matched pairs in the original analysis with a Gower distance larger than our threshold indicates that they were not well matched initially.

Overmatching is often caused by matching on nonconfounders including intermediates in the causal pathway from exposure to outcome; it can result in reduced statistical efficiency and biased associations. We removed possible intermediates from exposures used for matching and matched only on confounders to avoid overmatching. We also found most risk factors identified in the original analysis and reported in the literature[2] to be associated with illness in our re-analysis; this strongly suggests that our analysis did not lead to overmatching. Even though our matching resulted in only 40% of cases being included on average in a re-matched dataset, our ability to find additional associations compared with the original analysis indicates no evidence of loss of statistical efficiency. This is likely partly because the improvement in confounding control and the reduction in the number of independent variables in the model outweighed the decline in sample size.

Despite reduced sample size, we were able to assess more than 100 rare exposures. The conversion of many confounders to matching strata provided the nearest-neighbors matching approach with an improved ability to examine rare exposures. Given sample size limitations, in most case–control studies the number of participants with any given rare exposure is expected to be small. Consequently, conditional logistic regression models that produce reliable estimates can include only a limited number of confounders. Rare exposures are important because they can be associated with very high individual risk even if the amount of illness resulting from these exposures in the population is low, for example, consumption of chicken liver and raw clams. Different public health approaches are more appropriate for rare exposures (e.g., public education) than more common ones (e.g., regulatory changes).

The validity of our method is further supported by the fact that many previous case–control studies have reported risk factors similar to those we found, including consumption of chicken, undercooked meat,[22,24] and seafood,[22] and eating outside the home[2,6,23,24] as risk factors.[2,6,7,22,24–29] Previous case–control studies reported that handling of raw poultry meat[4,6,21] and poor kitchen hygiene habits[2,6,22] were risk factors, whereas good kitchen hygiene habits[4,24] have been reported to be protective. Many studies have also found that contact with animals and the environment were risk factors.[2,6,7,21,22,24,25,27] However, none of these studies provided the information on the full range of exposures in our analysis or the details on the risks of particular exposures enabled by our method. Previous studies support our findings regarding exposures that were protective in the original analysis but risky in our analysis. For example, a systematic review of case–control studies[2] found that eating poultry at home was associated with increased risk (OR = 1.3, 95% CI = 0.99, 1.6). Based on our criteria for the association, we only found that not handling meat and cleaning hands after handling meat were associated with a reduced risk of *Campylobacter* infection. We did not find any exposures associated with reduced risk that were unexpected or that had been identified as risk factors by other studies.

Nearest-neighbors matching is likely to be most effective when there are more controls than cases available making it more likely that most cases can be matched to at least one control. Inclusion of most cases in analyses may help avoid selection bias, improve the generalizability of findings, improve accuracy and precision of estimates, and reduce occurrences where neither cases nor controls are exposed for rare exposures. An increase in the number of controls in a case–control study must be balanced with practical factors,

such as cost, but because nearest-neighbors matching permits tight matching of cases and controls at the analysis stage, the method may allow investigators to obtain controls without the burden of finding those that match cases during study enrollment.

Our study has limitations. Although our findings were consistent with the literature with respect to the directions of association, our estimates of ORs could be biased, because cases included in our re-matched analysis were selected based on Gower distance rather than a random mechanism. We were not able to evaluate possible bias directly because this real-life study is subject to measurement error, unmeasured confounders, and a lack of knowledge about the true associations. Also, selecting results based on statistical significance likely overestimates the true strength of association.[30] We used data from an old case–control study and so our results might not be applicable to the exposures most relevant today. However, the use of old data does not negate its usefulness in evaluating the nearest-neighbors matching approach and even permits reflection on our new findings in the light of subsequent history. For example, because of continued reports of illness, chicken liver has become a priority for the USDA-FSIS[31,32]; our analysis provides support for reducing contamination. In addition, Firth's bias-reduced penalized-likelihood logistic regression was not designed for analyzing matched case–control data. Although this method tended to produce slightly different ORs compared with conditional logistic regression, the discrepancies had little influence on our conclusions. Firth's regression had the advantage of not generating infinite ORs, as occurs with the Mantel-Haenszel method or conditional logistic regression when only a small number of cases or controls were exposed.

