# Where do hypotheses come from?

May 10, 2005

PulseNet /Foodborne Epi National meeting
Seattle, Washington

Robert V Tauxe, M.D., M.P.H.

Foodborne and Diarrheal Diseases Branch,
Centers for Disease Control and Prevention,

# Stages of an outbreak investigation

- Stage 1: Detecting a cluster in the first place
- Stage 2: Developing hypotheses about exposure
- Stage 3: Testing the hypotheses
- Stage 4: Reconstructing how and where contamination is likely to have occurred
- Stage 5: Prevention - Intervening in current outbreak and preventing future similar ones

# Analytic epidemiology is the testing of hypotheses

- Axiomatic that:
  - If do not have the right hypothesis, then will not get the right answer.
  - If do not have an hypothesis at all, just doing a blind case-control study is fraught with failure.
  - Major issue for enteric pathogens transmitted through food, as the number of possible vehicles, and the number of control points is enormous.

- Worth reflecting on where hypotheses come from


Restaurant
Chain
A
STRAW
BERRY
SHT CAKE

# Can draw on lots of information to generate the hypotheses

- If we know the pathogen
  - Reservoir
  - Biology
  - History of previous outbreaks
- Look at reported series of outbreaks in eFORS
- Look at case-control studies of sporadic cases
- Look at isolates from animals and foods
- Look at food recall history from UDSA and FDA

- *Vibrio parahaemolyticus* outbreaks 1998-2004: 96% of outbreaks are associated with seafood

CDC

# *E. coli* O157:H7 infections
# Series of reported outbreaks 1982-2002*

| Route | Outbreaks | Percent of outbreaks | Cases | **Percent of cases** |
|---|---|---|---|---|
| Food | 183 | 52% | 5269 | **61%** |
| Person-to-person | 50 | 14% | 651 | **8%** |
| Water | 31 | 9% | 1545 | **18%** |
| Animal contact | 11 | 3% | 319 | **4%** |
| Lab-related | 1 | <1% | 2 | **<<1%** |
| Unknown | 74 | 21% | 812 | **9%** |
| Total | 350 | 100% | 8598 | **100%** |

***Rangel et al. EID 11:603-9, 2005**

CDC

# *E. coli* O157:H7 infections
# Series of reported outbreaks 1982-2002*

| Food vehicle | Outbreaks | Percent of outbreaks | Cases | **Percent of cases** |
|---|---|---|---|---|
| Ground beef | 75 | 41% | 1760 | **33%** |
| Other beef | 11 | 6% | 563 | **11%** |
| Produce | 38 | 21% | 1794 | **34%** |
| Dairy product | 7 | 4% | 300 | **6%** |
| Other | 10 | 5% | 206 | **4%** |
| Unknown | 12 | 23% | 646 | **12%** |
| Total | 183 | 100% | 5269 | **100%** |

***Rangel et al. EID 11:603-9, 2005**

CDC

# Even if we know the pathogen

- New patterns will emerge:
  - Globalization of food supply
  - Bacteria can invade new niches
  - Non-food sources (trendy pets, bottled water…)
  - Ethnic groups import their own tastes
  - Steady increase in produce-associated outbreaks
  - *Salmonella* Heidelberg has now appeared in egg-associated outbreaks, just like *S.* Enteritidis

- Expect the unexpected

CDC

# Can draw on lots of information to generate the hypotheses

- If we know exposure is associated with a place
  - Mass gathering (conference, wedding, church supper)
  - Restaurant (or restaurant chain)
  - Grocery store (or grocery chain)

