

Tilastotiedettä ja elintarvikeriskien mallinnusta

Erikoistutkija , dos. Jukka Ranta, riskinarvioinnin tutkimusyksikkö, Evira

Bacteria are hitchhikers.



Don't Give Ribs a Tangle!



Transparency 1

Riskinarviointi on osa riskianalyysiä ja sen tavoitteena on tuottaa uutta tietoa riskinhallinnan tarpeisiin. Ennen riskinarviointia tehtävän riskiprofiilin tavoitteena on kerätä ja kuvata olemassa olevaa tietoa vaaroista ja niiden riskeistä. Riskinarviointi on eri tahojen eduista riippumatonta ja pyrkii tieteen kriteerit täyttävin menetelmin sellaisen objektiivisen tiedon tuottamiseen, jonka avulla on mahdollista tehdä riskinhallintapäätöksiä.

Riskinarvioinnin tutkimus- ja kehitystyö

Riskinarvioinnin tutkimusyksikön tutkimus- ja kehitystyö painottuu erityisesti käytettäviin arviointimenetelmiin ja niiden soveltamiseen elintarvikkeiden ja alkutuotannon (eläintaudit ja kasvinterveys) arvioinneissa. Tavoitteena on kehittää terveysriskien arvioinnin rinnalla myös taloudellisten vaikutusten arviointia yhteistyössä alan muiden toimijoiden kanssa.



12.11.2015

Riskinarvioinnin tutkimusyksikön henkilöstö elokuussa 2011.

Minkälaisia sovelluskohteita?

- **Yksittäiset tutkimusprojektit:** kuvaileva data-analyysi ja estimointi. Rajoittuu usein yhteen selkeään datasettiin.
 - Regressiomallit, aikasarjat, esiintyvyyksien ja pitoisuuksien estimointi, selittävien tekijöiden vaikutus, kahden tai useamman populaation vertailu, varianssikomponenttimallit.
- **Laajemmat riskinarviointihankkeet:** systeemin tai laajemman kokonaisuuden mallintaminen. Useita tietolähteitä, ei ennalta rajattua datasettiä.
 - Tuotanto- ja elintarvikeketjujen simulointimallit, hierarkkiset Bayesmallit.

<http://www.evira.fi/portal/fi/tietoa+evirasta/esittely/toiminta/riskinarviointi/>

Elintarvikeriskit: kemialliset ja mikrobiologiset

- **Kemialliset:** Saantiarviot. Pitoisuusmittaukset elintarvikkeista tai raaka-aineista. Kulutustietoja, esim. ruokapäiväkirja-aineistot. Saanti = pitoisuus × annoskoko. ”Akuutti” vs ”krooninen”. Raja-arvojen ylittyminen? Pitoisuudet eivät *usein* muutu ennen kulutusajankohtaa.
- **Mikrobiologiset:** Esiintyvyys ja pitoisuus saattavat muuttua olosuhteiden myötä (esim. säilytys- /kypsennyslämpötila). Prediktiiviset mikrobiologiset mallit. Tuotantoketjujen mallintaminen vaiheittain. Dataa yleensä ketjun aiemmista vaiheista → ennustettava kuluttajan riskiä.
- **Molemmat:** interventioiden vaikutus riskiin?

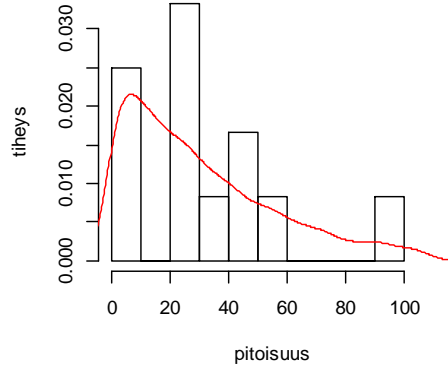
Tavoite: tuottaa tietoa päätöksenteon tueksi, esim. suosituksia varten, tai valvontaohjelmien suunnitteluun, tai EU-käyttöön.

Saantiarviot

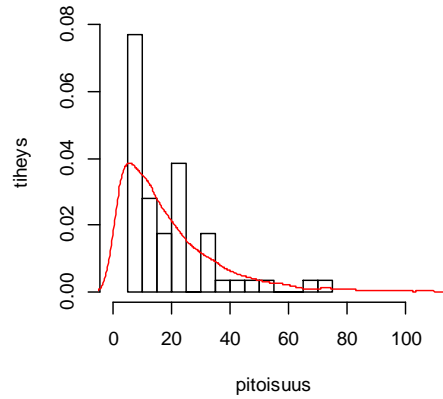
- **Useimmiten: kemiallisten aineiden saanti. Osaltaan sovellettavissa myös mikrobiologisiin.**
 - Suomalaisen aineiston mallintaminen. Pitoisuudet usein pieniä, aineistoa vaihtelevia määriä eri elintarvikkeista ja raaka-aineista.
 - Pitoisuuksien ja annoskokojen vaihtelut? Arvioon liittyvät epävarmuudet? Suuren pitoisuuden todennäköisyys? Harvinaisempien elintarvikkeiden kulutus? Yksilön riski? Populaation keskimääräinen riski?

Esimerkki: nitraatti

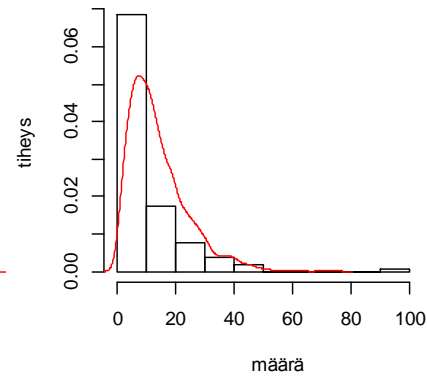
leikkelemakkara



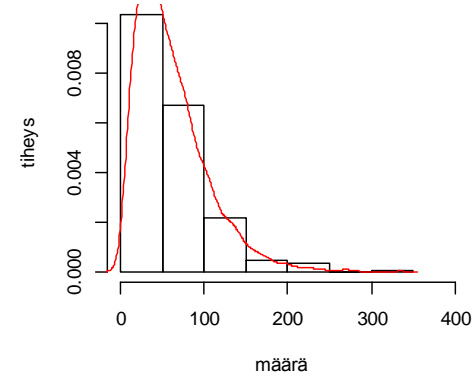
ruokamakkara



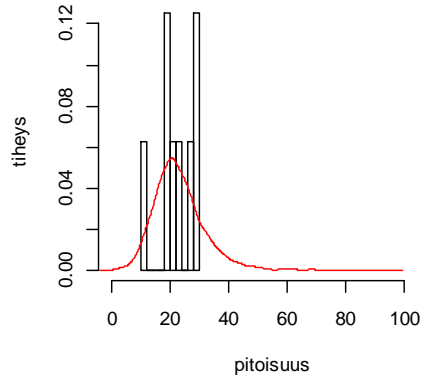
leikkelemakkara, 3v



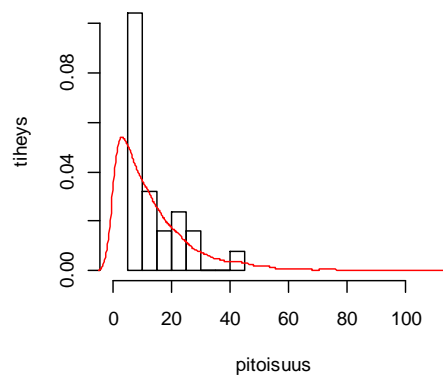
ruokamakkara, 3v



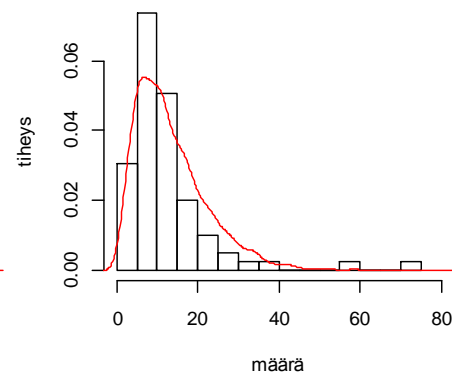
kalkkunaleikkele



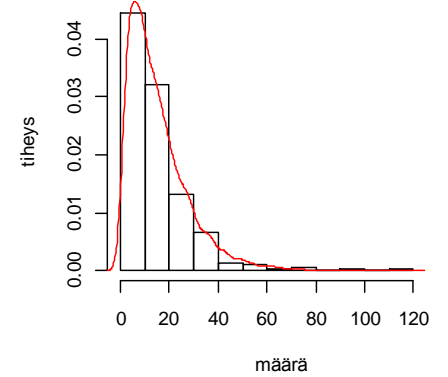
kinkkuleikkele



kalkkunaleikkele, 3v



kinkkuleikkele, 3v



Pitoisuuksien mallintaminen

- **Parametriset jakaumat:** pitoisuus on positiivinen luku. Log-normaalijakauma, gammajakauma, eksponenttijakauma, weibull,... sekoitejakaumat,...
- **Detektoraja:** Tavallisesti aineistossa on paljon pitoisuuksia jotka jäävät alle detektorajan, <LOQ. Todellisia nolliä vai $0 < x < \text{LOQ}$? Vasemmalta sensuroidut havainnot?

Pitoisuuksien mallintaminen

- **Esim. normaalijakauma log-pitoisuuksista**
→ parametrien (μ, σ) estimointi.

$$\begin{aligned} P(\text{DATA} \mid \mu, \sigma^2) &= P(x_1, \dots, x_n, x_{n+1} < L_1, \dots, x_{n+k} < L_k \mid \mu, \sigma^2) \\ &= \prod_{i=1}^n N(x_i \mid \mu, \sigma^2) \prod_{j=1}^k F(L_j \mid \mu, \sigma^2) \end{aligned}$$

- **Uskottavuusestimointi:** etsi parametriarvot, joilla $P(\text{DATA} \mid \mu, \sigma^2)$ maksimoituu, ja laske luottamusväli.
- **Bayes-estimointi:** valitse priorijakauma $p(\mu, \sigma^2)$ ja laske posteriorijakauma Bayesin kaavasta:
 $p(\mu, \sigma^2 \mid \text{DATA}) = p(\text{DATA} \mid \mu, \sigma^2)p(\mu, \sigma^2)/\text{vakio}.$

Pitoisuuksien mallintaminen

- **Bayes-mallien toteutus:** yleensä laskennallinen (Monte Carlo-menetelmät), esim. käyttäen OpenBUGSia.
- **Parametrien epävarmuusjakaumalle käyttöä:** predikttiiviset tulokset joissa huomioitu malliparametrien epävarmuus. Input-jakaumiksi muihin simuloitimalleihin, Bayes-verkot, 2D-simulointi.
- **Epävarmuuksien laskenta:** aleatorinen ja episteeminen epävarmuus
 - Aleatorista: $p(x | \mu, \sigma^2)$ \leftrightarrow Episteemistä: $p(\mu, \sigma^2 | \text{DATA})$
 - Esim. pitoisuuksien x populaatiojakauman 99%:n kvantiilipiste. Tämä on vakio jos μ, σ^2 tunnetaan. Kuvaa x :n vaihtelujakaumaa, tietyillä parametreilla.
→ epävarmuusjakauma kvantiilipisteelle seuraa parametrien epävarmuusjakaumasta.
 - Esim. ennuste satunnaiselle kuluttajalle: aleatorinen + episteeminen epävarmuus.

Mikrobien pitoisuuksia

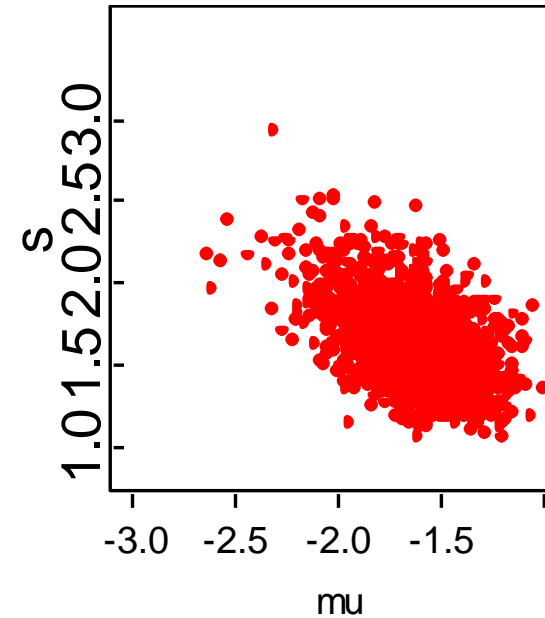
Esimerkki: Listeria

- Tarkkoja pitoisuuksia, vasemmalta & oikealta sensurointi, intervallisensurointi.
- Listeria monocytogenes kaupan savustetussa kalassa **Belgiassa** 2005-2007.
 - $N(x | \mu, \sigma^2)$ -jakaumamalli \log_{10} -arvoille.
 - $p(\mu, \sigma^2 | \text{DATA})$
 - $p(\text{ennustettu } x | \text{DATA})$
 - = $\iint N(x | \mu, \sigma^2) p(\mu, \sigma^2 | \text{DATA}) d\mu d\sigma^2$

P Busschaert, AH Geeraerd, M Uyttendaele, JF Van Impe. Estimating distributions out of qualitative and (semi)quantitative microbial contamination data for use in risk assessment. International Journal of Food Microbiology. 138 (2010), 260-269.

