Study design of an epidemiological research and sampling methodology for a risk analysis

-To ensure a risk analysis epidemiologically reliable-

Kohei Makita
BVSc, PhD. Email: K.makita@cgiar.org
Safe food, fair food project coordinator
International Livestock Research Institute (ILRI)

In Côte d’Ivoire, April 2009
A risk analysis and an epidemiological research

- Within a scope of safe food, fair food project
- A risk assessment using risk pathways
- Risk inputs are obtained from epidemiological researches
Safe food, fair food project

Building capacity to improve the safety of animal-source foods and ensure continued market access for poor farmers in sub Saharan Africa
Two approaches to Risk Assessment

Codex Alimentarius Committee

- Hazard identification
- Hazard characterization
- Exposure assessment
- Risk characterization

OIE

- Release assessment
- Exposure assessment
- Consequence assessment
- Risk assessment
Pathway maps

Cow
- T=2, C=0
- F=1.8, S=8.6
- W=0

Farm
- Tara
- Farm 1: 84 litres
- Other cows (7)

Path
- Vendor 1: 20 litres
- Co-operative: 50 litres

Consumer
- Self: 4 litres
- Near HH BC: 1 litre
- Near HH NKD: 0.5 litres
- Far HH YA: 1 litre
- Far HH B: 0.75 litres
- 25 other households: 16.75 litres

Households: 5/6
- 10 litres

Key
- Total plate count: Standard < 5,000
- Coliform count: Standard: <0.1
- Fat: Standard 3.5
- Solids not fat: Standard: 8.5
- Added water

Possible Critical Control Point (CCP)
## Risk mitigation

<table>
<thead>
<tr>
<th>Hazard Transmission</th>
<th>Risk mitigation strategies currently practiced (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecosystem to cow</td>
<td></td>
</tr>
<tr>
<td>Keep only one species</td>
<td>29% Treat cattle often 31%</td>
</tr>
<tr>
<td>Zero-graze</td>
<td>38% Don't keep calves 39%</td>
</tr>
<tr>
<td>Use own land only for feed</td>
<td>41% Use Artificial insemination 44%</td>
</tr>
<tr>
<td>Avoid common grazing</td>
<td>56% Vaccinate against brucellosis 1%</td>
</tr>
<tr>
<td>Keep local breeds</td>
<td></td>
</tr>
<tr>
<td>Milk shed to cow</td>
<td></td>
</tr>
<tr>
<td>Use feed/water trough</td>
<td>94% Stack manure 11%</td>
</tr>
<tr>
<td>Have concrete/stone floor</td>
<td>96% Have a waste disposal strategy 96%</td>
</tr>
<tr>
<td>Use bedding</td>
<td></td>
</tr>
<tr>
<td>Milk shed / dairy to milk</td>
<td></td>
</tr>
<tr>
<td>Have washable shed wall</td>
<td>100% Use just metal/ glass vessels 19%</td>
</tr>
<tr>
<td>Have metal/tin roof</td>
<td>96% Use piped water 75%</td>
</tr>
<tr>
<td>Store containers off floor</td>
<td>29% Keep premises clean 51%</td>
</tr>
<tr>
<td>Keep milk bar dry</td>
<td>45% Depose waste &gt;5m away 38%</td>
</tr>
<tr>
<td>Milk handler to milk</td>
<td></td>
</tr>
<tr>
<td>Use hot water to clean</td>
<td>18% Have no discharges/ wounds 97%</td>
</tr>
<tr>
<td>Use soap to clean</td>
<td>81% Have clean hands 79%</td>
</tr>
<tr>
<td>Wear protective clothing</td>
<td>1% Have clean/short nails 81%</td>
</tr>
<tr>
<td>Wash hands with soap before handling milk</td>
<td>59% Access to latrine 98%</td>
</tr>
<tr>
<td></td>
<td>Good personal hygiene 49%</td>
</tr>
<tr>
<td>Transport to milk</td>
<td></td>
</tr>
<tr>
<td>Don't drink unsold milk</td>
<td>10% Don't sell/store unsold milk 90%</td>
</tr>
<tr>
<td>Milk to consumer</td>
<td></td>
</tr>
<tr>
<td>Treat milk</td>
<td>50% Sell milk quickly (=6 hrs) 82%</td>
</tr>
<tr>
<td>Avoid drinking raw milk</td>
<td>93% Don't consume milk until 64% withdrawal period passed</td>
</tr>
<tr>
<td>Check milk quality by smell/taste</td>
<td>48%</td>
</tr>
</tbody>
</table>
Figure 4: Fault tree (simplified) of human illness as the result of exposure to C. parvum (type 2) in Dagoretti, Nairobi
Participatory epidemiology

