The line listing

Data management during outbreak investigations

FETP India

Competency to be gained from this lecture

Maintain a line listing to prepare the time, place and person analysis and generate hypotheses during an outbreak investigation

Key areas

- Listing cases
- Aggregating cases

Steps of an outbreak investigation

1. Determine the existence of an outbreak
2. Confirm the diagnosis
3. Define a case
4. Search for cases
5. Generate hypotheses using descriptive findings
6. Test hypotheses based upon an analytical study
7. Draw conclusions
8. Compare the hypothesis with established facts
9. Communicate findings
10. Execute prevention measures
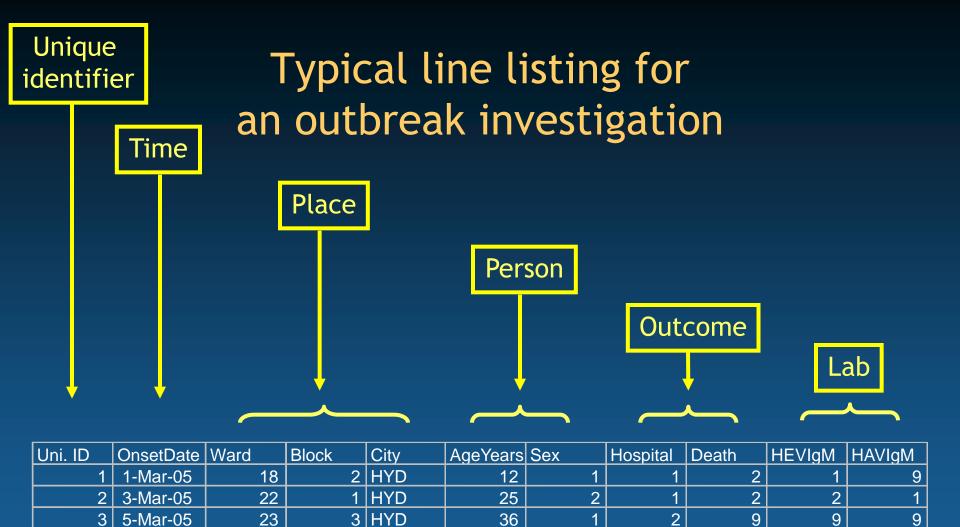
Listing cases

Steps of an outbreak investigation

1. Determine the existence of an outbreak
2. Confirm the diagnosis
3. Define a case
4. Search for cases
5. Generate hypotheses using descriptive findings
6. Test hypotheses based upon an analytical study
7. Draw conclusions
8. Compare the hypothesis with established facts
9. Communicate findings
10. Execute prevention measures

Listing cases

The line listing

- Constitutes and updates a database of cases
- Protects the confidentiality of the patients
- Prepares an easy, automated, descriptive analysis

Listing cases

Guiding principles for the line listing

- Is unique
  - Don't confuse yourself and others with multiple versions
- Contains a unique identifier for each record (case)
- Ensures confidentiality
- Contains essential information on each case
  - Time, place, person, other (e.g., clinical picture, laboratory)
- Can be updated as the investigation develops
- Allows regular, automated, computerized analysis
  - Aggregation

Listing cases

Typical line listing for an outbreak investigation

Unique identifier

Time

Place

Person

Outcome

Lab

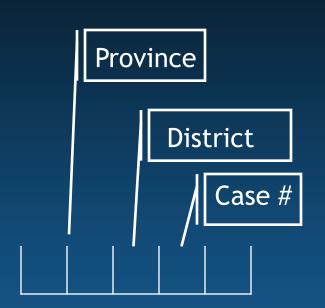| Uni. ID	OnsetDate	Ward	Block	City	AgeYears	Sex	Hospital	Death	HEVIgM	HAVIgM
1	1-Mar-05	18	2	HYD	12	1	1	2	1	9
2	3-Mar-05	22	1	HYD	25	2	1	2	2	1
3	5-Mar-05	23	3	HYD	36	1	2	9	9	9
4	6-Mar-05	-	-	SEC	23	2	1	1	1	2

