Possible Directions for Statistics at Massey University in the Context of My Own Research and the New Zealand Environment

Prof. Dr. Dankmar Böhning

Division of International Health

Institute for Social Medicine, Epidemiology, und Health Economics

Charité Medical School Berlin


Botanischer
Garten
Institute
Mensa
der
Freien Universität
Universitäts-
klinikum
Benjamin-
Franklin
Außenstelle
BAM
FU
EFB
Rudi-
Dutschke-Weg
Arnim-
Str.
Lieben-
stein-
str.
Limonen-
Malvenstr.
Fabeckstr.
Schwendener
Flied-
ner-
weg
Reichen-
Wichernstr.
Pat-
O.-Hahn-
Alt-
enstein-
Schweidter Allee
Tietzen-
Astern-
platz
Tulpen-
Eichen-
Schützen-
Hor-
Enzi-
M.-
Luther-K.
Kita
Bundes-
Kinder
hm
Willdenow-
1 Braillestr.
2 Wulffstr.
3 Waetzoldtst.
Rothe
Garten
148
101
X11·111
Botanischer
Garten
Bundes-


SHETLAND ISLANDS
ORKNEY ISLANDS
HEBRIDES
Rockall (U.K.)
North Atlantic Ocean
North Sea
Bergen
Oslo
Stockholm
Gävle
Tampere
Helsinki
Turku
St. Petersburg
Gulf of Finland
ALAND ISLANDS
Tallinn
ESTONIA
RUSSIA
Stavanger
Göteborg
Skagerrak
Kattegat
Gotland
Baltic Sea
Öland
Riga
LATVIA
Aberdeen
Glasgow
Edinburgh
UNITED
Belfast
Isle of Man (U.K.)
Irish Sea
Dublin
IRELAND
Leeds
Manchester
Liverpool
KINGDOM
Cardiff
Birmingham
Amsterdam
NETH.
Rotterdam
London
Celtic Sea
English Channel
Guernsey (U.K.)
Jersey (U.K.)
DENMARK
Copenhagen
Malmö
Bornholm
Kaliningrad
RUSSIA
LITHUANIA
Vilnius
Vitsyebsk
Mahilyow
Minsk
BELARUS
Hrodna
Homyel'
Gdańsk
Hamburg
Bremen
Berlin
Poznań
Warsaw
POLAND
Brest
Łódź
Wroclaw
Rivne
Oder
Elbe
Vistula
Essen
Cologne
Leipzig
Brussels
Lille
BEL.
Bonn
GERMANY
Frankfurt am Main
Luxembourg
LUX.
Prague
CZECH REPUBLIC
Brno
Kraków
L'viv
UKRAINE
Chernivtsi
SLOVAKIA
CARPATHIAN MTS.
Bratislava
Chisinau
Iaşi
MOLDOVA
Cluj-Napoca
Budapest
Vienna
AUSTRIA
HUNGARY
ROMANIA
Paris
Seine
Loire
Nantes
Strasbourg
Rhine
Stuttgart
Munich
Danube
LIECH.
Zürich
Bern
Vaduz
SWITZ.
Geneva
FRANCE
Bay of Biscay
Ljubljana
SLOVENIA
Zagreb
Bucharest
Lyon
MASSIF CENTRAL
Milan
Venice
Turin
Po
Bordeaux
BOSNIA AND HERZEGOVINA
Belgrade
Varna
Danube
CROATIA
Genoa
SAN MARINO
Sarajevo
SERBIA AND MONTENEGRO
BULGARIA
Bilbao
Toulouse
MONACO
Florence
Ligurian Sea
Andorra la Vella
Marseille
Sofia
Porto
PYRENEES
Zaragoza
ANDORRA
ITALY
Adriatic Sea
Skopje
F.Y.R.O.M.
Corsica
Rome
Madrid
Barcelona
VATICAN CITY
Tirana
ALB.
Thessaloniki