In conclusion, the nearest-neighbors matching approach appears to be more efficient than conventional multivariate logistic analysis in the analysis of case–control data with many variables, many of which are interdependent. The approach allows examination of highly granular and rare exposures and may inform the development of specific and actionable public health measures and recommendations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

1. Centers for Disease Control and Prevention. Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet 2015 Surveillance Report (Final DATA). U.S. Department of Health and Human Services, Centers for Diseases Control and Prevention; 2017.
2. Domingues AR, Pires SM, Halasa T, Hald T. Source attribution of human campylobacteriosis using a meta-analysis of case-control studies of sporadic infections. Epidemiol Infect. 2012;140:970–981. [PubMed: 22214729]
3. Friedman CR, Hoekstra RM, Samuel M, et al. ; Emerging Infections Program FoodNet Working Group. Risk factors for sporadic Campylobacter infection in the United States: a case-control study in FoodNet sites. Clin Infect Dis. 2004;38 Suppl 3:S285–S296. [PubMed: 15095201]
4. Pogreba-Brown K, Baker A, Ernst K, Stewart J, Harris RB, Weiss J. Assessing risk factors of sporadic Campylobacter infection: a case-control study in Arizona. Epidemiol Infect. 2016;144:829–839. [PubMed: 26828241]

5. Mughini Gras L, Smid JH, Wagenaar JA, et al. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. PLoS One. 2012;7:e42599. [PubMed: 22880049]
6. Rosner BM, Schielke A, Didelot X, et al. A combined case-control and molecular source attribution study of human Campylobacter infections in Germany, 2011–2014. Sci Rep. 2017;7:5139. [PubMed: 28698561]
7. Stafford RJ, Schluter PJ, Wilson AJ, Kirk MD, Hall G, Unicomb L; OzFoodNet Working Group. Population-attributable risk estimates for risk factors associated with Campylobacter infection, australia. Emerg Infect Dis. 2008;14:895–901. [PubMed: 18507899]
8. Cologne JB, Shibata Y. Optimal case-control matching in practice. Epidemiology. 1995;6:271–275. [PubMed: 7619935]
9. Guisan A, N.E. Z. Predictive habitat distibution models in ecology. Ecol Modell. 2000;135:147–186.
10. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci. 2003;43:1235–1248.
11. Hawkins SS, Stern AD, Baum CF, Gillman MW. Compliance with the Baby-Friendly Hospital Initiative and impact on breastfeeding rates. Arch Dis Child Fetal Neonatal Ed. 2014;99:F138–F143. [PubMed: 24277661]
12. Shirts BH, Bennett ST, Jackson BR. Using patients like my patient for clinical decision support: institution-specific probability of celiac disease diagnosis using simplified near-neighbor classification. J Gen Intern Med. 2013;28:1565–1572. [PubMed: 23645451]
13. Centers for Disease Control and Prevention. Foodborne Diseases Active Surveillance Network (FoodNet). Available at: https://www.cdc.gov/foodnet/index.html Accessed November 18, 2019.
14. Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971;27:5.
15. Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley and Sons; 1987.
16. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80:27–38.
17. Heinze G, Ploner M, Beyea J. Confidence intervals after multiple imputation: combining profile likelihood information from logistic regressions. Stat Med. 2013;32:5062–5076. [PubMed: 23873477]
18. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985;122:904–914. [PubMed: 4050778]
19. Cramer H. Mathematical methods of statistics. Princeton University Press; 1946.
20. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2019.
21. Davis MA, Moore DL, Baker KN, et al. Risk factors for campylobacteriosis in two washington state counties with high numbers of dairy farms. J Clin Microbiol. 2013;51:3921–3927. [PubMed: 24025908]
22. Doorduyn Y, Van Den Brandhof WE, Van Duynhoven YT, Breukink BJ, Wagenaar JA, Van Pelt W. Risk factors for indigenous Campylobacter jejuni and Campylobacter coli infections in The Netherlands: a case-control study. Epidemiol Infect. 2010;138:1391–1404. [PubMed: 20223048]
23. Mossong J, Mughini-Gras L, Penny C, et al. Human campylobacteriosis in luxembourg, 2010–2013: a case-control study combined with multilocus sequence typing for source attribution and risk factor analysis. Sci Rep. 2016;6:20939. [PubMed: 26860258]
24. MacDonald E, White R, Mexia R, et al. Risk factors for sporadic domestically acquired campylobacter infections in Norway 2010–2011: a national prospective case-control study. PLoS One. 2015;10:e0139636. [PubMed: 26431341]
25. Buettner S, Wieland B, Staerk KD, Regula G. Risk attribution of Campylobacter infection by age group using exposure modelling. Epidemiol Infect. 2010;138:1748–1761. [PubMed: 20598208]
26. Bassal R, Ovadia A, Bromberg M, et al. Risk factors for sporadic infection with campylobacter Spp. Among children in Israel: a case-control study. Pediatr Infect Dis J. 2016;35:249–252. [PubMed: 26569191]