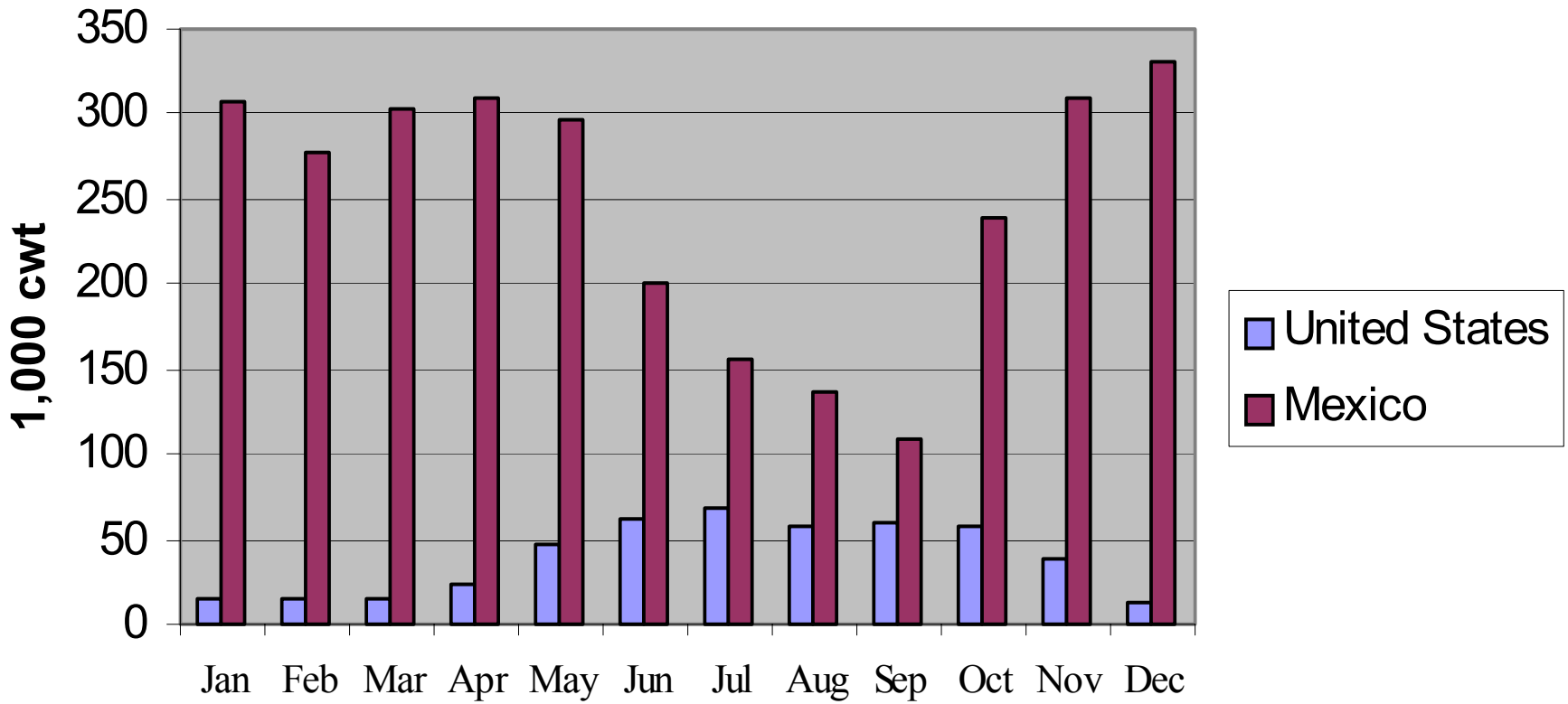
- If cases have a place in common,
  - Then the list of possible items is constrained.
  - Still need the menu, or the shopping list

# Can draw on lots of information to generate the hypotheses

- If we know the epi curve
  - Sharp and tight:  a single production lot, a short shelf life
  - Spread out: suggests a multiple production lots or a product with a long shelf life

- If we know the geography and season
  - Scattered across multiple states,
  - No obvious link -  then presume a single item in trade
  - Seasonal variation in food source

## Figure 2–Shipments of green onions, 2002




Source: Fresh Fruit and Vegetable Shipments, Agricultural Marketing Service, USDA