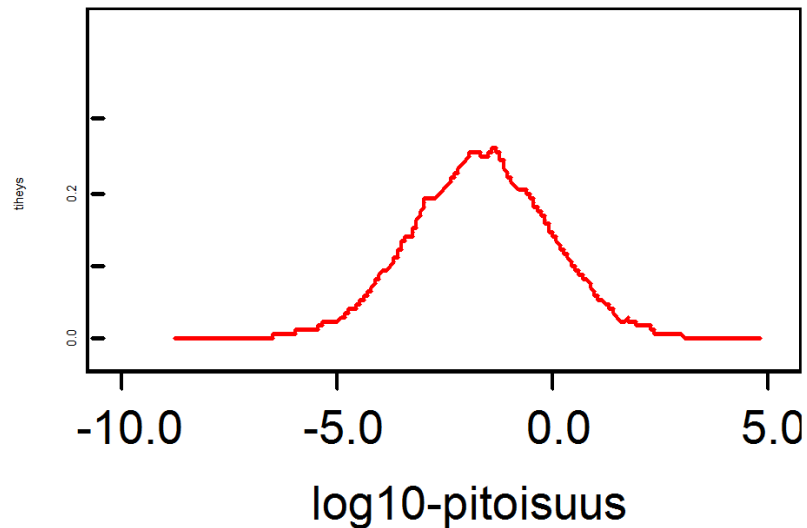
Data voi olla tällainen:	
Näytteiden määrä	Pitoisuus (cfu/g)
54	<0.04
2	<100
26	0.04-10
1	15
8	0.04-100
2	>100
1	<1
1	>1
7	0.04-1
1	1-100

- Parametrien (μ, σ) jakauma:

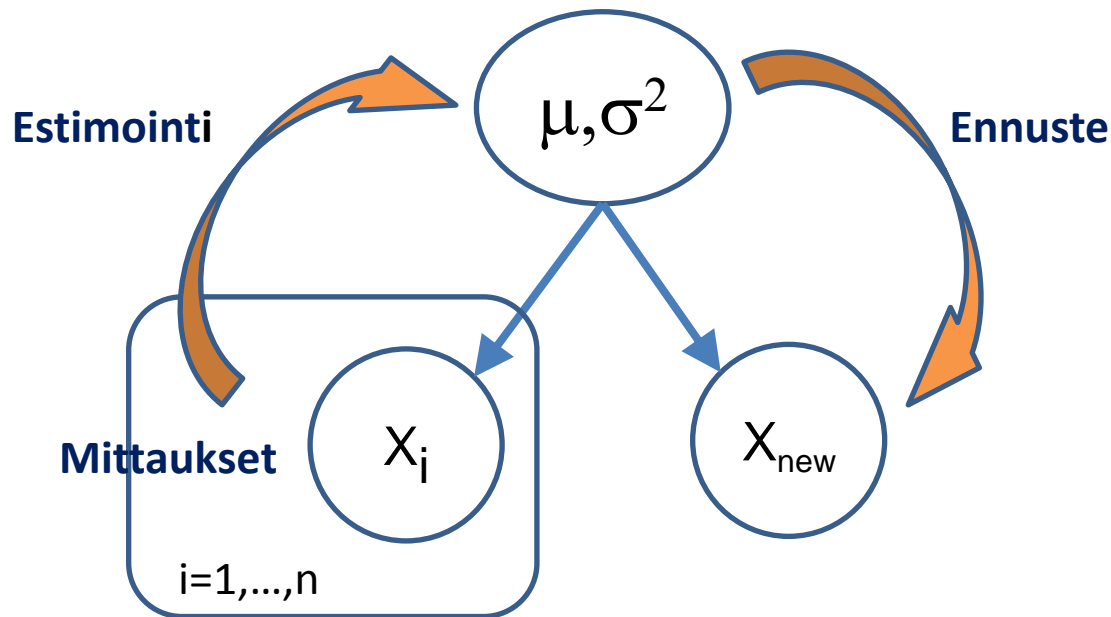


- Ennuste satunnaiselle annokselle

$p(\log_{10}\text{-cfu} \mid \text{DATA})$



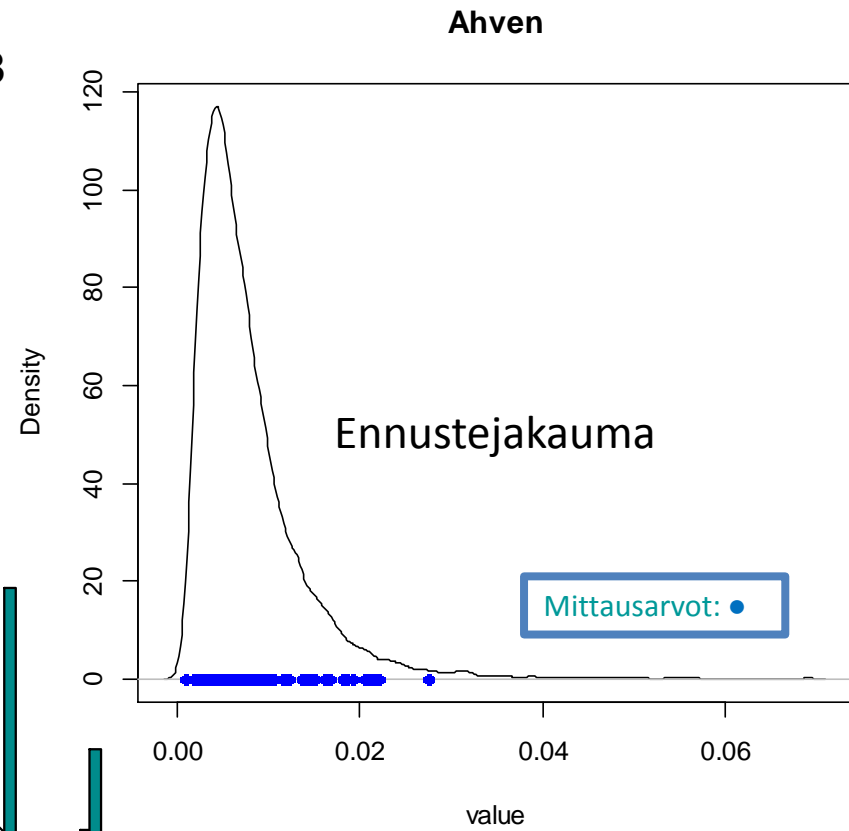
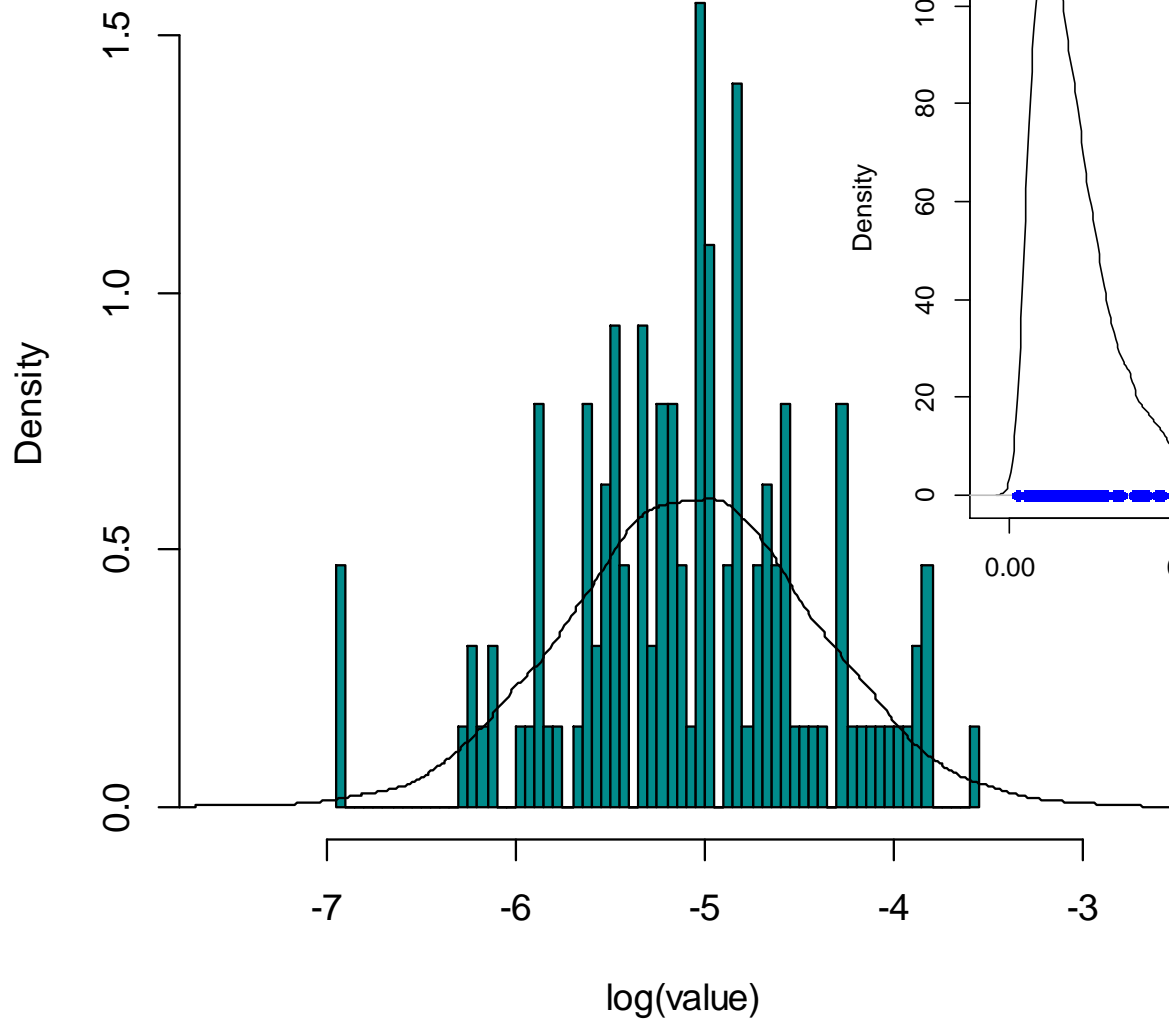
- Voidaan kuvata myös graafisena mallina:
- $P(x_i | \mu, \sigma)$ $i=1, \dots, n$ (eksaktit havainnot x_i)
- $P(x_{\text{new}} | \mu, \sigma)$ (ennustettava x_{new})



Aineiston määrä ja ominaisuudet vaihtelevat paljon tapauskohtaisesti!
→ vaikutus tilastolliseen päättelyyn?

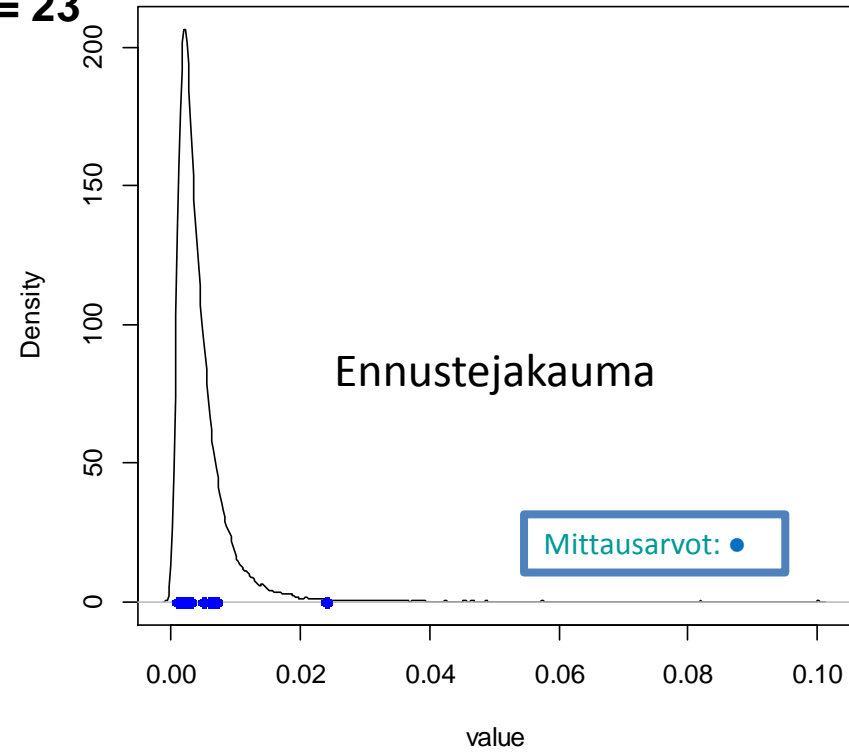
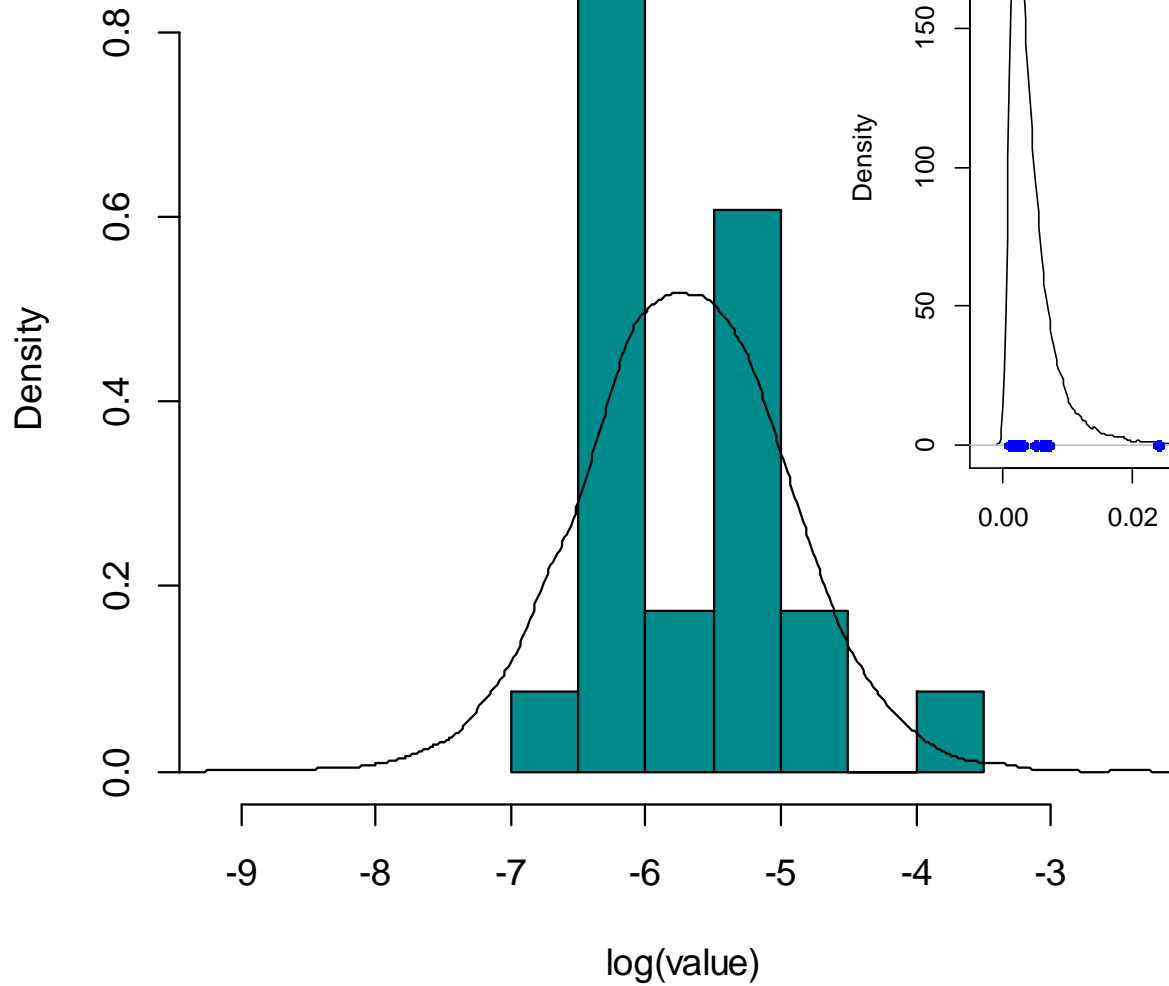
Esimerkki: kadmium