Listing cases

The identifier in the line listing

- Unique
  - Computerized index
  - Quality assurance
- Confidential
  - Identifying information is kept separate, preferably not in electronic format

Listing cases

Using codes within the unique identifier

- The unique identifier may contain information

- Each digit or set of digits refer to specific information

  - Example:
    - Two first digit: Province
    - Third digit: District
    - Fourth and fifth digit: Case number

Province

District

Case #

Listing cases

Keeping the identifying information separate from the line listing

- Unique identifier in the database
  - E.g., 1, 2, 3
- Separate list with identifying information
  - Links unique database identifiers with identify
  - Contains identifying information
    - Name, address, phone etc...
  - Is managed as a hard, paper copy
  - Is kept under lock and key

Listing cases

Line listing and paper list with identifying information

- ## Electronic line listing
  - ### Public, can be analyzed

Uni. ID	OnsetDate	Ward	Block	City	AgeYears	Sex	Hospital	Death	HEVIgM	HAVIgM
1	1-Mar-05	18	2	HYD	12	1	1	2	1	9
2	3-Mar-05	22	1	HYD	25	2	1	2	2	1
3	5-Mar-05	23	3	HYD	36	1	2	9	9	9
4	6-Mar-05	-	-	SEC	23	2	1	1	1	2

- ## Separate list with identifying information
  - ### Confidential, cannot be analyzed

Uni. ID	Name	Address	Phone
1	Salif Keita	234 Defense colony	044 12 34 56 78
2	K Suresh	234 Nugambakam High Road	044 87 65 43 21
3	A.D. Singh	234 Ring road	044 98 87 65 41
4	R.T. Balaj	678 Vadapalani	044 14 25 36 78

Listing cases

Information about time in the line listing

- Multiple time variables may be needed
  - Onset
  - Hospitalization
  - Death
- Precision needs may vary
  - Hour of onset (gastroenteritis outbreak)
  - Date of onset (Hepatitis outbreak)
  - Month of seroconversion (HIV)
- Store data in appropriate format
  - The software needs to recognize the dates

Listing cases

Information about place in the line listing

- Requires coding
  - Province, district or other administrative unit
- Can be analyzed by the computer
  - ✘ 234 Defense Colony square
  - ✔ WARD variable: 23
  - ✔ BLOCK variable: 2
- Needs to relate to a population denominator
  - Denominator must be available for the level used
  - Identical coding will facilitate rates calculation

Listing cases

Information about persons in the line listing

- Requires coding
  - 1 for male, 2 for female
- Can be analyzed by the computer
- Needs to relate to a population denominator
  - Denominator must be available for the information examined
    - Recode age using age groups for which denominator information is available
  - Identical coding will facilitate rates calculation

Listing cases

Information about outcome in the line listing

- Requires coding
  - Death:
    1 for Yes, 2 for No, 9 for unknown
- May need to be updated regularly
  - Measles-associated death is a death within 30 days of rash onset
  - Information not available before 30 days
- Can be analyzed by the computer

Listing cases

Information about laboratory in the line listing

- Requires coding
  - Measles IgM:
    1 for Yes, 2 for No, 9 for unknown
- May be available at a later stage
- May be handled as a separate file that is electronically related to the line listing using the unique identifier

Listing cases

Steps of an outbreak investigation

1. Determine the existence of an outbreak
2. Confirm the diagnosis
3. Define a case
4. Search for cases
5. Generate hypotheses using descriptive findings
6. Test hypotheses based upon an analytical study
7. Draw conclusions
8. Compare the hypothesis with established facts
9. Communicate findings
10. Execute prevention measures

Aggregating cases

Count, divide and compare (CDC)

- Count
  - Aggregate the cases in the line listing by characteristic (e.g., age, sex, residence)
- Divide
  - Divide the number of cases by the relevant denominator
- Compare
  - Compare incidence across groups