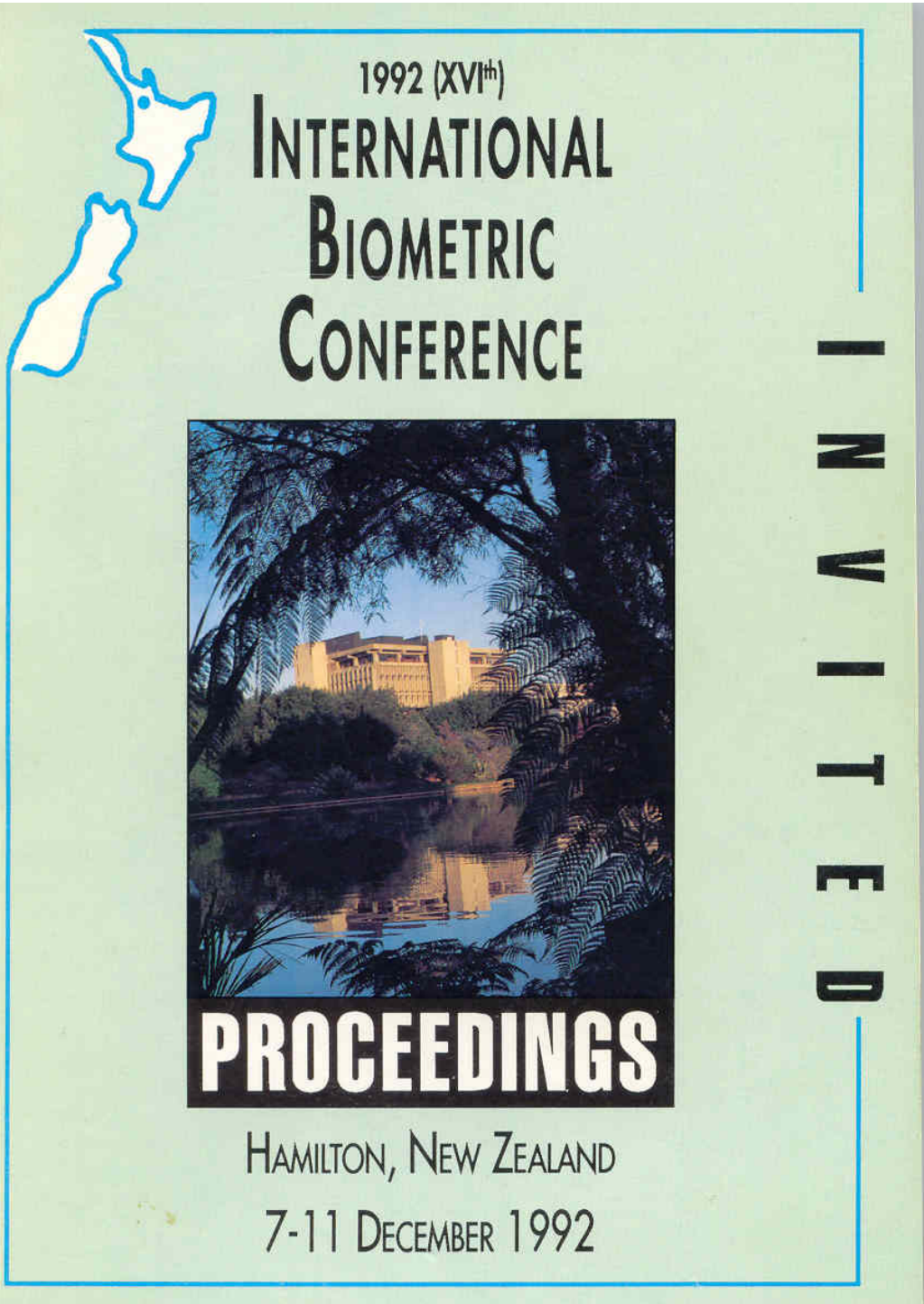
Division of International Health: Staff (currently)

- Prof. Dr. Dankmar Böhning
- Dr. Ekkehart Dietz
- Ronny Kuhnert (DFG)
- Ms. Sasivimol Rattanasiri (BMZ)
- Ms. Beatrice Chew (Secretary)
- Ms. Ina Schöttle (Research Assistant)

my motivation ...


MCN

... but in fact


1992 (XVIth)
INTERNATIONAL
BIOMETRIC
CONFERENCE
PROCEEDINGS
HAMILTON, NEW ZEALAND
7-11 DECEMBER 1992
INVITED

Overview

- History
- General Topics
- Current Areas of Interest and in Perspective
  - Profile Likelihood in Multicenter Studies
  - Capture-Recapture based on Counting Distibutions
  - Evaluation of Cumulative Evidence for Freedom of Disease with Application to BSE

Overview

- History
- General Topics
- Current Areas of Interest and in Perspective
  - Profile Likelihood in Multicenter Studies
  - Capture-Recapture based on Counting Distibutions
  - Evaluation of Cumulative Evidence for Freedom of Disease with Application to BSE

Personal Background

- Studies
  - Mathematics (main) and social sciences (Bielefeld and Berlin)
- Degrees
  - M.Sc. (optimal design) Dr. (algorithms)
  - Habil. (medicine: epidemiology/biometry)
- Cooperation
  - Numerous Institutions in Europe, USA, Australia, Thailand, and Philippines
- Visiting
  - 85-86 Statistics, PennState
  - 96 Psychology, Vienna
  - 98-99 Statistics, Munich
  - 04 International EpiLab, Copenhagen
  - Several Visits to Philippines and Thailand

History

- 1982: after completion of my PhD take up junior position at the Institute of Social Medicine
- 1992: v. l. in Medical Statistics and Epidemiolgy
- 2000: Award of the Title of *Professor*
- Several co-workers 1990-2004: Dietz, Kuhnert, Malzahn, Schlattmann, Stallmann, Schleinitz, ...