Ahven, n = 128



Esimerkki: kadmium

Muikku ja siika, n = 23



Pitoisuus on vain yksi osa pulmaa. Entä **dose-response** & **exposure** ?

- **Annos-vaste:** kirjallisuudesta löytyy joillekin patogeeneille aiemmin tutkittu dose-response. (Harvemmin uutta dataa).
 - *“Before regulatory agencies can conduct risk assessments for food-borne pathogens, dose-response models must be developed to predict infection at low doses”*. D Holcomb et al. Comparison of six dose-response models for use with food-borne pathogens. Risk Analysis, vol 19, n 6, 1999. 1091-1100.
- **Altistumisen arviointi:** kuinka monta kontaminoitunutta annosta? Kuinka suuria annoksia? Miten prosessointi ja säilytys vaikuttaa?
 - Dataa alkutuotannosta ja/tai elintarvikeketjusta.
 - Dataa kulutuksesta (ruoankäyttötutkimuksia).

Esimerkki: EHEC (Pouillot *et al.* 2010): 2D Monte Carlo.
(<https://r-forge.r-project.org/projects/riskassessment/>)

Mikä EHEC? *Escherichia coli* O157:H7

HS 28.8.2013 : <http://www.hs.fi/kotimaa/a1377578768425>

*Seitsemän sairastunutta lasta sai ehecin vakavan muodon, niin sanotun **HUS**:n eli hemolyyttis-ureemisen oireyhtymän.*

"Vakava tauti, joka johtaa pahimmillaan kuolemaan", sanoo lastenklinikan ylilääkäri, lasten infektiotautien erikoislääkäri Jussi Mertsola Turun yliopistollisesta keskussairaalaista.

"Suomessakin on kuollut lapsia tämän takia."

Mertsola kuvailee ehecin vakavaa muotoa elimistön myrkytystilaksi, kun bakteerin tuottamat myrkyt vaikuttavat verisuonten seinämiin. Hoitaminen ei ole helppoa, koska esimerkiksi antibioottien hyödytä ei ole selvää näyttöä. Pahimpaan muotoon liittyy vakava munuaisten toimintavaje, joka vaatii keinomunuais- eli dialyysihoitoa.



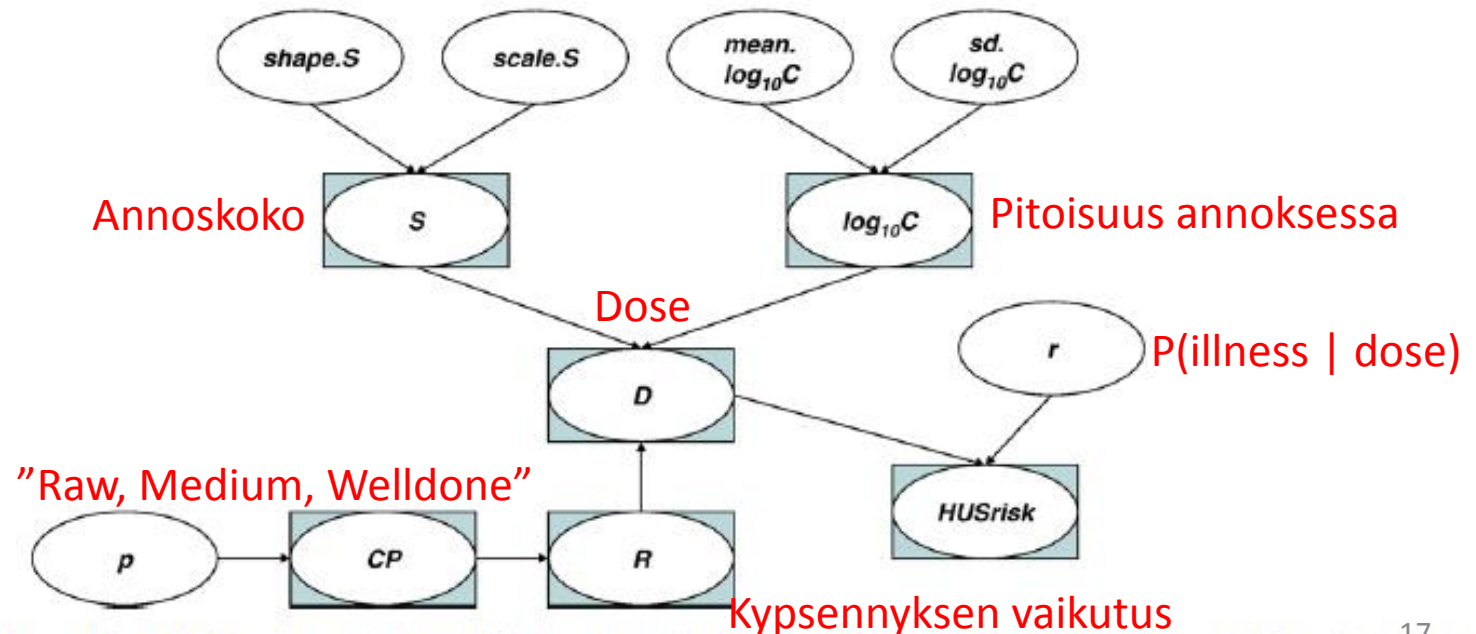
Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages

Régis Pouillot^a, Marie Laure Delignette-Muller^{b,c,*}

^a 7403 Wyndale lane, Chevy Chase, MD 20815, United States

^b Université de Lyon, F-69000, Lyon; Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

^c Université de Lyon, F-69000, Lyon; VetAgro Sup Campus Vétérinaire de Lyon, F-69280 Marcy l'Etoile, France



12.11.2015
 Fig. 1. Directed acyclic graph of the example: *Escherichia coli* O157:H7 in ground beef. See text and nomenclature for a description of the parameters and their relationships. Ovals stand for uncertain parameters and ovals-in-rectangles stand for variable and uncertain parameters.

Simulointimalli (Pouillot *et al.*)

- **S** = annoskoko (grammaa), kuluttajatutkimus n=254. Jakauma: log-normal, gamma tai Weibull.
- **log₁₀C** = pitoisuus kontaminoituneessa kypsentämättömässä jauhelihasa. Jakauma: yleensä N-jakauma log-pitoisuudelle.
 - Huom: vasemmalta sensurointi, oikealta sensurointi, välisensurointi !
 - Usein vain vähän dataa, jos pitoisuudet matalia/harvinaisia
- **CP** = kypsennystapa (Raw/Medium/Welldone): kategorinen jakauma.
- **R** = kypsennyksen aiheuttama vähenemä (log₁₀), ehdolla CP:
 - U(0,0.9) tai U(0.2,1.4) tai U(1.2,2.8)
- **D** = dose, jakauma Poisson(S*C*10^{-R})
- **r** = sairastumistodennäköisyys yhdestä bakteerista, jakauma: Pert(min 1e-4, mode 1.2e-3, max 1e-2))
- **HUSrisk** = 1-(1-r)^D
- **Uncertainty & Variability: 2D Monte Carlo.**
- *Mitä jakaumat esittävät? (ks. edellinen sivu)*

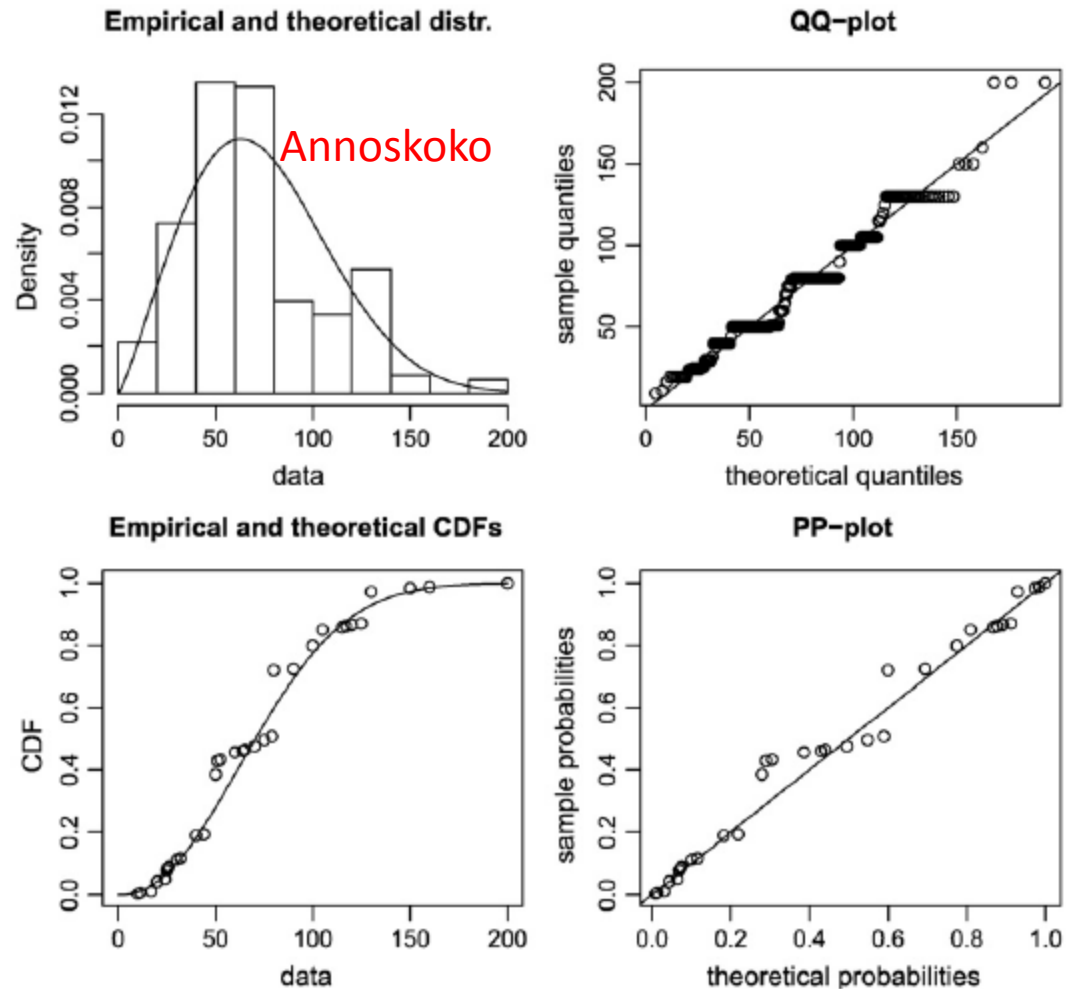


Fig. 4. Results of the fitting of a Weibull distribution to the serving size dataset. Clockwise: i) histogram of the empirical distribution (data) superimposed with the density function of the theoretical fitted distribution; ii) Q-Q plot, i.e. plot of the quantiles of the theoretical fitted distribution (x-axis) against the empirical quantiles of the data (y-axis); iii) P-P plot, i.e. for each value of the data set, the cumulative density function of the fitted distribution (x-axis) is plotted against the empirical cumulative density function (y-axis); iv) empirical density function of the data superimposed with the cumulative density function of the theoretical fitted distribution.