27. Fajó-Pascual M, Godoy P, Ferrero-Cáncer M, Wymore K. Case-control study of risk factors for sporadic Campylobacter infections in northeastern Spain. Eur J Public Health. 2010;20:443–448. [PubMed: 20034929]

28. Komba EV, Mdegela RH, Msoffe PL, Nielsen LN, Ingmer H. Prevalence, antimicrobial resistance and risk factors for thermophilic campylobacter infections in symptomatic and asymptomatic humans in Tanzania. Zoonoses Public Health. 2015;62:557–568. [PubMed: 25753615]

29. Mellou K, Sourtzi P, Tsakris A, Saroglou G, Velonakis E. Risk factors for sporadic Campylobacter jejuni infections in children in a Greek region. Epidemiol Infect. 2010;138:1719–1725. [PubMed: 20492748]

30. Lane DMD, W P. Estimating effect size: bias resulting from the significance criterion in editorial decisions. Br J Math Stat Psychol. 1978;31:107–112.

31. Chicken liver - resources for illness prevention. Available at: https://www.fsis.usda.gov/ChickenLiver. Accessed November 18, 2019.

32. Lanier WA, Hale KR, Geissler AL, Dewey-Mattia D. Chicken liver-associated outbreaks of campylobacteriosis and salmonellosis, United States, 2000–2016: identifying opportunities for prevention. Foodborne Pathog Dis. 2018;15:726–733. [PubMed: 30192164]

TABLE 1.

Characteristics of Study Participants

	Cases (n= 1079)	Controls (n= 1223)
Male (%)	53	34
White (%)	83	80
Age group in years (%)		
2–<6	7	7
6–<12	5	4
12–<18	5	5
18–<40	41	41
40–<60	29	30
≥60	13	13
Location of residence (%)		
Urban area	35	37
Suburban area	36	36
Town	13	13
Rural area, not a farm	9	9
Farm in rural area	7	4
Education (%)		
Less than high school graduate	12	10
High school graduate	23	22
Some college or college graduate	53	56
Master's or doctoral degree	11	12
Annual household income, $ (%)		
≤15,000	12	13
>15,000-≤30,000	16	20
>30,000-≤60,000	34	35
>60,000-≤ 100,000	23	20
>100,000	15	11

TABLE 2.

Characteristics Related to Nearest-Neighbors Matching for Logistic Regression-Estimated Probabilities From 0.50 to 1 With an Interval of 0.05

	Logistic Regression-Estimated Probabilities										
Characteristics	**0.50**	**0.55**	**0.60**	**0.65**	**0.70**	**0.75**	**0.80**	**0.85**	**0.90**	**0.95**	**1.00**
Gower distance threshold	0.106	0.104	0.102	0.099	0.096	0.094	0.090	0.086	0.081	0.072	0.043
Average sample size of re-matched datasets	1384	1380	1375	1373	1369	1360	1355	1344	1332	1247	395
Number of exposures with 100 re-matched datasets	382	377	377	377	376	378	379	372	378	372	254
Number of exposures associated with disease	78	79	77	79	88	92	98	111	126	109	11

TABLE 3.