History

- 1981: Professor Schelp became Head of Institute for Social Medicine with emphasis on Epidemiology
- Professor Schelp active (at that time) with several intervention projects in NE Thailand

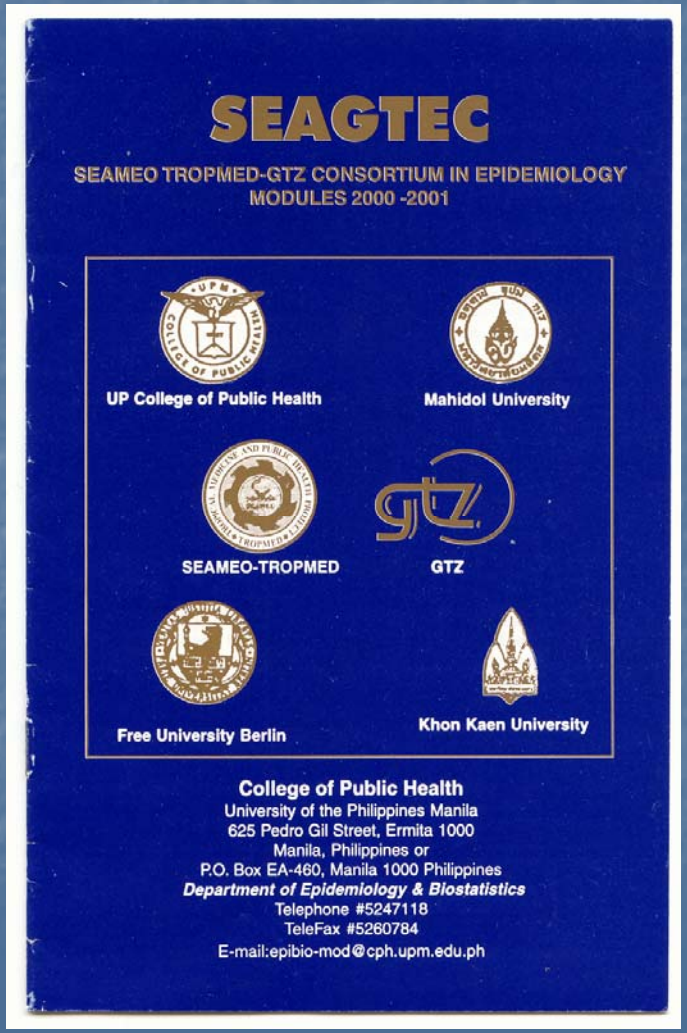
History

Besides cooperating in several projects in SE Asia

one major activity 1990-2000:

M.Sc. in Epidemiology

at UP Manila under participation of the universities of Mahidol (Bangkok, Th), Khon Kaen(Th), FU Berlin, UP Manila (Ph)



Cooperation Projects with SE ASIA

- Partner: Faculty for Public Health, Mahidol University, Bkk, Thailand
- Prof. Chukiat Viwatwongkasem (Counterpart)
- Funding: DFG, BMZ und National Research Council of Thailand (NRCT)

