Assessing the risk of illness from food-borne pathogens – some thoughts

Alicia Carriquiry - Iowa State University

Workshop on Overaching Issues in Risk Analysis
Ames, October 27-29, 2005

Outline

- Farm-to-fork models:
  - Event models with many nodes
  - Scarce data and selective sampling
  - Noise and measurement error
- Example: $Salmonella$ in Finnish beef cattle, Ranta et al., 2005.
- Interesting statistical issues in exposure assessment.

The state of the art

- The cost of foodborne illnesses is very high. Focus on food safety in recent years.
- Monte Carlo simulations (using software such as @Risk) as a tool to account for uncertainties in the value of risk model parameters.
- There is research on specific model components (e.g., Mosier and Craig, earlier talk).
- Hierarchical models fitted within a Bayesian framework have recently been proposed and have promise.
- Some excellent work recently published by Ranta and others at National Veterinary and Food Research Institute, Finland.

Challenges

- Scenario pathways and event trees often used to model risk.
- A farm-to-fork model can be very extensive and include:
  - Food production component
  - Distribution/storage component
  - Preparation/consumption component (exposure).
- Each component, in turn, may be composed of many possible events.
- Within each component, we need to know:
  - What can go wrong (events).
  - What is the (conditional) probability of each event
  - What are the consequences of each event.

Challenges (cont'd)

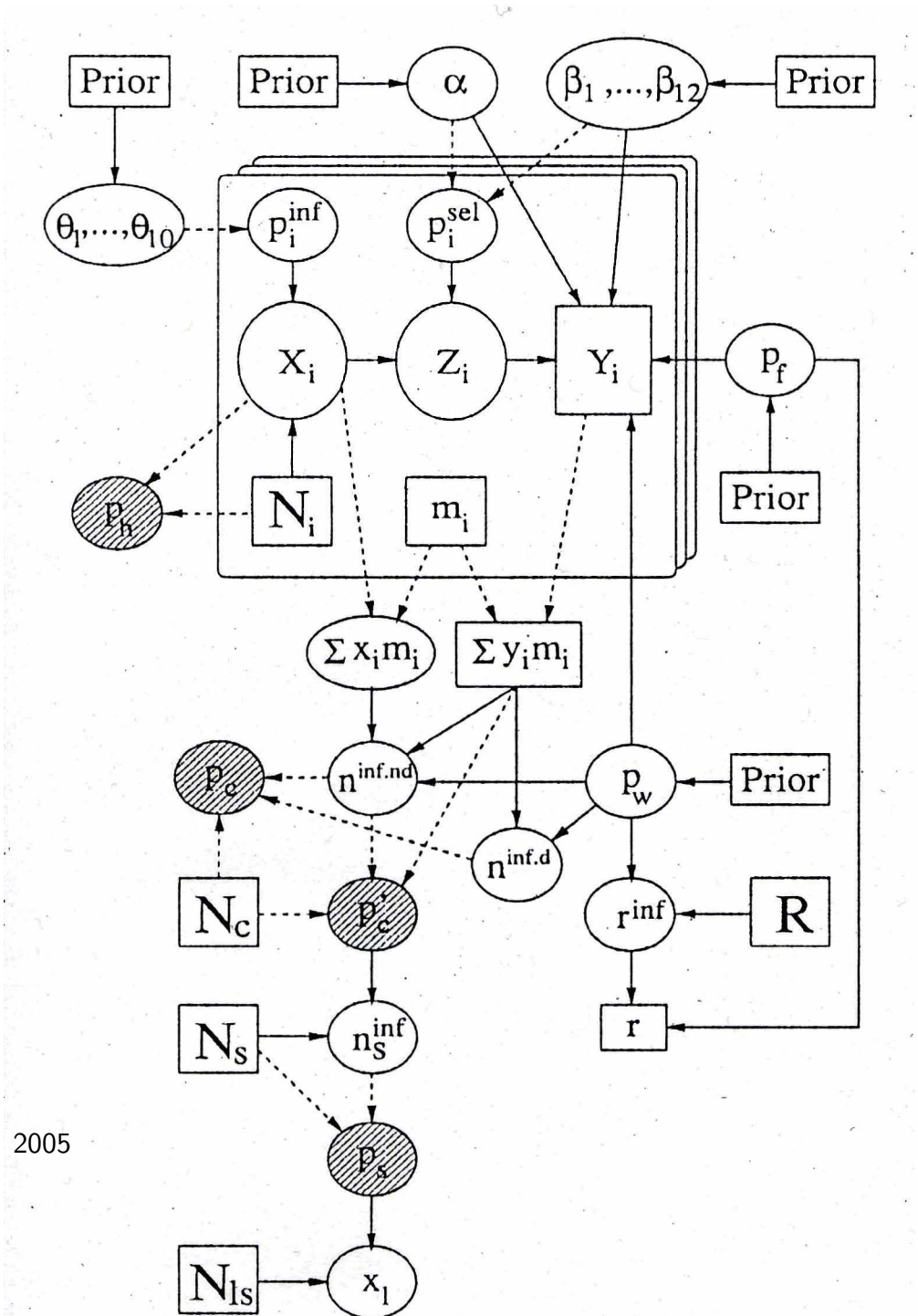
- As an example, estimating human exposure to *Salmonella* from contaminated eggs in the home requires knowledge of:
  - Probability that a purchased egg will be contaminated (during production, transportation or storage).
  - Recipes of foods and beverages that include raw or undercooked eggs.
  - Usual consumption, by age groups, of each of those foods and other food preparation information.
  - Distribution of likely doses of the organism consumed. Depends on initial contamination, food preparation, contamination in the home, and other.
  - Probability of illness as a function of dose. Varies across individuals and across time within individuals.