Comparison of Odds Ratios and 95% Confidence Intervals for 16 Exposures That Were Associated With Campylobacteriosis in the Published Analysis[a], by Original Analysis and the Nearest-Neighbors Matching Results

	Original Analysis[b]		Nearest-Neighbors Matching[c]	
Exposure	**OR**	**95% CI**	**OR**	**95% CI**
Ate chicken prepared at a restaurant	2.0	1.5, 2.6	3.4	2.3, 5.0
Ate non-poultry meat prepared at a restaurant[d]	1.6	1.3, 2.0	2.3	1.7, 3.3
Had contact with animal stool	1.4	1.1, 1.9	2.9	2.0, 4.3
Had pet puppy	3.7	2.1, 6.6	5.2	2.4, 11
Had contact with farm animals[e]	1.6	1.1, 2.3	3.6	2.1, 6.3
Ate turkey prepared at a restaurant	1.7	0.9, 3.2	5.0	1.8, 14
Drank untreated water from a lake, river, or stream	3.1	1.6, 6.1	2.5	0.9, 6.8
Ate undercooked or pink chicken	2.0	1.2, 3.4	4.0	1.9, 8.2
Ate raw seafood	1.7	1.0, 2.9	2.8	1.4, 5.5
Drank raw milk	4.5	1.5, 14	3.0	0.8, 12
Ate nonpoultry meat prepared at home[d,f]	0.7	0.5, 0.9	1.1	0.8, 1.6
Ate chicken prepared at home[f]	0.6	0.5, 0.8	1.6	1.1, 2.4
Ate fresh berries bought at a store	0.7	0.5, 0.9	0.8	0.6, 1.1
Female sex	0.5	0.4, 0.6	0.5	0.4, 0.7
Ate fried chicken	0.5	0.4, 0.7	1.4	0.9, 2.3
Ate turkey prepared at home[f]	0.8	0.6, 1.1	1.1	0.6, 2.0

[a] The multivariate analysis in Friedman et al. (3).

[b] The conditional logistic regression model that includes all 16 exposures listed in this table was applied to each of the 100 imputed datasets. Estimates were combined using Rubin's rules.

[c] For each of the 16 exposures, the nearest-neighbors matching approach was implemented based on all measured exposures in each of the 100 imputed datasets to create re-matched datasets, and each re-matched dataset was analyzed using a conditional logistic regression model that includes the exposure of interest and re-matched strata. Estimates were combined using Rubin's rules. This table presents the combined results for the 16 exposures.

[d] Nonpoultry meat was defined as beef, veal, lamb, pork, and venison.

[e] Farm animals were defined as chicken, turkey, cattle, goat, sheep, lamb, horse, and pig. We did not stratify the analysis by age. These results are for persons aged >2 years.

[f] Prepared at one's own home.

CI, confidence interval; OR, odds ratio.

TABLE 4.

Risk and Protective Factors Associated With Campylobacteriosis, Percentages Exposed, Odds Ratios From the Nearest-Neighbors Matching Approach and Population Attributable Fractions for Exposures

Exposure	% of cases exposed	% of controls exposed	OR[a]	95% CI[b]	PAF, %	95% CI[b,c]
Exposures identified in the published analysis[d]						
Ate chicken prepared at a restaurant	41	22	3.9	2.6, 6.0	31	26, 34
Ate nonpoultry meat prepared at a restaurant[e]	51	34	2.6	1.8, 3.7	31	23, 37
Had contact with animal stool	26	13	3.4	2.2, 5.1	18	15, 21
Had petpuppy	7.4	1.7	6.3	2.9, 14	6.2	4.8, 6.9
Had contact with farm animals[e]	14	5.2	4.3	2.4, 7.9	11	8.3, 12
Ate turkey prepared at a restaurant	4.1	1.0	6.1	2.2, 18	3.4	2.2, 3.8
Drank untreated water from a lake, river, or stream	3.3	1.3	2.9	1.0, 8.3	2.2	0.1, 2.9
Ate undercooked or pink chicken	7.1	2.3	4.7	2.2, 10	5.6	3.9, 6.4
Ate raw seafood	6.5	2.6	3.2	1.5, 6.8	4.5	2.3, 5.6
Drank raw milk	2.0	0.5	3.5	0.8, 15	-	-
Ate nonpoultry meat prepared at home[e]	72	71	1.1	0.7, 1.7	-	-
Ate chicken prepared at home	61	57	1.7	1.1, 2.7	26	8.0, 39
Ate fresh berries bought at a store	20	24	0.8	0.5, 1.1	-	-
Female sex	50	67	0.4	0.3, 0.6	-	-
Ate fried chicken	20	18	1.5	0.9, 2.4	-	-
Ate turkey prepared at home	8.8	7.3	1.2	0.6, 2.1	-	-
Exposures identified only in the nearest-neighbors matching approach						
Ate ≥5 types of meat (poultry and nonpoultry)[e]	63	42	2.8	2.0, 4.0	40	31, 47
Location of food preparation						
Ate meals prepared on an outdoor grill at a large gathering	3.1	1.0	4.2	1.4, 13	2.4	0.9, 2.9
Ate ≥2 types of poultry and nonpoultry meat prepared at a restaurant[e]	39	21	3.8	2.5, 5.7	29	24, 32
Ate poultry meat						
Chicken	89	77	3.3	2.0, 5.6	62	45, 73
Chicken wings	14	7.7	3.3	1.8, 6.1	9.8	6.3, 12

