



# KEY POINTS

for Successful Foodborne Outbreak Detection and Investigation

## USING BINOMIAL PROBABILITY CALCULATIONS DURING CLUSTER INVESTIGATIONS

**Scenario:** You have had 10 *Salmonella* cases closely related by WGS reported to your health department in the last week. Initial interviews do not reveal any obvious commonalities (e.g., restaurants, gatherings); therefore, you suspect a retail food item. Indeed, a majority of cases in this cluster do report eating a couple of different food items, but you aren't certain how unusual the reported consumption rates might be. So, how should you proceed to advance hypotheses?

Population based case-control studies have historically been used in these situations. However, recruiting representative controls for these studies is very resource intensive, and with the increasing use of cell phones, often very difficult. As a result, the use of binomial probability calculations (Bernoulli trials) has emerged as a way to quickly assess potential exposures during cluster investigations. Here are some key points for using binomial probability calculations during cluster investigations, based on the Minnesota Department of Health's (MDH) experience.

### Data Needed to Calculate Binomial Probabilities

- Total number of cases in the cluster
- Number of cases in the cluster who ate the food item of interest
- The background consumption rate of that food item
  - Getting background consumption rate data is the trickiest part of this whole process. Rarely do perfectly suitable (i.e., specific, current, and geographically representative) background food consumption data exist for a given food item. Usually, we have to use various less than perfect sources (see below) and recognize the limitations.

### Sources of Background Food Consumption Rate Estimates

- [FoodNet Population Survey, 2018 - 2019](#)
- Aggregated Case Surveillance Data
  - Minnesota data for sporadic *Salmonella* and *E. coli* O157 cases

### Considerations for Choosing Background Food Consumption Rate Estimates

- Recent data are preferred, as food consumption patterns can change over time.
- Characteristics of the cluster cases
  - If cases have distinct gender, age, or racial/ethnic characteristics, background estimates should be derived from similar populations (like matching in traditional case-control studies)
  - Regional variability - food consumption patterns can vary across states/regions
- Seasonal variation in consumption of particular food items
- Use of sporadic case data for background rates could bias towards the null

- If you are using sporadic case interview data for background consumption estimates, observed consumption rates for known vehicles (e.g., ground beef for *E. coli* O157) may be higher than in to the general population; this could limit your ability to find a statistically significant association even if that food item is the source of the outbreak.
- If no data are available for the specific food item of interest, you can use data for a similar food item, or make an educated guess (it is best to be conservative and err on the high side). The beauty of using this model is that you can plug in a variety of background estimates and see how they might change (or not change) your conclusion.

### Interpretation and Next Steps

- Interpretation of a binomial probability calculation goes like this: by chance alone, how likely are we to find x of n people (or more) eating a given food item? In the example below, we have a cluster of 10 cases, and we are evaluating a food item with an estimated background consumption rate of 25%. Let's say 8 of our 10 cluster cases reported eating that food item. The probability of at least 8 cases eating that food item by chance alone is 0.004. Is this enough to conclude that the food item is the cause of the outbreak? Not usually, but what it does indicate is that efforts to confirm the hypothesis (e.g., through informational trace backs, food testing) are reasonable to pursue.

	# cases with exposure	% cases with exposure	Cumulative Probability	Cumulative Probability (exponential)	P (exactly that many hits)
	10	100%	0.0000	9.537E-07	0.00000
sample size	9	90%	0.0000	2.956E-05	0.00003
10	8	80%	0.0004	4.158E-04	0.00039
	7	70%	0.0035	3.506E-03	0.00309
	6	60%	0.0197	1.973E-02	0.01622
background rate	5	50%	0.0781	7.813E-02	0.05840
25.0%	4	40%	0.2241	2.241E-01	0.14600
	3	30%	0.4744	4.744E-01	0.25028
	2	20%	0.7560	7.560E-01	0.28157
	1	10%	0.9437	9.437E-01	0.18771
	0	0%	1.0000	1.000E+00	0.05631

### Can the Binomial Probability Method be used in Restaurant-specific Studies?

When you have an outbreak at a specific restaurant, the binomial probability method can be used, but it is not appropriate to use general background consumption rate data, such as those derived from the FoodNet Population Survey, for comparison. Instead, data sources need to be specific to that restaurant, and could include data from restaurant sales records such as online orders or transaction records. If data from patron interviews (including well-meal companions, reservation lists, or credit card receipts) is available, a traditional case-control study design can be used.

### Binomial Calculator Resources

- Epi Info 7.0 – under StatCalc - Binomial
- [Oregon Binomial Probability Worksheet](#)

#### Additional Resources:

Hazelnut Case Study

KFC case Study

CIFOR, Third Edition of the CIFOR Guidelines for Foodborne Disease Outbreak Response, Chapter 5 (see page 104 for 5.4 Test Hypotheses)  
<https://cifor.us/downloads/clearinghouse/CIFOR-Guidelines-Complete-third-Ed.-FINAL.pdf>