Event tree models

- Can be useful to estimate the probability of an end-event occurring. An end-event is, for example, illness in the population.
- What/if scenarios can be tested: how is risk reduced if certain policies or regulations are implemented?
- Often, risk estimates are critically sensitive to estimated probabilities of intermediate events in the process. Scarce data are available for estimation.
- Dependencies among tree branches can be overlooked. Multiplying probabilities of different events implies independence and can lead to unrealistically low risk estimates.
- Risks are difficult to estimate precisely, but *relative risks* often useful.
- Example: Ranta et al., *Risk Analysis*, 2005.

Salmonella in cattle in Finland

- We illustrate complexity showing just a few of the steps in the model.
- Objective: estimate prevalence of *Salmonella* in live and slaughtered cattle in Finland.
- Multi-step model:
  1. First estimate prevalence in slaughtered animals with no information from live herds
  2. Second, combine herd and animal-level models.
- Made use of animal-level data collected in abbatoirs, herd-level data collected in each municipality and national (aggregated) data.


Prior
Prior
α
β1 ,...,β12
Prior
θ1,...,θ10
p_i^inf
p_i^sel
X_i
Z_i
Y_i
p_f
p_h
N_i
m_i
Prior
Σ x_i m_i
Σ y_i m_i
p_c
n^inf.nd
p_w
Prior
n^inf.d
N_c
p'_c
r^inf
R
N_s
n_s^inf
r
p_s
N_ls
x_l

Slaughtered animal model

- x is true # of infected animals, y is # infected in N tested, p_s is true prevalence in slaughtered animals, p_l is sensitivity of test.

$$y|x, p_l \sim \text{Bin}(x, p_l),\ x|N, p_s \sim \text{Bin}(N, p_s),$$

and with uniform priors on (p_l, p_s),

$$\pi(p_s, p_l, x|y, N) \propto \binom{N}{x} p_s^x (1-p_s)^{N-x} \binom{x}{y} p_l^y (1-p_l)^{x-y} \pi(p_s, p_l).$$

- Data collected from herds is then used to better determine $\pi(p_s)$.
- Literature and expert opinion for choosing $\pi(p_l)$.

Combining herd and animal models

- Ranta et al. estimated prevalence at three levels:

  1. p_h: prevalence in population of herds.
  2. p_c: prevalence among live animals.
  3. p_s: prevalence among slaughtered animals.