Exposure	% of cases exposed	% of controls exposed	OR[a]	95% CI[b]	PAF, %	95% CI[b,c]
Number of times ate chicken wings (vs. 0)						
1	9.6	6.2	2.0	1.1, 3.7	-	-
≥2	3.1	1.7	2.0	0.7, 5.7	-	-
Rotisserie chicken	8.4	4.5	2.2	1.1, 4.4	4.6	0.9, 6.5
Microwaved chicken	5.3	2.6	2.6	1.1, 6.2	3.3	0.5, 4.5
Outdoor-grilled chicken	21	16	2.5	1.4, 4.5	12	5.8, 16
Outdoor-grilled chicken prepared at a large gathering	1.2	0.2	17	1.3,1100	1.2	0.3, 1.2
Chicken liver	1.4	0.3	7.1	1.4, 41	1.2	0.4, 1.4
Number of times ate chicken liver (at least one vs. none)	1.3	0.3	8.6	1.6, 55	1.1	0.5, 1.3
≥2 types of chicken meat	59	41	2.7	1.9, 3.7	37	28, 43
Chicken prepared at restaurant						
Chicken stir-fry prepared at a restaurant	5.6	1.7	4.8	2.0, 11	4.5	2.9, 5.1
Chicken wings prepared at a restaurant	7.9	3.5	2.8	1.4, 5.9	5.1	2.1, 6.5
Chicken fingers, nuggets, or patties prepared at a restaurant	12	7.5	2.0	1.2, 3.5	6.3	1.9, 8.9
Rotisserie chicken prepared at a restaurant	7.3	3.2	2.8	1.4, 5.7	4.7	2.1, 6.1
≥2 types of chicken meat at a restaurant	15	5.5	5.0	2.7, 9.5	12	9.3, 13
Chicken prepared at home						
Chicken liver prepared at home	1.3	0.2	9.6	1.7, 61	1.2	0.5, 1.3
Pink chicken						
Outdoor-grilled chicken that was pink inside	1.6	0.4	5.1	1.1, 25	1.3	0.1, 1.5
Fried chicken that was pink inside	1.3	0.5	5.5	1.1, 31	1.1	0.2, 1.3
≥2 types of turkey meat	7.4	3.7	2.3	1.1, 4.6	4.2	0.9, 5.8
Oven roasted or baked turkey that was pink inside	0.7	0.1	11	1.1, 130	0.6	0.1, 0.7
Turkey prepared in someone else's home	2.6	0.6	4.9	1.3, 20	2.0	0.5, 2.4
Turkey prepared at a restaurant						
Ground turkey prepared at a restaurant	0.8	0.1	23	1.5, 2000	0.8	0.3, 0.8
Turkey oven roasted or baked at a restaurant	2.8	0.9	3.7	1.2, 12	2.0	0.5, 2.6
Other poultry-related factors						
Stuffing prepared inside poultry	3.5	1.4	3.0	1.2, 7.8	2.4	0.5, 3.1
Other poultry such as Cornish game hen, duck, or goose	2.8	1.1	3.5	1.2, 9.8	2.0	0.5, 2.5
Raw meat contact and kitchen practices						