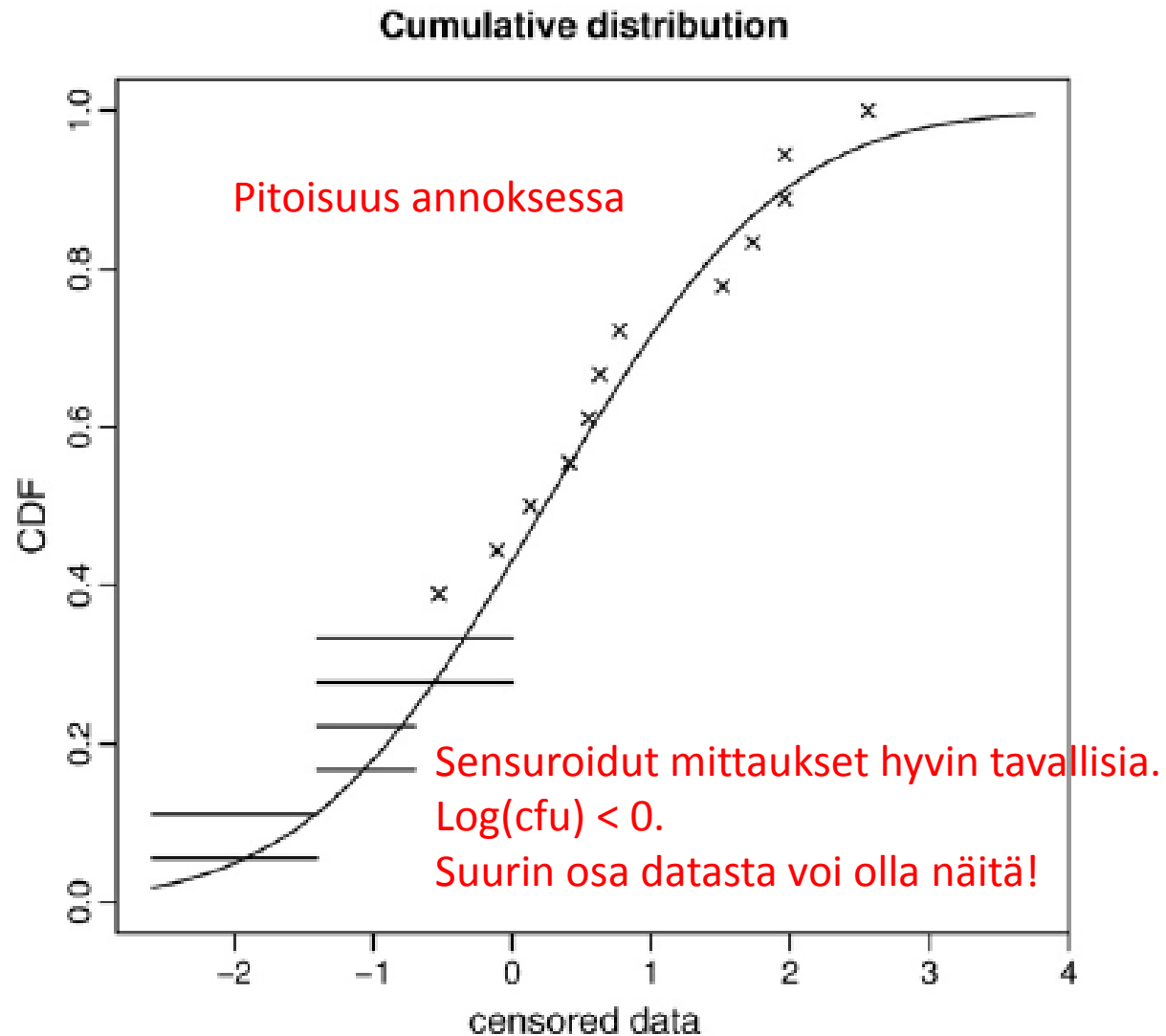


Fig. 5. Goodness-of-fit graph for the fit of a lognormal distribution censored dataset of the decimal logarithm microbial concentration from the example. Crosses stand for non censored data. Segments lie between the limits of the censoring interval.

2D Monte Carlo

R. Pouillot, M.L. Delignette-Muller / International Journal of Food Microbiology 142 (2010) 330–340

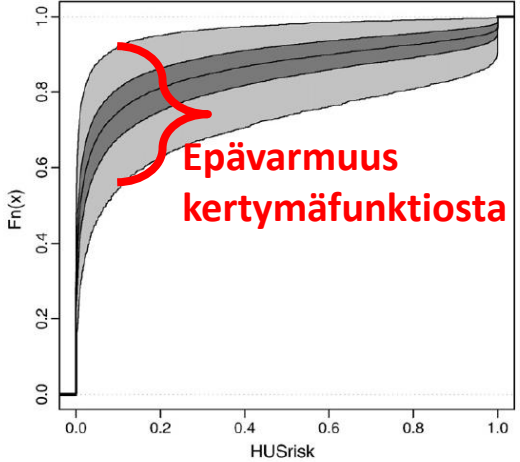
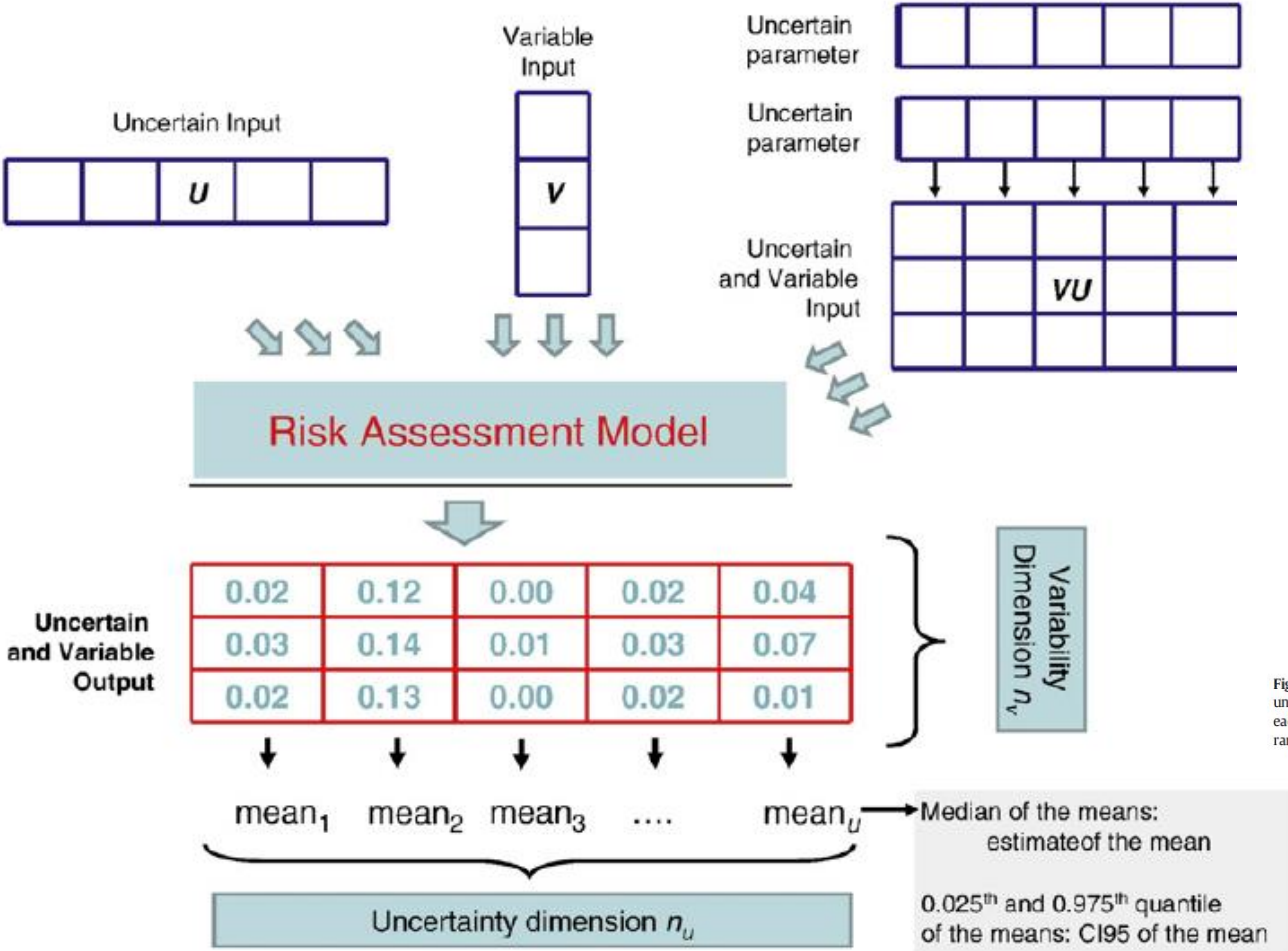
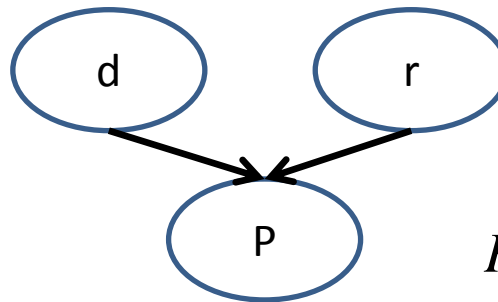


Fig. 8. Variability cumulative distribution plots of the output of the *E. coli* model, with uncertainty indicated: light gray band corresponds to the 95% uncertainty range on each quantile of variability and dark gray band corresponds to the 50% uncertainty range on each quantile of variability.



$$P(\text{inf} | r, d) = 1 - (1 - r)^d$$

Pitoisuudet ja annoskoot mallinnettu?

Hyvä!  Annosvaste –malli.

$P(\text{illness} | \text{dose})$

C Haas: Conditional dose-response relationships for microorganisms: development and application. Risk Analysis, Vol 22, n 3, 2002. 455-463.

1. Oletus: d solua annoksessa, jokainen voi sairastuttaa: $P(\text{inf})=r$

- Infektion todennäköisyys, ehdolla d: **Single-hit-model**

$$P(\text{inf} \mid r, d) = 1 - (1 - r)^d$$

- Jos infektiöalttius r vaihtelee yksilöiden välillä $r_i \sim \text{Beta}(\alpha, \beta)$ (=epävarmuus yhden yksilön kohdalla):

$$P(\text{inf} \mid \alpha, \beta, d) = \int_0^1 [1 - (1 - r)^d] \text{Beta}(r \mid \alpha, \beta) dr$$

$$= 1 - \frac{\Gamma(\alpha + \beta)\Gamma(\beta + 1)}{\Gamma(\beta)\Gamma(\alpha + \beta + d)} = E([1 - (1 - r)^d] \mid \alpha, \beta, d)$$

2. Seuraavaksi: tuntematon solumäärä (d)

- Oletus $d \sim \text{Poisson}(\lambda)$, $r = \text{sama kaikille yksilöille}$
→ exponential dose-response model

$$P(\text{inf} \mid r, \lambda) = \sum_{d=0}^{\infty} [1 - (1 - r)^d] \text{Poisson}(d \mid \lambda) = 1 - e^{-r\lambda}$$

- Tässä λ on keskimääräinen solumäärä annoksessa, r on infektiodennäköisyys yhdestä solusta.

3. Yhdistetään edelliset 1&2

- Oletus: $d \sim \text{Poisson}(\lambda)$, ja r vaihtelee yksilöiden välillä

→ beta-poisson dose-response model

$$P(\text{inf} \mid \alpha, \beta, \lambda) = \sum_{d=0}^{\infty} \int_0^1 [1 - (1-r)^d] \text{Beta}(r \mid \alpha, \beta) dr \text{Poisson}(d \mid \lambda)$$

$$= \sum_{d=0}^{\infty} \left[1 - \frac{\Gamma(\alpha + \beta)\Gamma(\beta + d)}{\Gamma(\beta)\Gamma(\alpha + \beta + d)} \right] \frac{\lambda^d}{d!} e^{-\lambda}$$

- Tämä voidaan kirjoittaa lyhyemmin, ns. Kummer's confluent hypergeometric function (KCHF)... Ja myös likiarvoisesti lyhyenä kaavana.

Kooste: single-hit DR -mallit

$P(\text{inf} \mid r, d)$ $P(\text{inf} \mid \alpha, \beta, d)$ $P(\text{inf} \mid r, \lambda)$ $P(\text{inf} \mid \alpha, \beta, \lambda)$	Vakio d	Satunnainen d
Vakio r	$1 - (1 - r)^d$ Binomial	$1 - \exp(-r\lambda)$ exponential
Satunnainen r	$1 - \frac{\Gamma(\alpha + \beta)\Gamma(\beta + 1)}{\Gamma(\beta)\Gamma(\alpha + \beta + d)}$ Beta-binomial	$1 - M(\alpha, \alpha + \beta, -\lambda)$ Beta-Poisson

Sairastuuko siitä?

$P(\text{inf} | \text{mallin parametrit, dose}) = ?$

Comparison of Six Dose-Response Models for Use with Food-Borne Pathogens

Holcomb et al.