Aggregating cases

Using time, place and person information to generate hypotheses

- Time
  - Epidemic curve
- Place
  - Spot map
  - Incidence by area
- Person
  - Incidence by age, sex, etc.
  - Trawling questionnaire of cases
  - Interview of outliers

Aggregating cases

Preparing the epidemic curve with a good line listing

1. Choose time unit
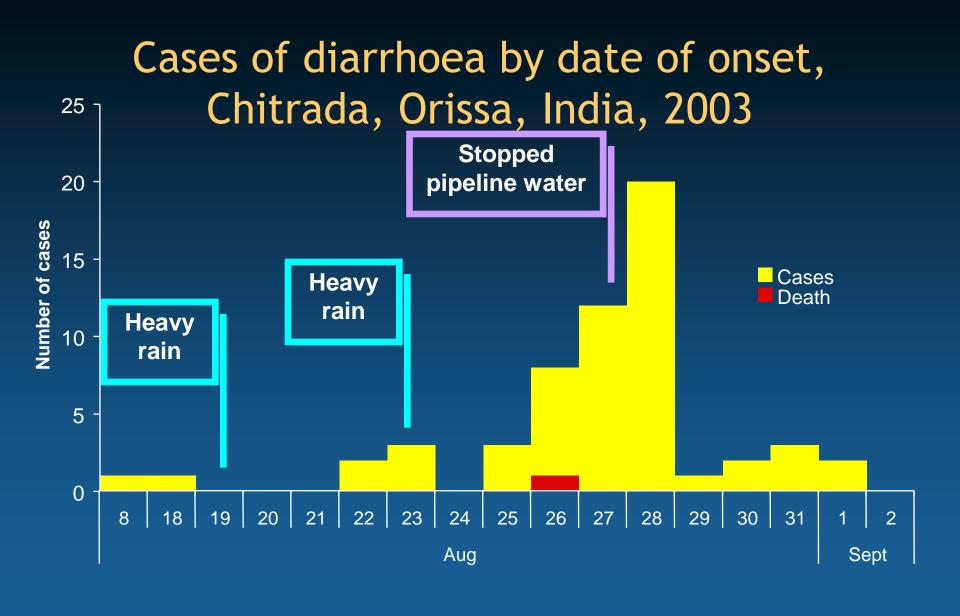   - Hours, days, weeks, month
   - Usually 1/4 of incubation period
2. Recode onset into desired time unit
   - Date of onset may need to be recoded in weeks, months etc..
3. Calculate the number of cases by time unit
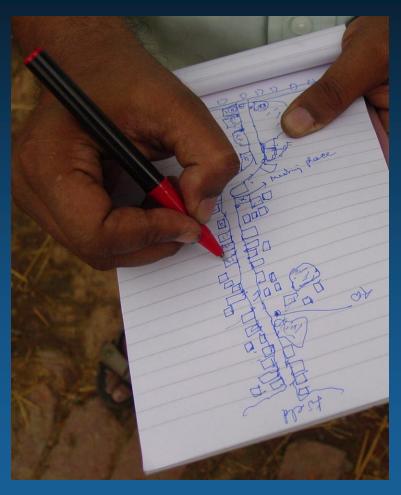   - Obtain an automated table with number of cases by time unit

   OR
   - Generate aggregated file with number of cases by time unit (Better)
4. Construct the epidemic curve

Aggregating cases

Cases of diarrhoea by date of onset, Chitrada, Orissa, India, 2003

PLACE: Drawing a map

- Spot map
  - Plot of cases (spot) on a rough map in the field
  - Does not account for population density
- Incidence by area
  - Calculate incidence by area (ward, block)
  - Draw map with colour scale proportional to the incidence

Aggregating cases

Linking the line listing and the population denominator by location

- ## Electronic line listing

Uni. ID	OnsetDate	Ward	Block	City	AgeYear	Sex	Hospital	Death	HEVIgM	HAVIgM
1	1-Mar-05	18	2	HYD	12	1	1	2	1	9
2	3-Mar-05	22	1	HYD	25	2	1	2	2	1
3	5-Mar-05	23	3	HYD	36	1	2	9	9	9
4	6-Mar-05	-	-	SEC	23	2	1	1	1	2