Exposure	% of cases exposed	% of controls exposed	OR[a]	95% CI[b]	PAF, %	95% CI[b,c]
A household member purchased raw chicken	50	40	2.3	1.6, 3.5	28	19, 35
Chicken juices leaked onto other foods in the grocery bag	2.5	0.8	3.8	1.1, 15	1.9	0.2, 2.4
Raw chicken stored in your refrigerator	30	22	2.2	1.5, 3.3	16	9.5, 21
Raw chicken was placed in refrigerator not in the store package	8.4	4.7	2.3	1.2, 4.2	4.7	1.5, 6.4
Raw chicken was prepared at home	64	55	4.2	2.5, 7.7	49	38, 55
Touched raw chicken during preparation	41	36	2.2	1.4, 3.6	23	12, 29
Raw chicken was cut into small pieces at home	15	14	2.2	1.2, 4.2	8.5	2.6, 12
A cutting board was used to cut raw chicken	13	11	2.2	1.2, 4.4	7.1	1.8, 9.9
Cutting board was washed with soap after cutting raw chicken	12	11	2.1	1.1, 4.2	6.5	0.9, 9.5
Use thigh joint looseness to determine when whole chicken fully cooked	2.5	0.8	4.0	1.2, 13	1.9	0.5, 2.4
Placed grilled chicken on the same plate used for raw chicken without washing	2.0	0.7	3.8	1.0, 14.4	1.5	0.1, 1.9
Ate nonpoultry meat[e]						
Nonpoultry meat	93	85	2.7	1.6, 5.0	59	34, 74
Nonpoultry meat prepared at a large gathering	3.6	0.5	10.2	3.1, 38	3.3	2.4, 3.5
Sausage prepared at a large gathering	0.7	0	24	1.2, 2900	0.6	0.1, 0.7
Hamburger prepared at a large gathering	1.7	0.3	9.7	2.0, 58	1.6	0.9, 1.7
Pork roast	7.9	4.6	2.2	1.2, 4.2	4.3	1.1, 6.0
Spareribs	8.1	4.8	2.1	1.1, 4.0	4.3	0.8, 6.0
Bacon	34	23	2.3	1.5, 3.4	19	12, 24
Venison	4.5	1.9	2.8	1.2, 6.7	2.9	0.7, 3.8
Venison prepared in someone else’s home	0.6	0.1	23	1.2, 2900	0.6	0.1, 0.6
≥3 types of nonpoultry meat	58	44	2.1	1.5, 2.9	30	19, 38
Nonpoultry meat prepared at a restaurant						
≥2 types of nonpoultry meat prepared at a restaurant	22	10	4.0	2.5, 6.5	17	13, 19
Bacon prepared at a restaurant	9.5	3.7	3.5	1.8, 6.7	6.8	4.4, 8.1
Steak prepared at a restaurant	9.2	4.8	2.5	1.4, 4.6	5.6	2.7, 7.2
Roast beef prepared at a restaurant	6.4	3.4	2.1	1.0, 4.4	3.4	0.1, 4.9
Lamb prepared at a restaurant	1.6	0.2	9.4	1.7, 69	1.5	0.6, 1.6
Sausage prepared at a restaurant	4.7	1.9	2.6	1.1, 6.1	2.9	0.5, 4.0
Ate seafood						

Epidemiology. Author manuscript; available in PMC 2023 May 22.