David L. Holcomb,¹ Mary A. Smith,^{1,2} Glenn O. Ware,³ Yen-Con Hung,²
Robert E. Brackett,² and Michael P. Doyle²

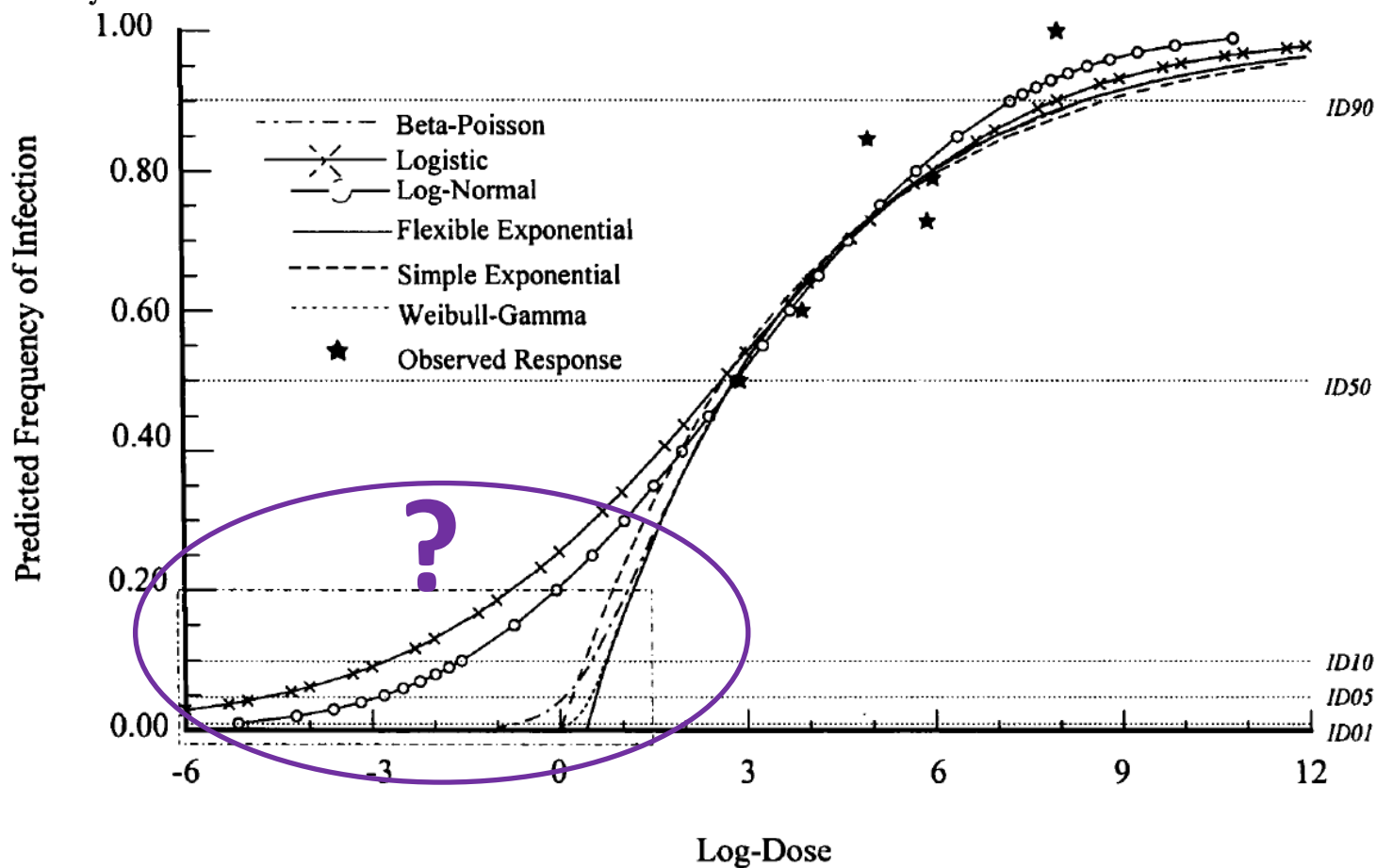



Fig. 1. Predicted frequency of infection versus log-dose for *Campylobacter jejuni*. All six of the models compared in this study had an acceptable fit to the microbial data from Black *et al.*⁽¹⁶⁾ (observed data overlaid on the plot).

- Ilman ”omaa” DR-dataa, DR-mallin parametrit α , β perustuvat yleensä kirjallisuudesta saatuihin estimaatteihin.
- Joillekin patogeeneille löytyy ”valmiiksi estimoitu” yleisesti hyväksytty konsensusmalli esim. FAO/WHO.
- Kampylobakteeri: beta-poisson malli: $\alpha=0.145$, $\beta = 7.59$.
- Vaste voi olla erilainen joillain väestöryhmillä, vanhukset, lapset, immuunipuutokset.
- Low-dose effect?

- Jos DR-dataa löytyy, voi toki mallintaa DR-mallin parametrien epävarmuuden sovittamalla mallin itse.
- Uutta DR-dataa vaikea saada. Löytyykö vapaaehtoisia?? 
- Esimerkki sikojen infektiosta:

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 2005, p. 2753–2755
 0099-2240/05/\$08.00+0 doi:10.1128/AEM.71.5.2753–2755.2005
 Copyright © 2005, American Society for Microbiology. All Rights Reserved.

Vol. 71, No. 5

Dose Determination for Acute *Salmonella* Infection in Pigs

A. T. Loynachan^{1,2} and D. L. Harris^{2,3*}

Department of Veterinary Microbiology and Preventive Medicine, College of Veterinary Medicine, Iowa State University,¹ Department of Animal Science, College of Agriculture, Iowa State University,² and Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University,³ Ames, Iowa

Received 1 June 2004/Accepted 23 November 2004

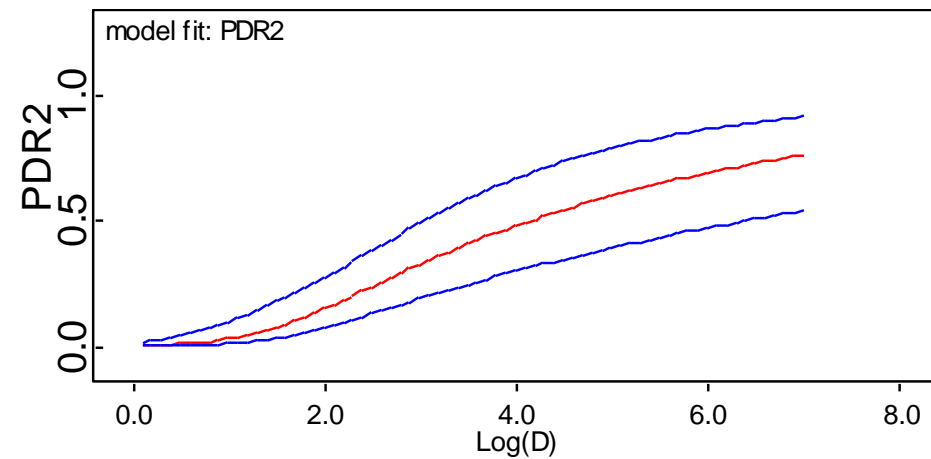
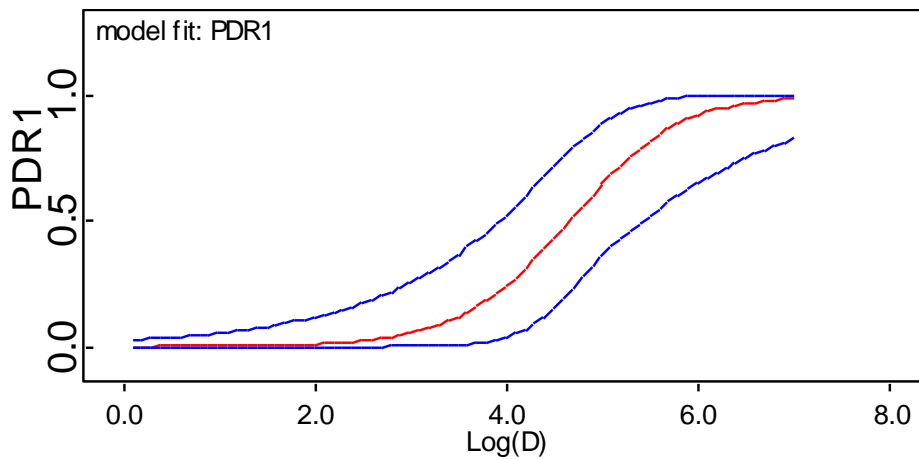
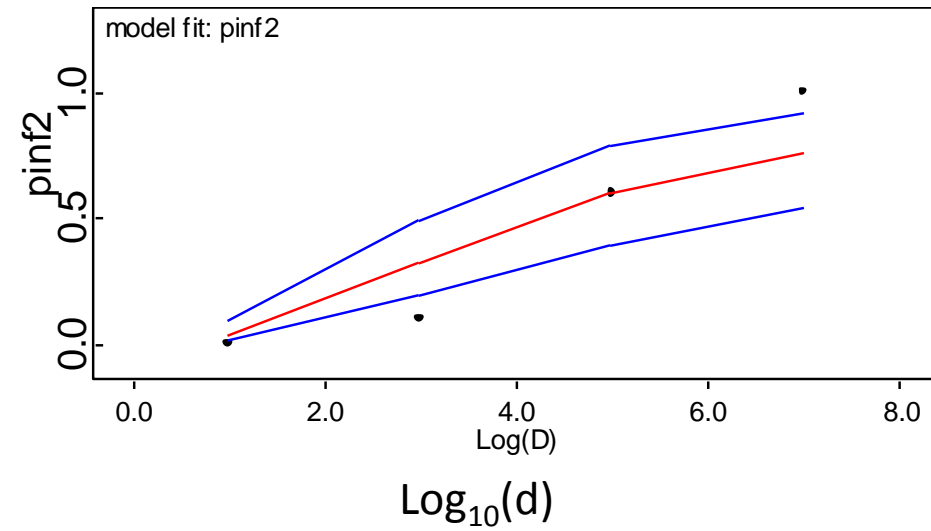
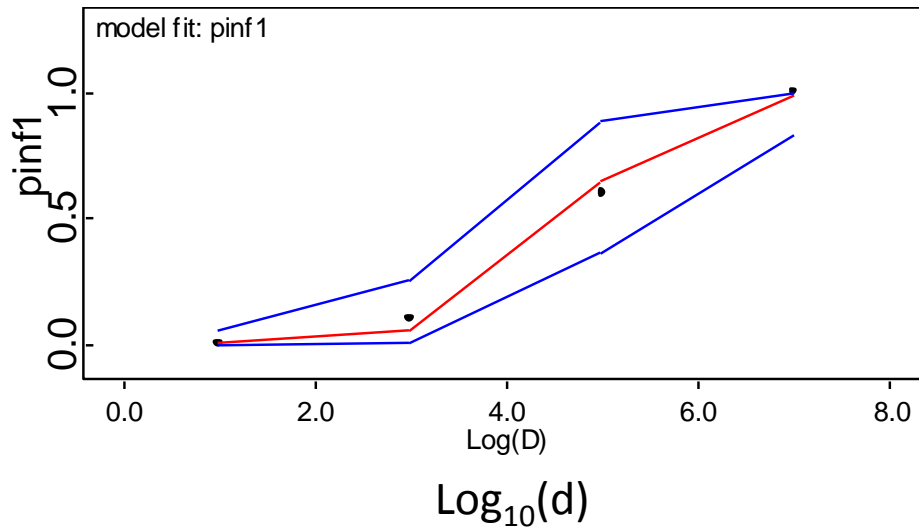
- Single-hit –malli vs. logit-malli? [$\log(p/(1-p))=a+b*\log(d)$]

TABLE 1. Incidence of *Salmonella* infection following intranasal challenges of pigs with various levels of *Salmonella* serovar Typhimurium

Sample type	No. of positive tissue samples						Negative controls (n = 6)
	Trial 1 (n = 5)			Trial 2 (n = 5)			
	10 ^{1a}	10 ³	10 ⁵	10 ³	10 ⁵	10 ⁷	
Alimentary							
Tonsil	0	0	3	1	3	5	0
Ileum	0	0	2	0	3	5	0
Cecum contents	0	0	2	0	2	5	0
Colon contents	0	0	0	0	0	5	0
% Positive	0	0	35	5	40	100	0
Nonalimentary							
Mandibular lymph node	0	0	1	0	1	1	0
Thymus	0	0	0	0	0	1	0
Lung	0	0	0	0	0	0	0
Liver	0	0	0	0	0	0	0
Spleen	0	0	0	0	0	0	0
Ileocecal lymph node	0	0	0	0	0	0	0
Muscle	0	0	0	0	0	0	0
Blood	0	0	0	0	0	0	0
% Positive	0	0	3	0	3	5	0
% Positive (all tissues)	0	0	13	2	15	37	0

^a Intranasal challenge dose of salmonellae.

logit vs. single-hit malli edellisestä taulukosta estimoituina



- Tunnetut jakaumafunktiot riittävät vain joihinkin tapauksiin, koska integrointia ei aina voi suorittaa analyyttisesti.
- Monte Carlo -laskennassa (tai MCMC) voidaan integroinnissa hyödyntää **ehdollista** muotoa todennäköisyydelle:

$$P(\text{inf} \mid \alpha, \beta, \lambda) = \sum_{d=0}^{\infty} P(\text{inf} \mid \alpha, \beta, d)P(d \mid \lambda) \approx \frac{1}{n} \sum_{k=1}^n P(\text{inf} \mid \alpha, \beta, d^{(k)})$$

 $d \sim$ Monte Carlo -otos jakaumasta $P(d \mid \lambda)$

Schmidt et al: Harnessing the theoretical foundations of the exponential and beta-poisson dose-response models to quantify parameter uncertainty using Markov-chain Monte Carlo. Risk Analysis, Vol 33, No 9, 2013. 1677-1693.

$$P(\text{inf} | \alpha, \beta, \lambda) = \sum_{d=0}^{\infty} P(\text{inf} | \alpha, \beta, d)P(d | \lambda) \approx \frac{1}{n} \sum_{k=1}^n P(\text{inf} | \alpha, \beta, d^{(k)})$$



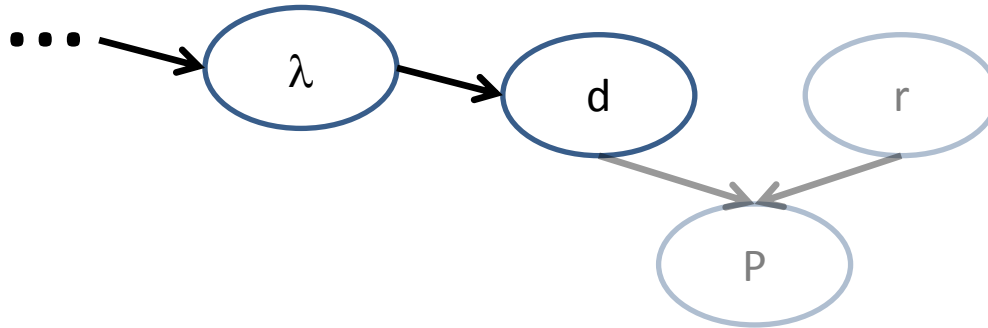
$d \sim$ Monte Carlo –otos jakaumasta $P(d | \lambda)$

- Tässä $d^{(k)}$ on otos jakaumasta $P(d | \lambda) = \text{Poisson}(\lambda)$, joka puolestaan voi perustua [prediktiiviseen mikrobiologiseen malliin](#), johon edelleen liittyy oletuksia edeltävistä tapahtumaketjuista ja niiden parametreista:
 $P(d | \lambda)P(\lambda | \theta)P(\theta) =$ **malli annokseen kulkeutuvien solujen määrästä.**
 → **ketjumalli** ”food pathway”, ”farm-to-fork”



Altistumisen pitkä reitti: prosessiketjujen mallinnus

- **Prevalenssi ja pitoisuus:** eri vaiheissa elintarvikeketjua, tuotannosta kuluttajalle, "farm-to-fork".
 - $P(\text{inf} \mid \text{contamination, dose})P(\text{dose} \mid \text{contamination})P(\text{contamination})$
- **Vaihtelu ja epävarmuus:** useita vaihtelun lähteitä (annoskoko, bakteerimäärä, tuotantoerien välinen ja sisäinen vaihtelu, bakteerien kasvu). Useita epävarmuuden lähteitä (mallien parametriepävarmuus, oletusten epävarmuus).



