

Estimating four different prevalence values of pig carcasses



- one for each pig-herd *Salmonella* level - when carcass swabs are pooled independently of original herd level.

Part of a risk assessment on multiresistant *Salmonella* DT104 in slaughter pigs

by Helle M. Sommer,

Søren Aabo, Bjarke B. Christensen, Peter Saadby,
Niels L. Nielsen, Birgit Nørrung and Danilo Wong

Purpose of the risk assessment



Estimate the relative changes in the number of MRDT104 carcasses due to proposed changes in the trade with pigs.

We compared the situation under *restricted trade* (at that time 'present') with the situation under *continued trade* (proposed future).

Salmonella is used as a surrogate for *Salmonella* Typhimurium DT104

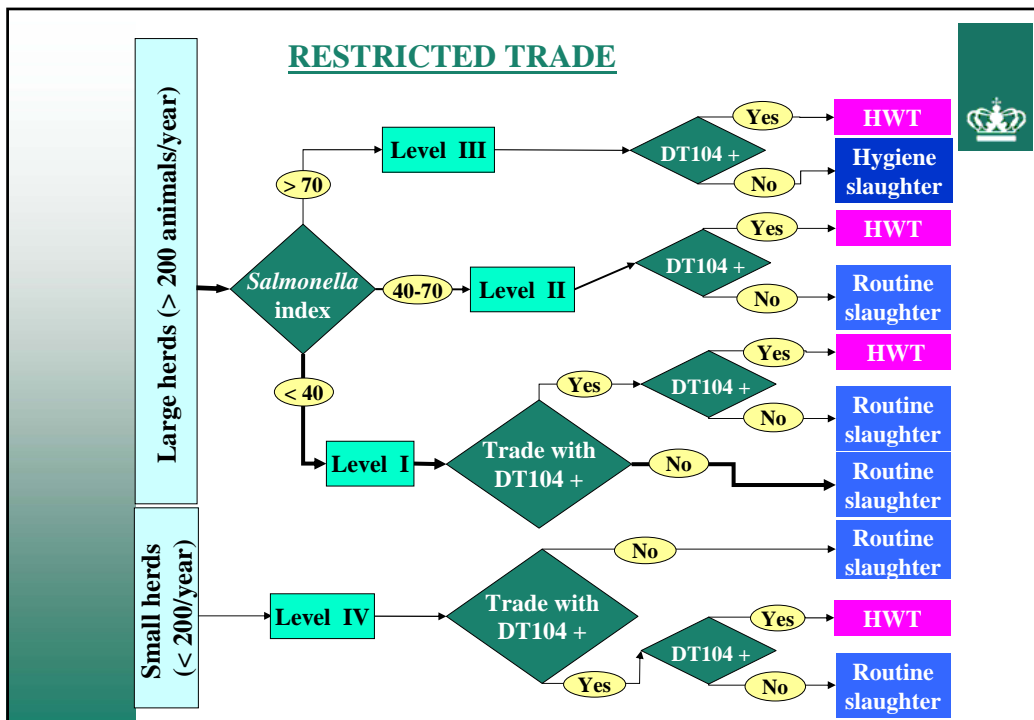
Focus of this presentation



Estimation of the *Salmonella* prevalence $p_{i,S,A}$ after slaughtering, for 4 different herd levels. The prevalence values are estimated from pooled data.

Why?

- estimating the transmission factors T_{ij}
- used in 'our' black box model of the slaughter-house process
- applied a black box - no interest in improving the processes - relative changes in the output



Black box model, restricted trade



Primary production

After slaughtering

No. of *Salm.* pos. animals · T_{ij} = No. of *Salm.* pos. carcasses

No. of DT104 pos. animals · T_{ij} = No. of DT104 pos. carcasses

T_{ij} is a transmission factor for Level i ($i = I, II, III, IV$) and slaughtering j (routine slaughter or slaughtering with HWT).

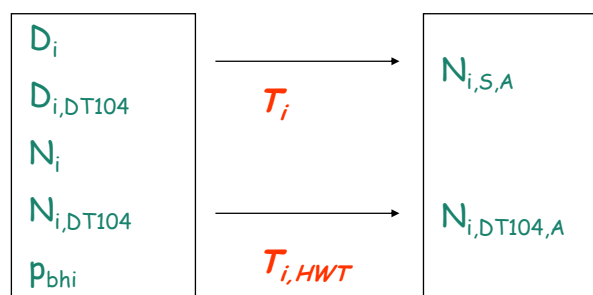
All together 8 transmission factors are estimated.