- To estimate p_h, used posterior predictive approach. Given observed number of infected herds, and total number of herds in 437 areas, derived posterior distribution of probability of infection θ_i for ith area. If $p_h =$ infected/total then

$$\pi(p_h|y) = \sum_i (N_i)^{-1} \int \pi(x_i|\theta_i)\pi(\theta_i|y_i)d\theta_i.$$

Detecting infected herds

- Observed number of positive herds is modeled as $y_i \sim$ Bin$(z_i, p_i^{h.sen})$.

- Probability of actually detecting infected herds depends on:

  1. Probability that an infected herd gets tested. Need to distinguish between herds that show clinical symptoms and those that do not. Estimate z_i, the number of infected herds tested.
  2. Probability that a tested infected herd gets positive results. Depends on: sensitivity of test, within-herd prevalence, number of tested animals within each herd.

- Next need to derive a model for $p_i^{h.sen}$, the overall sensitivity of testing method.

Detection (cont'd)

- Consider, for example, estimating the number z_i of infected herds in the ith region that are tested for $Salmonella$.

- Sampling schemes may be non-standard: herds with clinical symptoms sampled with higher probability than herds exhibiting no symptoms. Thus, estimate of z_i depends on the probability that infected herds are tested. If p_i^{sel} is probability that an infected herd gets tested, then

$$\begin{aligned} z_i|p_i^{sel}, x_i &\sim \text{Bin}(x_i, p_i^{sel}) \\ p_i^{sel} &= \Pr(\text{CS}|\text{infected}) + \Pr(\text{NCS}|\text{infected}) \end{aligned}$$

  where $\Pr(\text{CS}|\text{infected})$ is probability of testing based on clinical symptoms given that herd is infected.

Next: sensitivity of the test

- $p_i^{h.sen}$ may depend on reasons for conducting the test:
  - If herd shows CS, symptomatic animals are tested and then $p_i^{h.sen} = p_f$, the 'lab' sensitivity.
  - If testing is not due to CS, then a random sample of k animals are chosen and samples are pooled. Here,

$$p_i^{h.sen} = \sum_k (1 - (1 - p_{wi})^k) p_f \ \Pr(k),$$

    with p_{wi} the within-herd prevalence.

- Latter assumes that sensitivity of test is the same on single specimens and on pooled samples.

Why is Ranta model attractive?

- Model is (partially) comprehensive. Complete formulation involves several additional steps.

- Noteworthy is
  - Careful description of events and their probabilities at each step
  - Accounting for most (all?) of the factors that may affect the risk estimate
  - Hierarchical formulation of model that permits accommodating dependencies.

- Model is not farm-to-fork, transportation/storage and exposure components missing.

Other challenges: exposure step

- Assessing risk may require estimation of exposure to the hazard. E.g., how much pesticide from apples do children consume?

- Gross simplifications are often used: 'On the average, an apple has X mg of pesticide and the average child 4 - 8 years of age consumes 0.18 apples per day'. Tails are important!

- There is a distribution of pesticide content in apples and of usual apple consumption among children 4 - 8 and the mean (or median) is typically not a good summary of the distribution.

- Risk (most exposed) to pesticides in apples may depend on ethnic group, socio-economic status, region.

Exposure (cont'd)

- Data for estimating distribution of usual apple consumption consist of one or two observations of daily intake obtained from nationwide food consumption surveys.

- Must estimate distribution of probability of consumption of appleas among children and, conditional on consumption, amount consumed.

- For many foods, probability of consumption and amount consumed are not independent.

To conclude...

- Estimating the risk of end-events in the area of food safety typically requires large models with lots of nodes.

- Estimating the probabilities of events at the nodes can be difficult; see Mosier and Craig presentation and Ranta et al. publication.

- Many interesting statistical challenges:
  - Estimation of probabilities of rare events
  - Estimation based on adaptive and/or selective sampling
  - Combining data taken at different levels of aggregation and expert opinion.
  - Joint or marginal estimation to account for dependencies.
  - Calibrating and validating risk models in the presence of little or no data.