Models of food pathway

$$P(d | \lambda)P(\lambda | \theta)P(\theta)....$$

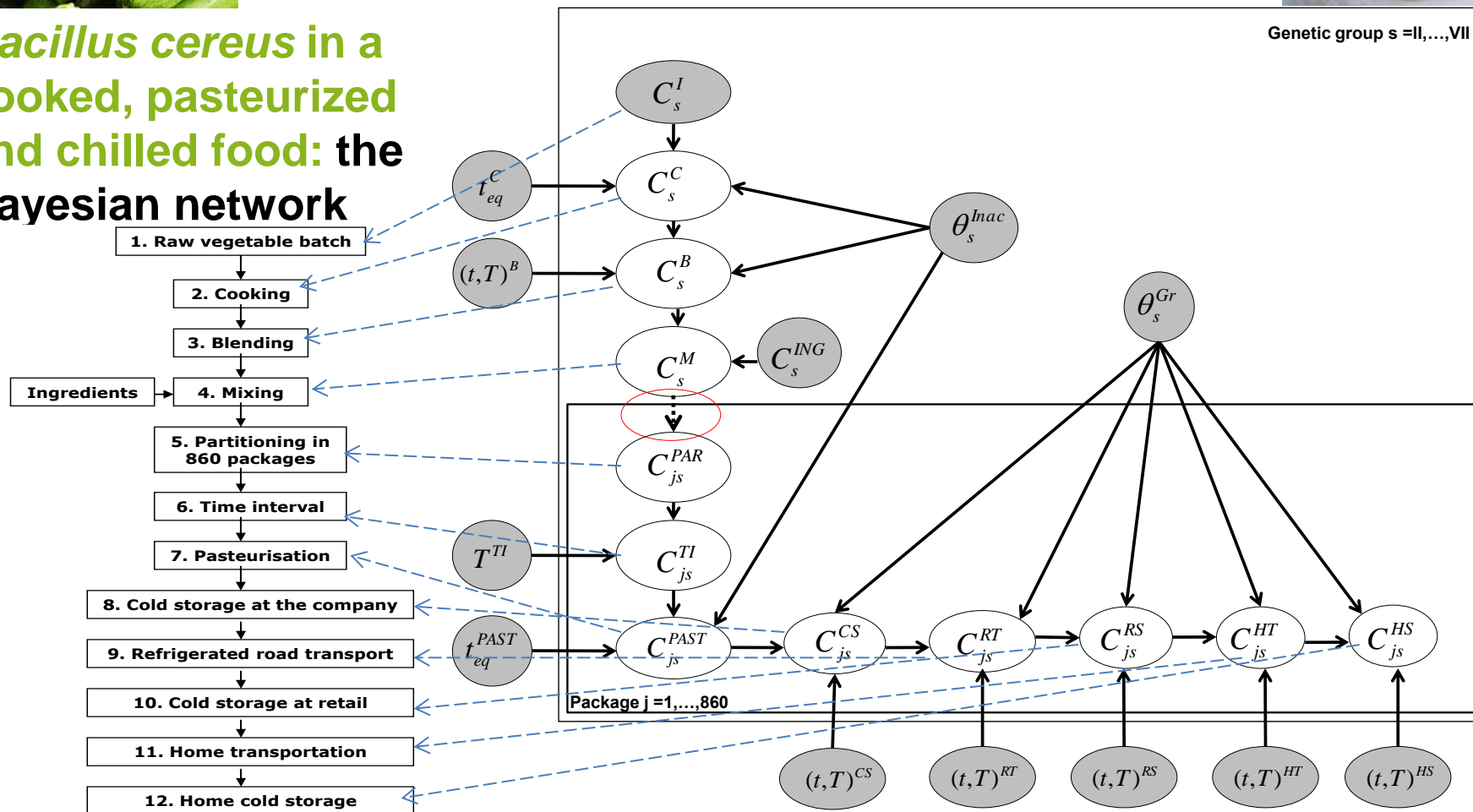
Esimerkki: C Rigaux et al: Inferring an Augmented Bayesian Network to Confront a Complex Quantitative Microbial Risk Assessment Model with Durability Studies: Application to *Bacillus Cereus* on a Courgette Purée Production Chain. Risk Analysis, Vol 33, n 5, 2013, 877-892.

Seuraava kalvo lainaus Isabelle Albertin esityksestä.



How to model the food pathway? Model in the case study

Bacillus cereus in a cooked, pasteurized and chilled food: the Bayesian network



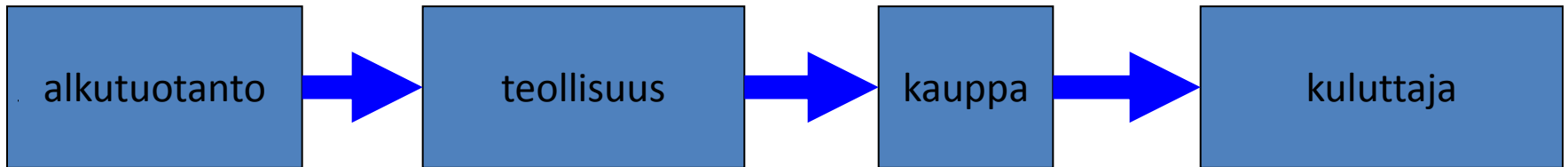
Original slide from: Isabelle Albert.
Met@risk unit, Paris.

DIET
AGRICULTURE
ENVIRONMENT



Altistusketju vaiheittain

- Usein modulaarinen rakenne:
 - Pitoisuuksien muutokset eri olosuhteissa (bakteerien kasvu & tuhoutuminen → **growth/inactivation models**).
 - Esiintyvyyksien muutokset (**mixing & partitioning**).



- Interventioiden vaikutus?
- Riskinarviointikysymyksen fokus?
- *Mistä vaiheista on dataa?*

Usein huomioituja asioita:

- Partitioning: tuotantoerän osittaminen.

- Bakteerien pitoisuus erässä 1cfu/g
- Erän koko 1000 kg
- Bakteerien määrä erässä 10^6 cfu
- Erä jakaantuu kymmeneen yhtä suureen osaan, bakteerien määrä osissa \sim Multinomial($p, 10^6$)



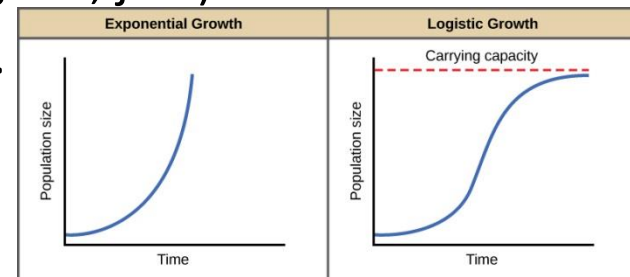
- Mixing: tuotantoerien yhdistäminen.

- Bakteerien pitoisuudet osissa X_i cfu/g
- Osien suhteelliset koot p_i
- Pitoisuus yhdisteessä $\sum p_i X_i$



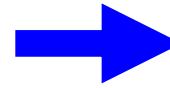
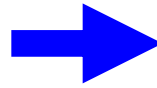
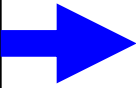
- Growth/inactivation:

- Bakteerien kasvumallit (Gompertz, logistic, jne.)
- Riippuu lämpötilasta, ajasta, PH, jne....



Altistusketju A-Ö

alkutuotanto



kuluttaja

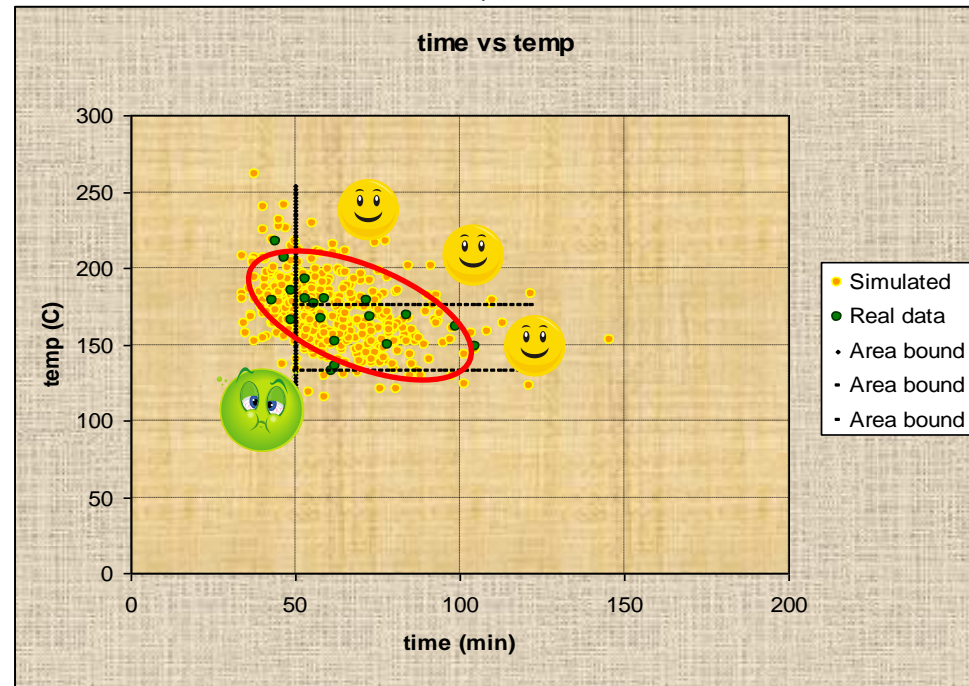
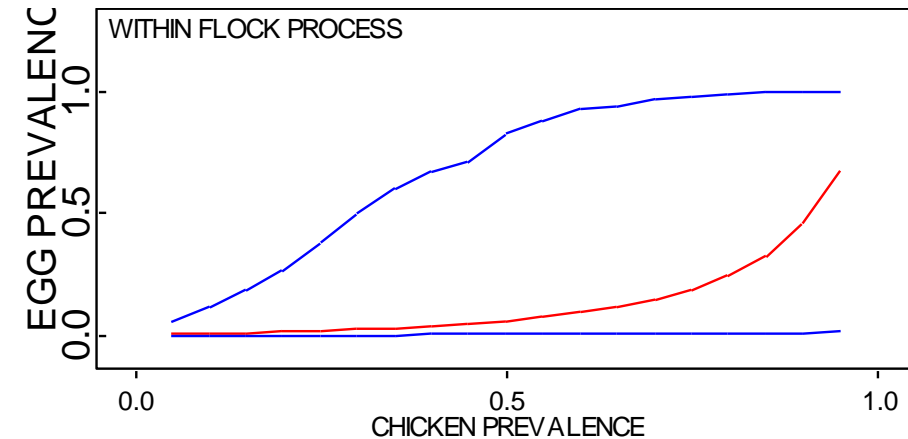


Esim: kananmunien kontaminoituminen parven kanoista

Esim: ruoan kypsennysaika ja lämpötila

$$\text{Logit}(p_2) = \alpha + \beta \text{logit}(p_1) + \varepsilon_i$$

$$\text{Log}(t,T) \sim \text{MN}(\mu, \Sigma)$$



- **Kysymys:** kuinka suurta kontaminaatio saa enintään olla ketjun alussa, jotta kuluttajariski jää alle tavoitearvon?

- Ottaen huomioon elintarvikeketjun muuttujien vaihtelu ja arvion epävarmuudet.

- esim. kampylobakteeri broilereissa:

- Kontaminoituneiden tuotantarvioiden esiintyvyys.

- Kontaminoituneiden lihaerien esiintyvyys.

- Esiintyvyys erän sisällä.

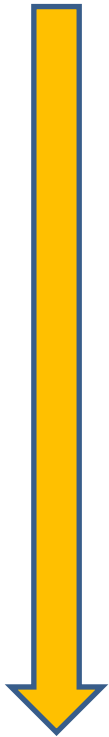
- Pitoisuuksien vaihtelu erän sisällä.