- Instrument
  - Fill the gap where there is no data
  - Decrease opposition to decisions
  - More efficient and effective than direct regulatory control

- End
  - Empowerment
  - Ownership
  - Subsidiarity
  - A right
An example of participatory risk assessment

- In village A, chicken are typically cooked and eaten at 12.00 am. Most is eaten immediately but some is kept for the evening meal. There is concern that this might be a disease risk. How long can the women safely store the food?

<table>
<thead>
<tr>
<th>Risk inputs</th>
<th>Distribution</th>
<th>Units</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial number of bacteria: B2</td>
<td>Poisson</td>
<td>cfu/g</td>
<td>Mean = 0.95 cfu per gram</td>
</tr>
<tr>
<td>Doubling time (hours): B3</td>
<td>Normal</td>
<td>Hours</td>
<td>Mean = 0.5, s.d = 0.02</td>
</tr>
<tr>
<td>Storage time (hours): B4</td>
<td>Triangular</td>
<td>Hours</td>
<td>Low = 4, high = 12, mode = 6</td>
</tr>
<tr>
<td>Number of bacteria at the end of storage time</td>
<td>cfu/g</td>
<td>=B2*2^(B4/B3)</td>
<td></td>
</tr>
</tbody>
</table>

Model: B2*2^(B4/B3)

Source: Grace D (2008) PRA training material
Risk inputs

- **Types of risk inputs**
  - Quantity of products – e.g. milk
  - Proportion – products, infection rate, pathway
  - Biological data – doubling time of bacteria
  - Counting data – coliform count
  - Probability distribution

- **Getting risk inputs – epidemiology**
  - Experimental studies
  - Observational studies

- **Valid risk inputs should represent the target population**
Study design of an epidemiological research
Definition of epidemiology

- Epidemiology is the study of disease in populations and of factors that determine its occurrence

Thrusfield M 2005, Veterinary Epidemiology
Types of observational studies

- **Cohort studies**
  - Changes over time
  - Smoking and cancer

- **Case-control studies**
  - Diseased and non-diseased animals

- **Cross sectional studies**
  - Prevalence or incidence at a time
Measure of association

<table>
<thead>
<tr>
<th></th>
<th>Diseased animals</th>
<th>Non-diseased animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

**Incidence:**
The number of *new cases* that occur in a known population over a specified period of time

**Prevalence:**
The number of *instances of diseases or related attributes* (e.g., infection or presence of antibodies) in a known population, at a designated time, without distinction between old and new cases

Source: Thrusfield M (2005) Veterinary Epidemiology 3rd Ed.
Measure of association - 1

<table>
<thead>
<tr>
<th></th>
<th>Diseased animals</th>
<th>Non-diseased animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Relative risk: RR
The ratio of the incidence of disease in exposed animals to the incidence in unexposed

\[
RR = \frac{a/(a+b)}{c/(c+d)}
\]

Source: Thrusfield M (2005) Veterinary Epidemiology 3rd Ed.
Measure of association - 2

Odds ratio: OR
The ratio of odds: ratio of the probability of an event occurring to the probability of it not occurring

<table>
<thead>
<tr>
<th></th>
<th>Diseased animals</th>
<th>Non-diseased animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

OR = \frac{ad}{bc}

Source: Thrusfield M (2005) Veterinary Epidemiology 3rd Ed.
Sampling methods and sample size
Sampling methods

- Non-probability sampling methods
  - Convenience sampling
  - Purposive selection

- Probability sampling methods
  - Simple random sampling
  - Systematic sampling
  - Stratified random sampling
  - Cluster sampling
  - Multistage sampling

Source: Thrusfield M (2005) Veterinary Epidemiology 3rd Ed.
Convenience sampling
Purposive selection