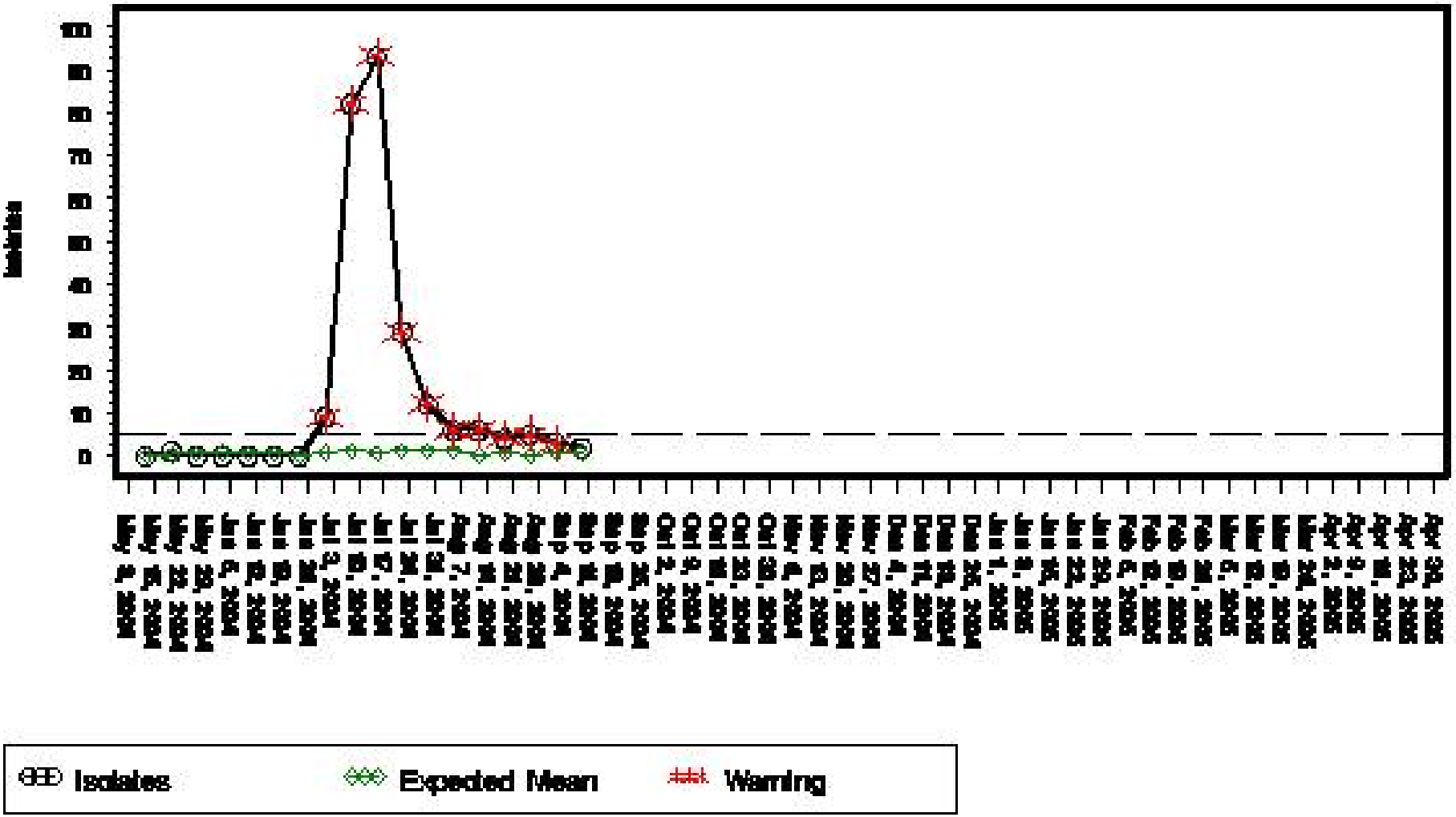
# Can draw on lots of information to generate the hypotheses

- If we know the age, sex and ethnicity of the people
  - Our eating patterns differ greatly
  - Atlas of Food Exposures to get an idea of just how much - FoodNet population survey covers a wide variety of different foods. (www.cdc.gov/foodnet)
  - Distribution of current outbreak vs historical pattern - we use Statistical outbreak detection algorithm (SODA)