- Pitoisuuksien vaihtelu erien välillä.

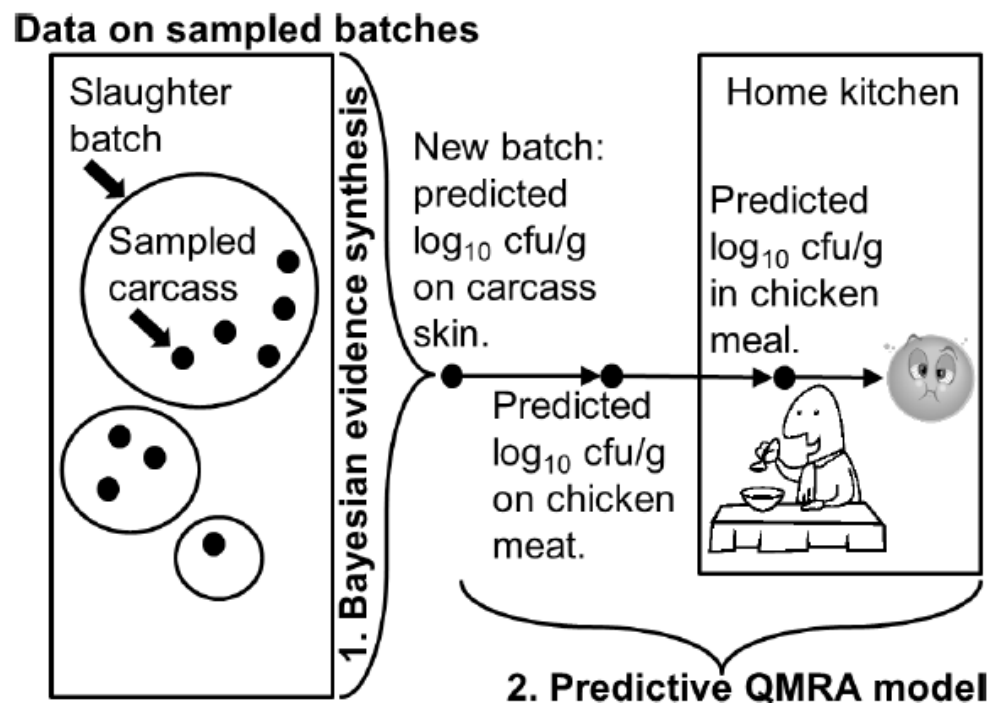
- Pitoisuuden muutos prosessoinnissa (carcass → fresh meat).

- Ristikontaminaation mahdollisuus raakalihasta valmiiseen annokseen.

- Annosten koko ja kappalemäärä, "final dose".



- → Tilastollinen ongelma:
- jokaisessa EU-maassa on seuranta-aineistoja broilereista.
- Lasketaan niihin perustuva arvio riskistä joka maalle.
- Arvion epävarmuuteen vaikuttaa kunkin datan määrä ja laatu.
- Tilastollisen mallin tulisi pystyä käsittelemään erilaisia dataja.
- Arvioidaan kuinka paljon riski suhteellisesti muuttuu jos asetetaan tietty mikrobikriteeri tiettyyn vaiheeseen ketjua.



Estimointia kahdesta hyvin erilaisesta aineistosta (Ruotsi) = "maakohtainen data"

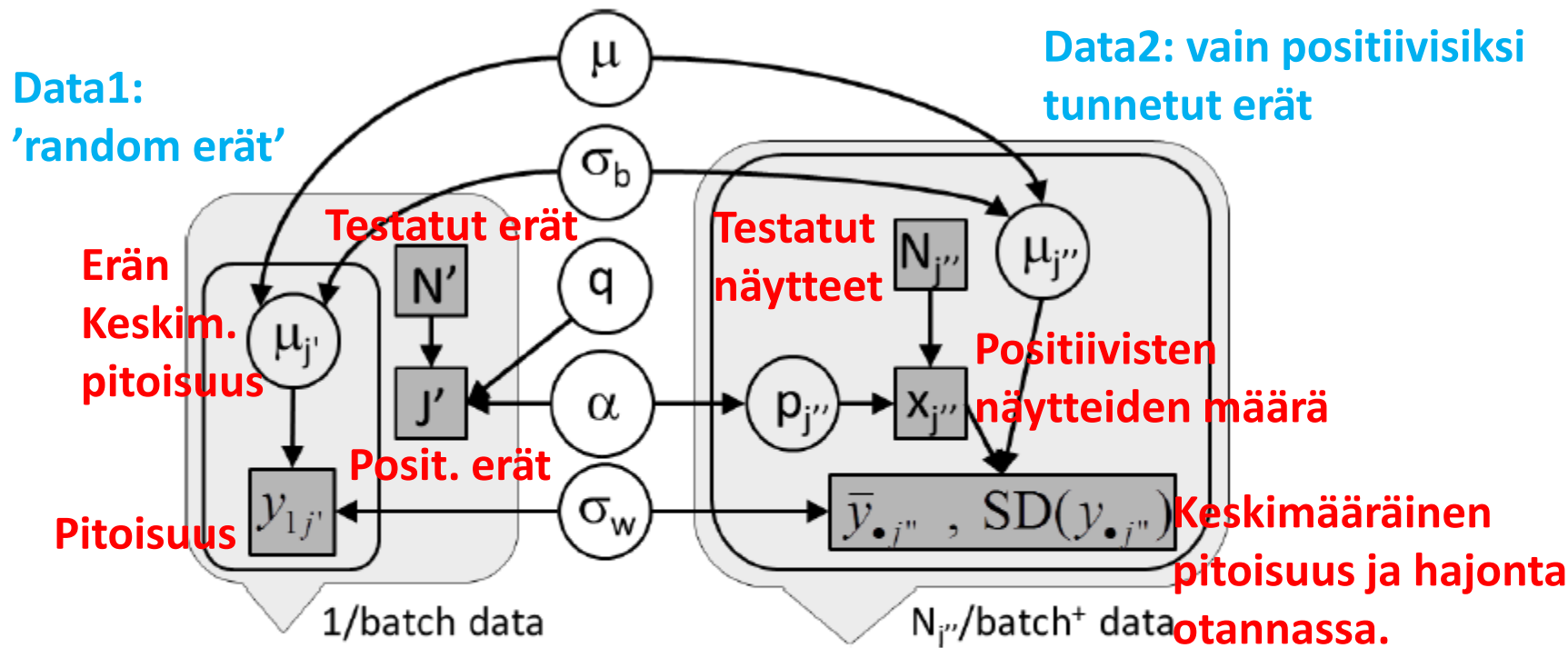
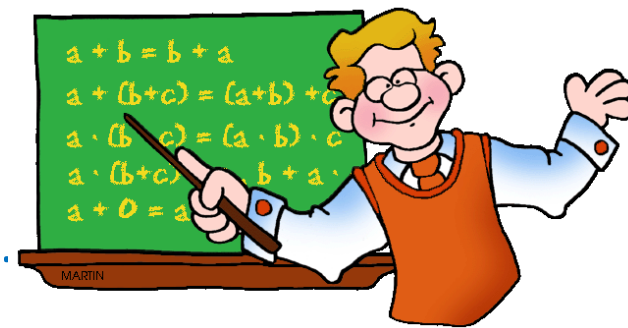


FIG 2. Directed acyclic graph of the final evidence synthesis model combining two qualitatively different data sets with common parameters in the middle. Incoming arrows denote conditional distributions as explained in the text.

Miten arvion tilastollinen epävarmuus lasketaan?

Lyhyt katsaus taustalla pyörivään matematiikkaan....



Parametrien posteriorijakauma, ehdolla Data1 tai Data2,
tai Data1&Data2 (= evidenssin synteesi).

Bayesin kaavan mukaan, laskenta OpenBUGSilla:

$$(1) \quad \pi(\mu, \tau_b, \tau_w, q, \mu_1, \dots, \mu_{J'} \mid J', N', y_{11}, \dots, y_{1J'}) \propto$$

$$\text{Binomial}(J' \mid N', q) \prod_{j'=1}^{J'} \text{N}(y_{1j'} \mid \mu_{j'}, \tau_w) \text{N}(\mu_{j'} \mid \mu, \tau_b) \pi(\mu) \pi(\tau_b) \pi(\tau_w) \pi(q),$$

$$(2) \quad \pi(\mu, \mu_1, \dots, \mu_{J''}, \tau_b, \tau_w, p_1, \dots, p_{J''}, \alpha \mid \{x_{j''}, N_{j''}, \bar{y}_{\cdot j''}, \text{SD}(y_{\cdot j''})\}_{j''=1, \dots, J''}) \propto$$

$$\prod_{j''=1}^{J''} \text{Binomial}(x_{j''} \mid N_{j''}, p_{j''}) \pi(\bar{y}_{\cdot j''}, \text{SD}(y_{\cdot j''}) \mid \mu_{j''}, \tau_w, x_{j''}) \text{N}(\mu_{j''} \mid \mu, \tau_b) \times$$

$$\text{Beta}(p_{j''} \mid \alpha, 2) \pi(\mu) \pi(\tau_b) \pi(\tau_w) \pi(\alpha).$$

Arvio esim. varianssiparametreista muuttuu datan mukaisesti:
Tilastotiede voi tarjota tähän ratkaisuja!

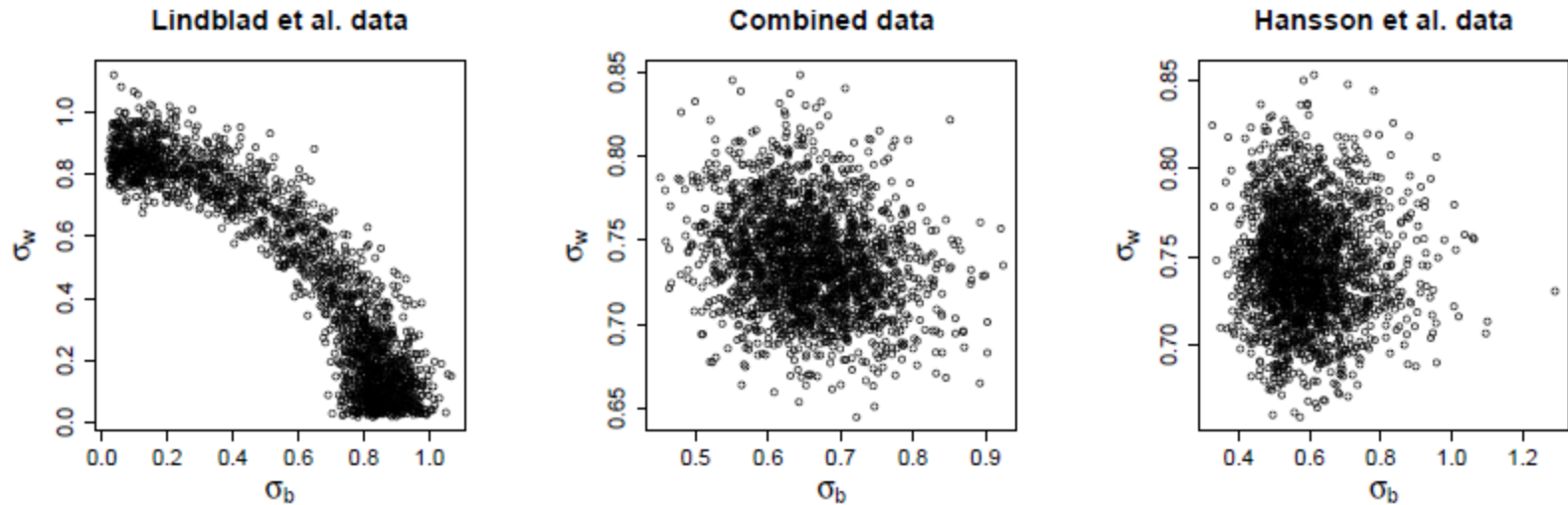


FIG 3. Marginal posterior distributions of (σ_b, σ_w) based on each data set alone (1/batch left, N_j''/batch^+ right) and the two data sets combined (middle).

TAVOITE: Riski: $P(\text{illness} | q, \mu, \sigma_w, \sigma_b, \alpha)$

= Sairastumistodennäköisyys per annos, ehdolla "kaikki ydinparametrit" jotka kuvaavat kyseisen maan tuotantoa.

"Ydinparametrit" (joiden epävarmuus riippuu maakohtaisesta datasta):

q = todennäköisyys että annoksen broiler on peräisin kontaminoituneesta erästä, ja erän status on $I_j \sim \text{Bernoulli}(q)$.

p = todennäköisyys että broiler on kontaminoitunut, jos se on em. erästä, erän sisäinen prevalenssi on $p_j \sim \text{Beta}(\alpha, 2)$.

μ_j = kontaminoituneen erän keskim. log-pitoisuus $\mu_j \sim N(\mu, \sigma_b^2)$.





y_{ij} = kontaminoituneen broilerin pitoisuus $y_{ij} \sim N(\mu_j, \sigma_w^2)$.

σ_w = erän sisäinen vaihtelu log-pitoisuudelle.