Sample size 5/ population size 17
Simple random sampling

Sample size 5/ population size 17
Systematic sampling

Random \downarrow 4th \downarrow 4th \downarrow 4th \downarrow 4th

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Sample size 5/20, sample interval 4
Stratified random sampling

Sample size 7/35 cows

Strata

Sampling units

Proportional allocation

1/5 2/10 3/15 1/5
Cluster sampling

Clusters = Primary sampling units

Sample size 2/4 farms

Unit of concern

?  ?  ?  ?
Cluster sampling

Sample size 2/4 farms

Clusters= Primary sampling units

Unit of concern

Sample all cows

5

15

?
Multistage sampling 1

Case 1: the herd size is not known ahead of time

Primary sampling units

Secondary units

Constant proportion

1/5

3/15
Multistage sampling 2

Case 2: the herd size is known ahead of time

Primary sampling units

Probability proportional selection

Secondary units

Fixed number

3/10

3/15
Sample size calculation

- Note 1: the formulae are different among cross sectional, case-control and cohort studies
- Note 2: the formulae are different also between random sampling and cluster sampling
- Note 3: again different among estimating prevalence, comparing means, medians and proportions of two groups, and disease detection

What you learn here are sample size calculation for
- Cross sectional study, (1) random sampling and (2) cluster sampling for estimating prevalence
- Detection of a disease
- Using a perfect test
Sample size calculation for cross sectional study, random sampling

\[ n = \frac{1.96^2 \times P_{exp}(1-P_{exp})}{d^2} \]

Where:  
- \( n \) = required sample size  
- \( P_{exp} \) = expected prevalence  
- \( d \) = desired absolute precision

Software to recommend
- Winepiscope 2.0 [Thrusfield et al., 2001]  
  http://www.clive.ed.ac.uk/winepiscope
- Epi Info [CDC., 2008]  
  http://www.cdc.gov/epiinfo

Source: Thrusfield M (2005) Veterinary Epidemiology 3rd Ed.
Sample size calculation for cross sectional study, one-stage cluster sampling

\[ V_c = c \left\{ \frac{K_1 c V}{T^2 (c-1)} - \frac{K_2 P(1-P)}{T} \right\} \]

Where: 
- \( c \) = number of clusters in the sample 
- \( T \) = total number of animals sampled 
- \( K_1 = (C-c)/C \) 

Where: 
- \( C \) = number of clusters in the population 
- \( K_2 = (N-T)/N \) 

Where: 
- \( N \) = total number of animals in the population 
- \( V = P^2(\Sigma n^2) - 2P(\Sigma nm) + (\Sigma m^2) \) 

Where: 
- \( P \) = sample estimates of overall prevalence 
- \( n \) = number of animals sampled in each cluster 
- \( m \) = number of diseased animals samples in each cluster

Software to recommend R 
http://www.r-project.org

Source: Thrusfield M (2005) Veterinary Epidemiology 3rd Ed.
Sample size calculation for cross sectional study, sampling for disease detection

\[ n = \left\{1-(1-P_1)^{1/d}\right\}\{N-d/2\}+1 \]

Where:  
\( n \) = required sample size  
\( N \) = population size  
\( P_1 \) = probability of finding at least one case in the sample  
\( d \) = minimum number of affected animals expected in the population

Software to recommend

Freecalc [Cameron and Baldock, 1998]  

Source: Thrusfield M (2005) Veterinary Epidemiology 3rd Ed.
Please note!

- These are the calculations using a perfect test (sensitivity and specificity = 1.0)!

- For imperfect test (sensitivity and specificity < 1.0), refer textbooks and use a software.
Number of animals to be sampled per herd

\[ n_I = \sqrt{\frac{\sigma^2_I}{\sigma^2_H} \times \frac{c_H}{c_I}} \]

- \( n_I \): number of animals to be sampled per herd
- \( \sigma^2_H \): between herd variance estimates
- \( \sigma^2_I \): within herd variance estimates
- \( c_H \): cost of sampling herds (e.g., fuel)
- \( c_I \): cost of sampling individuals (e.g., ELISA test)

Source: Dohoo et al. 2004 Veterinary Epidemiologic Research