# *Statistical Outbreak Detection Algorithm (SODA)*

## *Laboratory Confirmed Salmonella Isolates*

**Area = Pennsylvania  Serotype = Javiana**

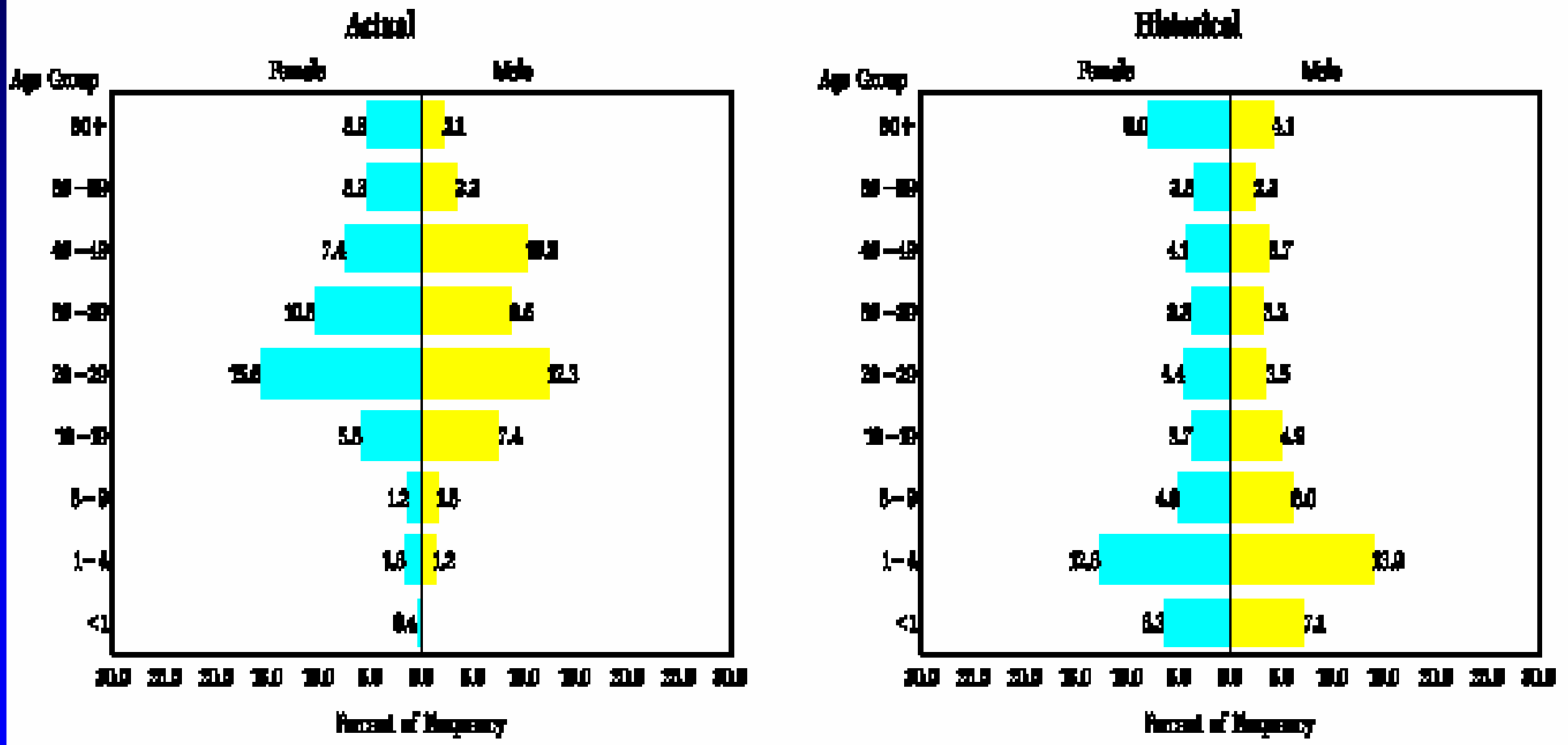


# *Statistical Outbreak Detection Algorithm (SODA)*

## *Laboratory Confirmed Salmonella Isolates*

**May 8, 2004 to May 8, 2005**

**Area = Pennsylvania** **Serotype = Javiana**




Values based on information reported through PHLIS

Values based on 1999–2003 National Information reported through PHLIS

Report Date: May 5, 2005

# Hypothesis generating interviews

- Strategies include:
  - Interviews with structured questionnaire with many food items on it: "trolling, trawling, or shotgun"
  - Intensive open-ended interviews about everything that went into patient's mouth in the last 5 days
  - Some combination of the two
  - All must be done the same way
  - In-depth interview with people in their homes, including refrigerator, pantry