- ## Computerized population denominator file

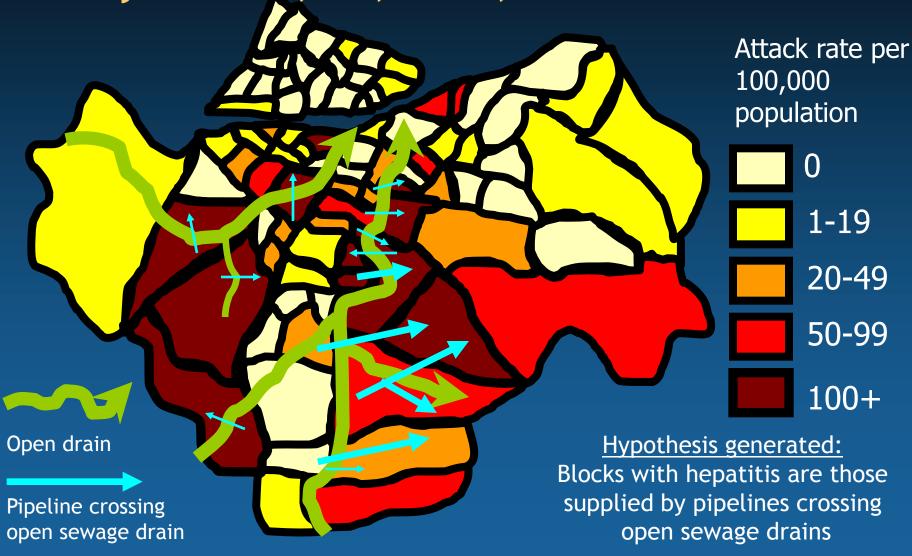| Ward	Block	2005 Population size
18	1	12345
18	2	45698
18	3	45632
18	4	1698

Aggregating cases

Preparing the calculation of incidences by geographical areas with a good line listing

1. Calculate the number of cases by area
   - Obtain an automated table with number of cases by area

   OR

   - Generate aggregated file with number of cases by area (Better)

2. Link with denominator
   - Manually enter number of cases by area in denominator file

   OR

   - Relate aggregated file to denominator file (Better)

3. Calculate rates automatically

Aggregating cases

Incidence by area

Incidence of acute hepatitis by block, Hyderabad, AP, India, March-June 2005

Attack rate per 100,000 population

0
1-19
20-49
50-99
100+

Open drain

Pipeline crossing open sewage drain

Hypothesis generated:
Blocks with hepatitis are those supplied by pipelines crossing open sewage drains

PERSON: Incidence by population groups

- Count
  - Count the cases in each age and sex group
- Divide
  - Obtain census denominators for each age and sex group
- Compare
  - Estimate the incidence for each:
    - Age group
    - Sex group

Aggregating cases

Preparing the calculation of incidences by individual characteristics with a good line listing

1. Calculate the number of cases with characteristic
   - Obtain an automated table with number of cases by age group, sex group etc..

   OR

   - Generate aggregated file with number of cases in each group (Better)

2. Link with denominator
   - Enter number of cases with characteristic in denominator file

   OR

   - Relate aggregated file to denominator file (Better)

3. Calculate rates automatically

Aggregating cases

Attack rate of measles by age and sex, Cuddalore, Tamil Nadu, India, 2004-2005

Characteristics		Number of cases	Population	Attack rate per 100,000
Age group	0-4	50	255,755	19.6
	5+	51	1,795,383	2.8
Sex	Male	48	1,032,938	4.6
	Female	53	1,018,200	5.2
Total		101	2,051,138	4.9

Aggregating cases

Take-home messages

- List cases using variables categorized so that they can be aggregated and related to a denominator

- Aggregate cases as needed to relate to the population denominator for the calculation of rates