Black box model, restricted trade



Primary production

After slaughtering



$D_{i(DT104)}$ Average delivery /(DT104) herd/month in Level i

$N_{i(DT104)}$ Average no. of (DT104) herds/month in Level i

p_{bhi} Between-herd-prevalence in Level i

$N_{i,S,A}$ Average no. of *Salm.* pos. carcasses after slaughter/month in Level i

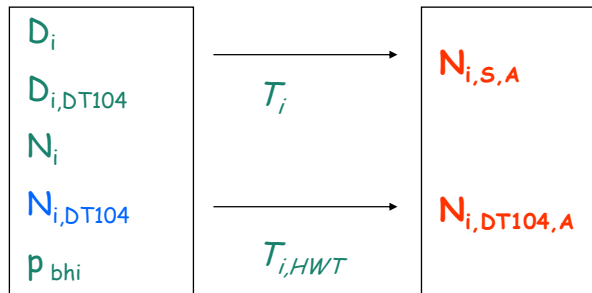
$N_{i,DT104,A}$ Average no. of DT104 pos. carcasses after slaughter/month in Level i

Black box model, continued trade



Primary production

After slaughtering



$D_{i(DT104)}$ Average delivery /(DT104) herd/month in Level i

$N_{i(DT104)}$ Average no. of (DT104) herds/month in Level i

p_{bhi} Between-herd-prevalence or relative no. of infected herds in Level i

$N_{i,S,A}$ Average no. of *Salmonella* pos. carcasses after slaughter/month in Level i

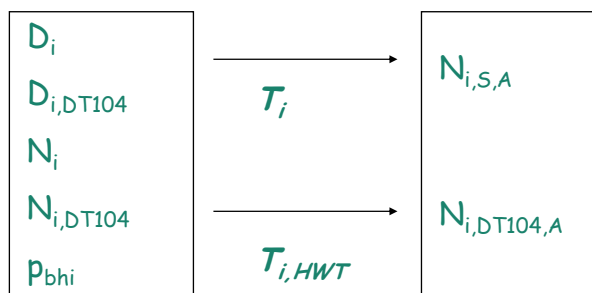
$N_{i,DT104,A}$ Average no. of DT104 pos. carcasses after slaughter/month in Level i

Estimating Transmission factors



Primary production

After slaughtering



No. of *Salmonella* pos. animals $\cdot T_{ij}$ = No. of *Salmonella* pos. carcasses

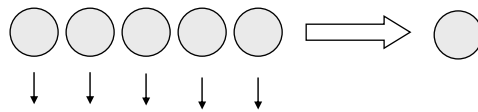
$$p_{bh,i} \cdot N_i \cdot D_i \cdot T_i = N_{i,S,A} = P_{i,S,A} \cdot N_i \cdot D_i$$

$$\Rightarrow T_i = P_{i,S,A} / p_{bh,i}$$

Estimating prevalence values



In average one individual swab sample is taken for every 600 carcasses. All together there are more than 6,000 pooled samples. One pooled sample consist of 5 individual samples from different herd levels. For example:

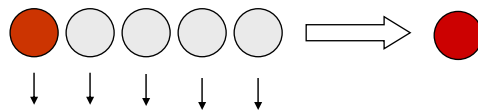


Level: II I I I IV

Binomial distributed with different probability of being positive depending on which level they are drawn from.

$$X_i \in B(n_i, p_i)$$

Estimating prevalence values



Level: I I I I I

Red is positive and p_I is the probability of a sample to be positive when the swab sample comes from a Level I herd. The binomial frequency function is:

$$\begin{aligned} P\{X=1|n=5\} &= p_I \cdot (1-p_I) \cdot (1-p_I) \cdot (1-p_I) \cdot (1-p_I) \cdot (n,x) \\ &= p_I \cdot (1-p_I)^4 \cdot (n,x) \end{aligned}$$

$$P\{X=2|n=5\} = p_I^2 \cdot (1-p_I)^3 \cdot (n,x)$$

$$p_{pool,5I} = P\{X=1|n=5\} + P\{X=2|n=5\} + P\{X=3|n=5\} + P\{X=4|n=5\} + P\{X=5|n=5\}$$

$$= 1 - P\{X=0\} = 1 - (1-p_I)^5$$

Estimating prevalence values



Estimating p_I from $p_{pool,5I}$ is simple:

$$p_I = 1 - (1 - p_{pool,5I})^{1/5}$$

5,130 pooled samples out of 6,051 were pure Level I samples, where 243 were positive giving the prevalence :

$$p_I = 1 - (1 - 243/5,130)^{1/5} = 0.97\%$$

Estimating the prevalence for Level II, III and IV were more complicated. Very few or non consisted of pure Level II, III or IV pools respectively. A different estimation strategy was applied involving conditional Maximum Likelihood. The prevalence values were not estimated simultaneously, but one at a time.