  - (Not recommended: Getting affected people together and asking them to discuss among themselves)

# Detailed open-ended interviews

- Can rule hypotheses out and may identify new ones
- Important when:
  - Poorly understood pathogen
  - Menu is unavailable
  - Unusual setting
  - Unusual population/ethnic group
  - Reason to suspect something unusual is occurring
- However:
  - Should be done by one person
  - Intensive - in-person if at all possible
  - Difficult to achieve in diffuse outbreaks
- The gold standard when other approaches fail

CDC

# How many hypothesis generating interviews?

- 5- 10 interviews is enough

- If plan to revise questionnaire for analytic stage
  - To focus detailed queries on lead hypotheses
  - To collect more data on sources
  - To shorten questionnaire and save time with interviews
  - Review after 5-10 HGIs, choose lead hypotheses, revise questionnaire, and proceed

- If plan to use trolling questionnaire for analytic stage, review after 5-10, make decision, and then proceed with case and control investigative interviews

# Important to consider the frequency of an hypothesized exposure in the population

- If the exposure is rare in the general population
  - It is easy to exclude with a few interviews
  - Doesn't take very many controls to test
  - Hard to develop in the first place
  - This is what makes trolling questionnaires long
- If the exposure is common
  - Unable to exclude it with a few interviews
  - Will take substantial statistical power to test
- If the exposure is universal
  - Very difficult to implicate at all, unless subtype exposure

# Consider the statistical power issue

- For a small cluster (≤5 cases)
  - A rare exposure can be tested with case-control
  - A common exposure is difficult to test
    - Subtype the exposure: brand, specific type, frequency
    - I.E. the hypothesis has to be very specific
    - If you are lucky someone will have something in their freezer
- For a larger cluster (>20 cases)
  - A rare exposure can be easily tested
  - A common exposure can be also be tested
- Of course - at the beginning you often do not know either the size or the hypothesis

# The approach is typically two staged

- Initial stage: a questionnaire that gathers
  - demographics
  - clinical information
  - data on the structure of the exposure
  - Some non-food related exposures
- Detailed exposure data
  - Trolling questionnaire
  - At least for major hypotheses,
    - Data on where and when eaten or purchased
- Some jurisdictions combine them as one interview

# The ideal world

- All *Listeria* cases to get full interview, all *E. coli* O157:H7, *Salmonella* cases get initial interview as soon as reported
- All those isolates are serotyped, subtyped by PulseNet
- Ready to conduct detailed exposure questionnaire for hypothesis generation and testing
- Use standard questionnaires
- FoodNet population survey matches the questionnaires, providing a pre-interviewed comparison group.

# What if we don't get a single strong food association?

- One item, but everyone ate it
- One item that is hard to remember
- Multiple dishes contaminated
- Maybe we missed it on the questionnaire
- Pseudo cluster of a common subtype

# When the first study of a cluster draws a blank

- Consider:
  - Contaminated common ingredient
  - Condiments, garnishes, spices
  - Something else in the kitchens
  - Infected foodhandlers

- If the cluster continues, take another look
- Start at the beginning with hypothesis generating interviews.

There is nothing more precious than a testable hypothesis

# Our challenge

- Detecting and interpreting a signal
- Hear about possible outbreaks from many sources
- Great increase in sensitivity with PulseNet
- Detecting more
  - Small clusters
  - Geographically dispersed clusters

- Central premise: A cluster defined by molecular is likely to have a common origin or source (Depends on frequency of that pattern)