Exposure	% of cases exposed	% of controls exposed	OR[a]	95% CI[b]	PAF, %	95% CI[b,c]
Clams	5.0	2.8	2.1	1.0, 4.3	2.6	0.1, 3.9
Raw seafood						
Oysters	1.4	0.5	4.6	1.1, 20	1.1	0.1, 1.3
Clams	1.3	0.1	39	4.6, 2500	1.3	1.0, 1.3
Had contact with animals						
Animals with diarrhea	13	2.9	6.5	3.5, 12	11	9.2, 12
Puppy	11	5.0	2.8	1.6, 5.1	7.0	4.0, 8.8
Puppy with diarrhea	2.8	0.3	12	2.8, 62	2.6	1.8, 2.8
≥2 kinds of animals	41	26	2.3	1.6, 3.2	23	16, 29
≥2 kinds of your pets	26	14	3.0	1.9, 4.7	17	12, 20
Farm animals						
Farm animal with diarrhea	3.9	0.3	26	6.3, 220	3.8	3.3, 3.9
Chicken	4.8	1.0	7.3	2.6, 22	4.1	2.9, 4.6
Chicken that was your farm animal	2.7	0.6	6.4	1.8, 25	2.3	1.2, 2.6
Chicken with diarrhea	0.9	0.1	23	1.7, 1700	0.9	0.4, 0.9
Turkey that was your farm animal	0.7	0.1	36	2.4, 3900	0.7	0.4, 0.7
Cow, bull, or steer	6.5	1.8	5.8	2.4, 15	5.4	3.8, 6.1
Cow, bull, or steer with diarrhea	2.2	0.2	21	3.4, 360	2.1	1.6, 2.2
Calf	4.5	0.7	15	4.2, 72	4.2	3.4, 4.4
Exposure	% of cases exposed	% of controls exposed	OR[a]	95% CI[b]	PAF, %	95% CI[b,c]
Calf that was your farm animal	2.4	0.5	6.5	1.5, 36	2.1	0.8, 2.4
Calf with diarrhea	2.3	0.2	23	4.1, 380	2.2	1.8, 2.3
Horse	5.5	2.5	2.7	1.2, 6.4	3.5	0.7, 4.6
Pig with diarrhea	0.5	0.1	23	1.0, 3500	0.5	0.01, 0.5
Environment						
Visited a farm	11	3.5	4.6	2.4, 8.7	8.7	6.6, 9.9
Animals present on the farm you visited	9.4	3.9	3.0	1.6, 5.5	6.3	3.6, 7.7
Lived on a farm	10	4.0	6.4	2.8, 16	8.5	6.4, 9.4
Animals present at the farm you lived on	7.9	3.3	4.5	2.0, 11	6.1	3.9, 7.1
Had contact with any animal manure	9.1	4.6	2.3	1.3, 4.2	5.2	1.9, 6.9

Exposure	% of cases exposed	% of controls exposed	OR[a]	95% CI[b]	PAF, %	95% CI[b,c]
Had contact with animal stool						
Farm animal stool	6.2	0.9	15	5.2, 47	5.8	5.0, 6.1
Cow, bull, or steer stool	3.5	0.3	22	4.9, 210	3.3	2.8, 3.5
Calf stool	2.4	0.3	15	3.1, 98	2.2	1.6, 2.4
Chicken stool	2.1	0.3	8.8	1.8, 51	1.9	0.9, 2.1
Stool of bird other than chicken and turkey	2.8	1.2	3.0	1.0, 9.2	1.9	0.05, 2.5
Puppy stool	4.0	0.8	6.2	2.1, 20	3.4	2.1, 3.8
Dog stool	9.1	5.3	2.1	1.1, 3.7	4.7	1.2, 6.6
Stool of animals other than farm and house animals [e]	2.1	0.5	4.8	1.2, 21	1.7	0.3, 2.0
Resided on a farm in rural area	8.7	3.6	7.0	3.0, 18	-	-
Protective factors						
After handling raw meat (vs. continued cooking without washing hands)						
Rinsed or wiped hands then continued cooking	16	11	0.4	0.1, 1.2	-	-
Washed hands with soap then continued cooking	72	76	0.2	0.1, 0.8	-	-
Do not handle raw meat	9.2	12	0.2	0.04, 0.6	-	-

[a] OR for an exposure with all measured confounders adjusted for using nearest-neighbor matching in each of the 100 re-matched datasets using Firth's bias-reduced penalized- likelihood logistic regression analysis; the 100 ORs were combined into a single OR reported here using the combination of penalized-likelihood profiles.

[b] Abbreviation for confidence interval.

[c] PAF was calculated for risk factors that are binary.

[d] The multivariate analysis in Friedman et al. (3).

[e] Non-poultry meat was defined as beef, veal, lamb, pork, and venison. Farm animals were defined as chicken, turkey, cattle, goat, sheep, lamb, horse, and pig. House animals were defined as kitten, cat, puppy, dog, and reptile.

[f] Other chicken is defined as chicken other than oven roasted or baked chicken, ground chicken, broiled chicken, rotisserie chicken, outdoor-grilled chicken, fried chicken, chicken stir-fry, microwaved chicken, chicken wings, chicken fingers, nuggets, and patties, chicken luncheon meat sliced at a deli, chicken sausage, chicken liver, chicken pot pie, and chicken salad.

CI, confidence interval; OR, odds ratio; PAF, population attributable fraction.