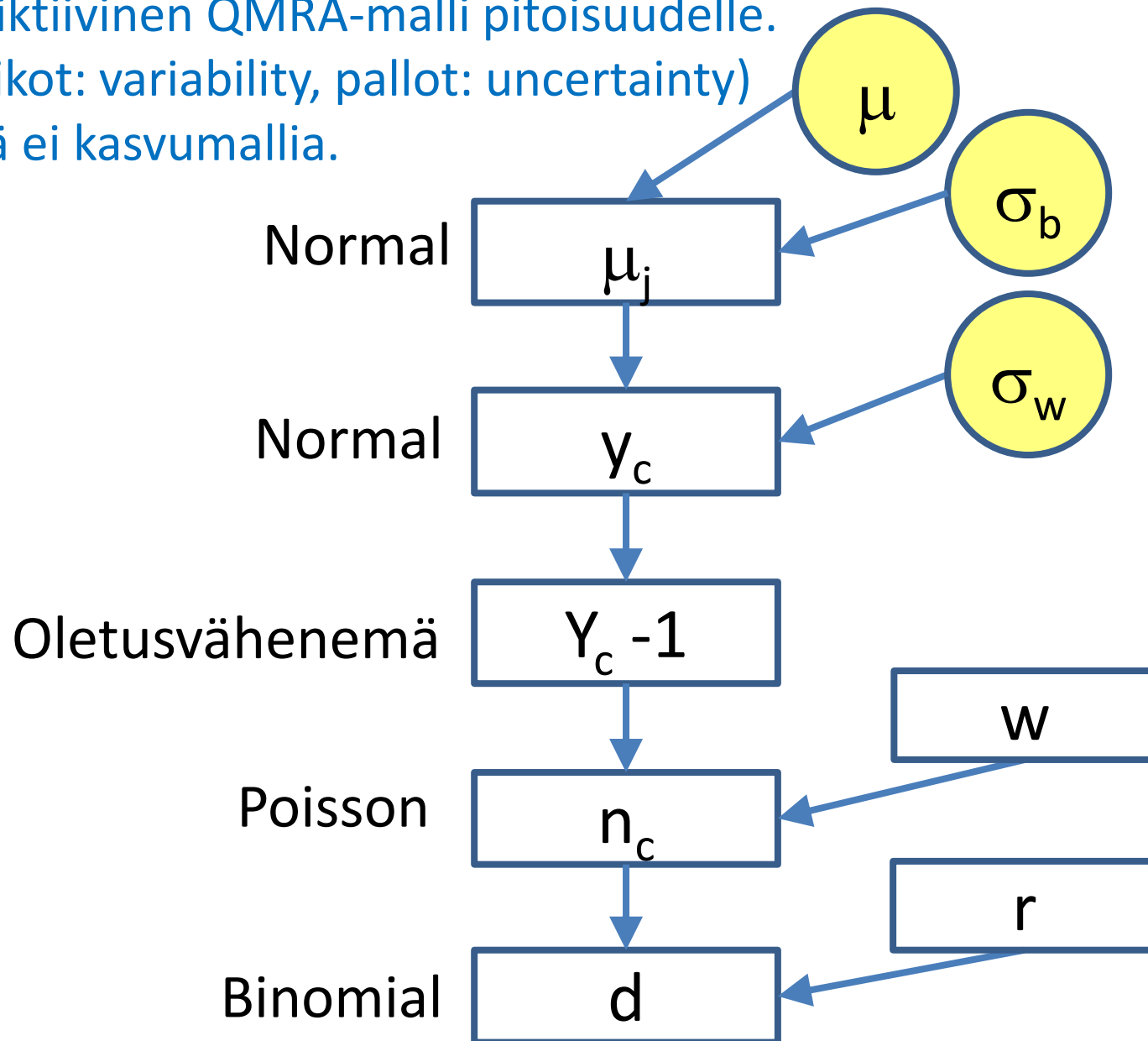
σ_b = erien välinen vaihtelu log-pitoisuudelle.

Miten evaluoidaan $P(\text{illness} | q, \mu, \sigma_w, \sigma_b, \alpha)$?

Mallin Monte-Carlo simulointia vaiheittain:

-  **Ydinparametreilla** $(q, \mu, \sigma_w, \sigma_b, \alpha)$ epävarmuusjakauma (ehdolla datat).
-  **Eräkohtaiset parametrit:** $\theta_j = (l_j, p_j, \mu_j)$. Näillä on ehdolliset jakaumat, ehdolla ydinparametrit.
-  **Annoskohtaiset parametrit** $\theta_s =$ annoskoko (w), solumäärä raakalihassa (n_c), ristikontaminaatio annokseen (r), lopullinen solumäärä annoksessa (d). Ehdollinen eräkohtaisille parametreille.
-  **”Risk per serving”**

Prediktiivinen QMRA-malli pitoisuudelle.
(laatikot: variability, pallot: uncertainty)
Tässä ei kasvumallia.



Tärkeä idea:

Erästä saatu lisätieto päivittää arviota k.o. erään liittyvästä riskistä.

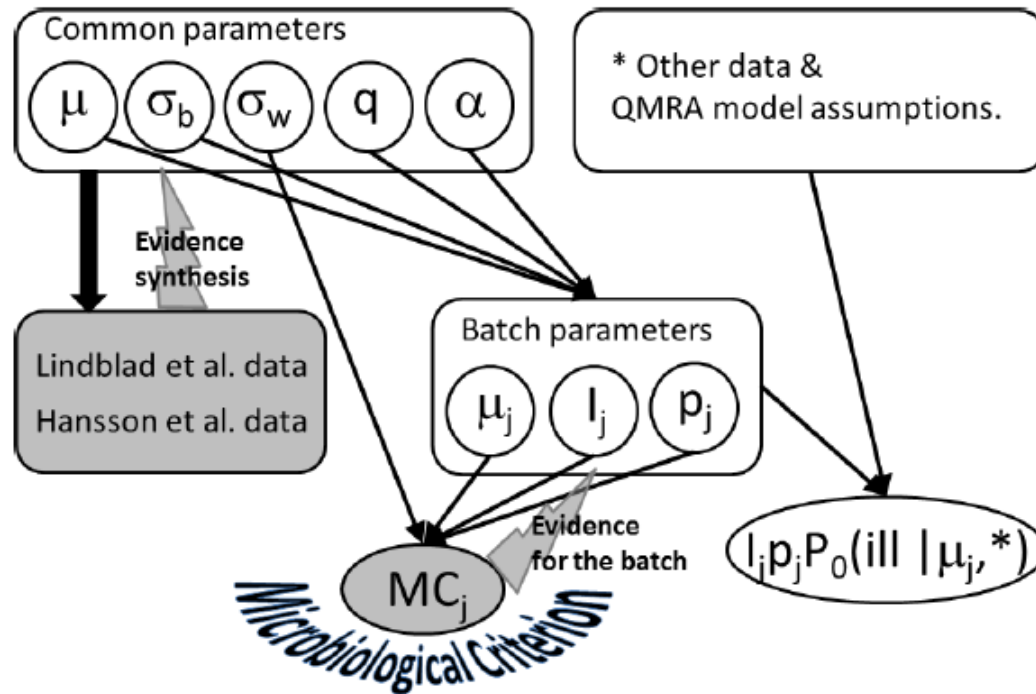


FIG 4. Directed acyclic graph of the Bayesian model combining the two carcass sample data sets for common parameters and predicting the batch parameters conditionally, now also based on the status of the batch ($MC_j = \text{accepted/rejected}$). The batch-specific illness probability (via QMRA model) then depends both on the carcass sample data and on the batch-specific status, for a generic new batch j to be predicted.

Mikrobikriteerit

- **Vaikutus riskiin, jos vain hyväksytyt erät kulutukseen?**

- *Ennuste erän testitulokselle ("n=5,c=1,m=1000")*

➤ *Millä todennäköisyydellä raja $c>1$, $m>1000$ ylittyy?*



- *Kun testitulos on selvillä*

➤ *Kysymys*: *mitä se kertoo k.o. erästä? Mitä riskistä?*

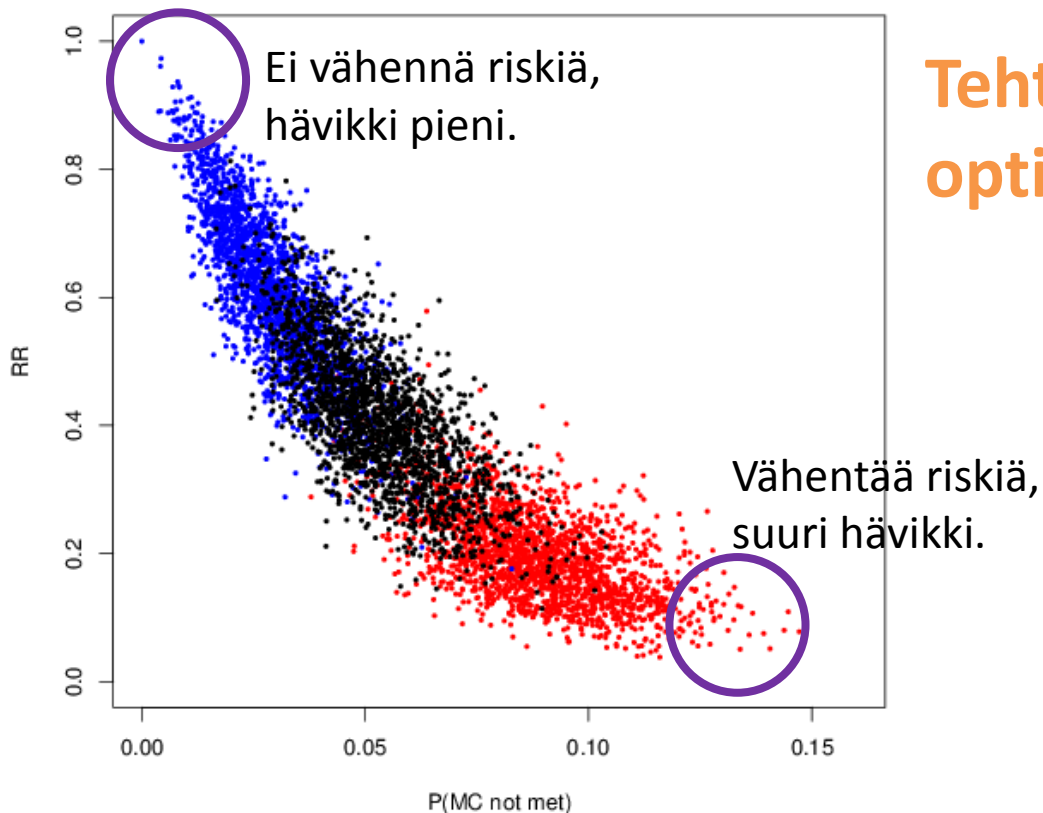
➤ *Tähän tarvitaan tilastollista päättelyä !*

➤ *Päivitetty arvio Bayes-mallin avulla → epävarmuusjakauma.*

Mikrobikriteerien arviointi ja epävarmuus

”consumer’s loss vs. producer’s loss”

RR = risk ratio



Tehtävä:
optimoi n/c/m !

P = tuotantoerän hylkäämistodennäköisyys.

Riskiperusteinen valinta: n/c/m ?

TABLE 2

Posterior means of $RR(q, \mu, \sigma_w, \sigma_b, \alpha)$, with various MC (n/c/m). Default in bold. $RR(q, \mu, \sigma_w, \sigma_b, \alpha)$ is risk conditional on acceptance of the batch (MC is met) divided by the risk conditional on “MC not applied”. Corresponding means of rejection percentages are shown in subscripts. The same results were obtained for RPR with a much faster computation.

	m=1000		m=100	
	n=5	n=10	n=5	n=10
c=0	0.20 _{9%}	0.10 _{11%}	0.01 _{14%}	0.00 _{14%}
c=1	0.42 _{5%}	0.22 _{8%}	0.06 _{12%}	0.01 _{14%}
c=2	0.61 _{3%}	0.34 _{6%}	0.14 _{10%}	0.03 _{13%}
c=3	0.78 _{1%}	0.45 _{5%}	0.30 _{7%}	0.05 _{12%}
c=4	0.92 _{0%}	0.56 _{3%}	0.58 _{4%}	0.09 _{11%}

Yhteenvetoa

- Annosvaste –malleja tarvitaan, jos halutaan arvioida väestöriskejä, eikä vain altistumista.
- Altistumisen mallit tarvitaan koska annokset ja pitoisuudet vaihtelevat.
- DATA: esiintyvyyys- ja pitoisuustietoja, ketjun jostakin vaiheesta + kulutustietoja.
- Prediktiiviset mallit: lopullisen pitoisuuden ennustaminen.
 - (vrt. kemiallinen pitoisuus on usein vakio, mikrobinen ei).
- Absoluuttinen riski vs. suhteellinen riski ?
- Interventioiden vaikutuksen arviointi, kustannus-hyöty?
- Tilastotieteen tehtävä: kertoa arvion epävarmuus !

Holcomb *et al.* Comparison of six dose-response models for use with food-borne pathogens. Risk Analysis, Vol 19, No 6, 1999. 1091-1100.

Pouillot *et al.* Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. Int. J of Food Microbiol. 142, 2010. 330-340.

Haas. Conditional dose-response relationships for microorganisms: development and application. Risk Analysis, Vol 22, No 3, 2002. 455-463.

Schmidt *et al.* Harnessing the theoretical foundations of the exponential and beta-poisson dose-response models to quantify parameter uncertainty using Markov-chain Monte Carlo. Risk Analysis, Vol 33, No 9, 2013. 1677-1693.

Loynachan & Harris. Dose determination for acute salmonella infection in pigs. Appl and environmental microbiol. Vol 71, No 5, 2005. 2753-2755.

Rigaux *et al.* Inferring an Augmented Bayesian Network to Confront a Complex Quantitative Microbial Risk Assessment Model with Durability Studies: Application to *Bacillus Cereus* on a Courgette Purée Production Chain. Risk Analysis, Vol 33, No 5, 2013, 877-892.

Ranta *et al.* A Bayesian approach to the evaluation of risk-based microbiological criteria for *Campylobacter* in broiler meat. ***Annals of Applied Statistics. Vol 9, No 3, 2015. 1415-1432.***

→ supplementary material sisältää yksityiskohdat ja BUGS-koodin!