$$p_I(5,130) \quad p_{II}(624) \quad p_{III}(149) \quad p_{IV}(148)$$

Estimating prevalence values



Pure Level II, III or IV pools were rare or did not exist. In combination with Level I there were several pooled samples. Estimating the prevalence for Level II could be based on pooled samples with:

$$5 \text{ II} \quad (n_{II}=5)$$

$$1\text{I}+4\text{II} \quad (n_I=1, n_{II}=4)$$

$$2\text{I}+3\text{II} \quad (n_I=2, n_{II}=3)$$

$$3\text{I}+2\text{II} \quad (n_I=3, n_{II}=2) \quad (Y=5, N=49)$$

$$4\text{I}+1\text{II} \quad (n_I=4, n_{II}=1)$$

$$p_{pool,I,II} = 1 - (1 - p_I)^{n_I} \cdot (1 - p_{II})^{n_{II}} \cdot (1 - p_{III})^{n_{III}} \cdot (1 - p_{IV})^{n_{IV}}$$

$$p_{pool,3I,2II} = 5/49 = 1 - (1 - p_I)^3 \cdot (1 - p_{II})^2$$

Estimating prevalence for Level II



5 II (n_{II}=5) (Y=0, N=3)

1I+4II (n_I=1, n_{II}=4) (Y=0, N=1)

2I+3II (n_I=2, n_{II}=3) (Y=0, N=11)

3I+2II (n_I=3, n_{II}=2) (Y=5, N=49)

4I+1II (n_I=4, n_{II}=1) (Y=33, N=560)

$$p_{pool,5II} = 0/3 = 1 - (1-p_{II})^5 \quad \rightarrow p_{II} = 0$$

$$p_{pool,1I,4II} = 0/1 = 1 - (1-p_I)^1 \cdot (1-p_{II})^4 \quad \rightarrow p_{II} = -2.4\%$$

$$p_{pool,2I,3II} = 0/11 = 1 - (1-p_I)^2 \cdot (1-p_{II})^3 \quad \rightarrow p_{II} = -6.5\%$$

$$p_{pool,3I,2II} = 5/49 = 1 - (1-p_I)^3 \cdot (1-p_{II})^2 \quad \rightarrow p_{II} = 3.8\%$$

$$p_{pool,4I,1II} = 33/560 = 1 - (1-p_I)^4 \cdot (1-p_{II})^1 \quad \rightarrow p_{II} = 2.2\%$$

Conditional Maximum Likelihood



$$L(p_i | p_{v<i}) = \prod_{j=1}^{m_j} f_{ij}(y_{ij}, p_i | p_{v<i})$$

Probability of the pooled sample to be positive

$$= \prod_{j=1}^{m_j} \left(\underbrace{1 - (1-p_I)^{n_{I,j}} \cdot (1-p_{II})^{n_{II,j}} \cdot (1-p_{III})^{n_{III,j}} \cdot (1-p_{IV})^{n_{IV,j}}}_{\text{Probability of the pooled sample to be negative}} \right)^{Y_{ij}}$$

Probability of the pooled sample to be negative

i = index for the level being estimated, j = index for the different pool combinations. p_i is found by maximizing $L(p_i)$. The maximum likelihood is conditional since other prevalence values for levels lower than i are fixed.

Conditional versus non-conditional



Conditional maximum likelihood

$$L(p_I) \quad L(p_{II} | p_I) \quad L(p_{III} | p_I, p_{II}) \quad L(p_{IV} | p_I, p_{II}, p_{III})$$

Estimating the prevalence values simultaneously from a maximum likelihood function:

$$L(p_I, p_{II}, p_{III}, p_{IV})$$

Simultaneously estimating of all 4 prevalence values requires an optimization function. The optimization is easiest done on the log-likelihood function.

$$\frac{\partial(\log L(p_I, p_{II}, p_{III}, p_{IV}))}{\partial p_I \partial p_{II} \partial p_{III} \partial p_{IV}} = 0$$

Conditional versus non-conditional



Conditional maximum likelihood

$$L(p_I) \quad L(p_{II} | p_I) \quad L(p_{III} | p_I, p_{II}) \quad L(p_{IV} | p_I, p_{II}, p_{III})$$

No. of pools
used in the
estimation

↓ ↓ ↓ ↓
5,130 624 149 148

Extra no. of
pools in the
data material

917 138 30 0

Part used in
the estimat.

85% 82% 83% 100%

Expecting small changes in the estimates when using a simultaneous estimation procedure - smallest for p_I and larger for p_{II} , then p_{III} and then p_{IV} .

Prevalence estimates



Prevalence	p_I	p_{II}	p_{III}	p_{IV}
Unadjusted <i>Salmonella</i> prevalence estimates	0.97 %	2.11 %	2.79 %	3.52 %
<i>Salmonella</i> prevalence estimates	1.76 %	3.84 %	5.07 %	6.40 %

Adjusted for the loss of sensitivity when pooling samples.

End of presentation
- time to wake up :o)