Capture-Recapture Procedures in Public Health

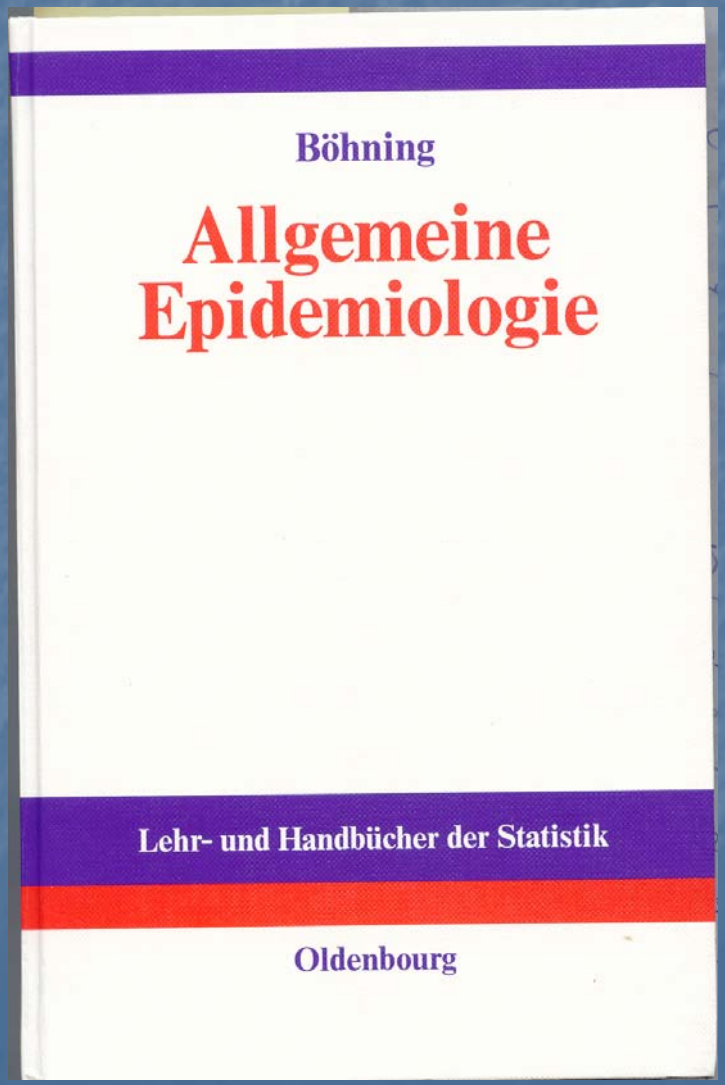
Surveillance Project on Illicit Drug Use in Thailand using Truncated Counting Distributions

Overview

- History
- General Topics
- Current Areas of Interest and in Perspective
  - Profile Likelihood in Multicenter Studies
  - Capture-Recapture based on Counting Distibutions
  - Evaluation of Cumulative Evidence for Freedom of Disease with Application to BSE

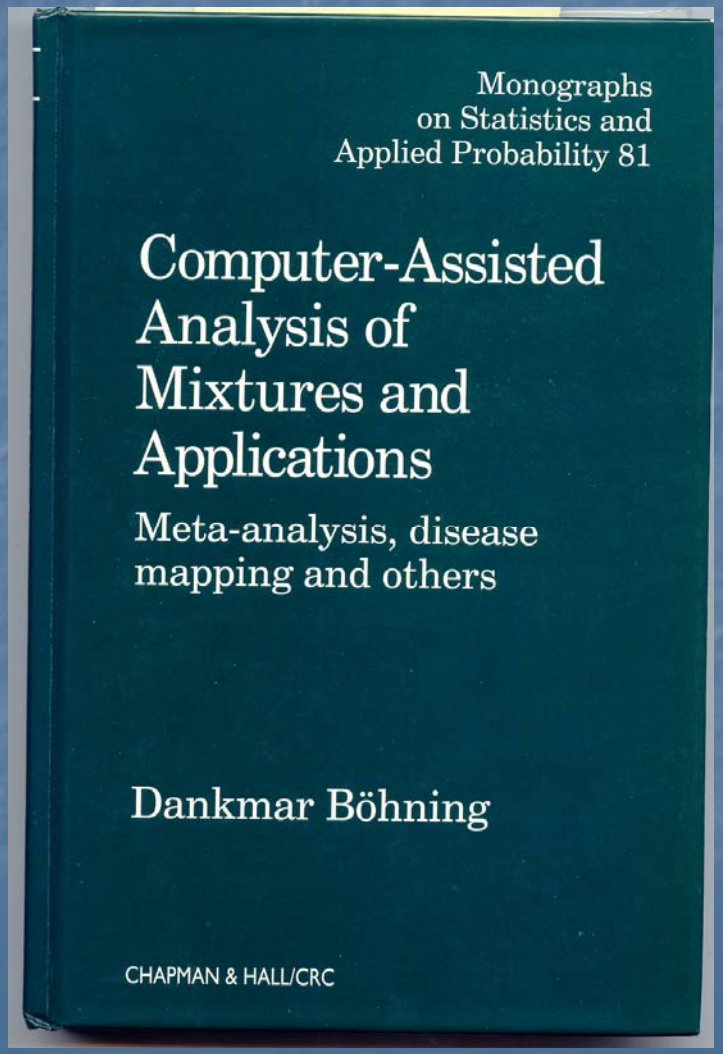
General Topics

- General Epidemiology
- Problems of Inference in Epidemiology
- Epidemiologic Modelling



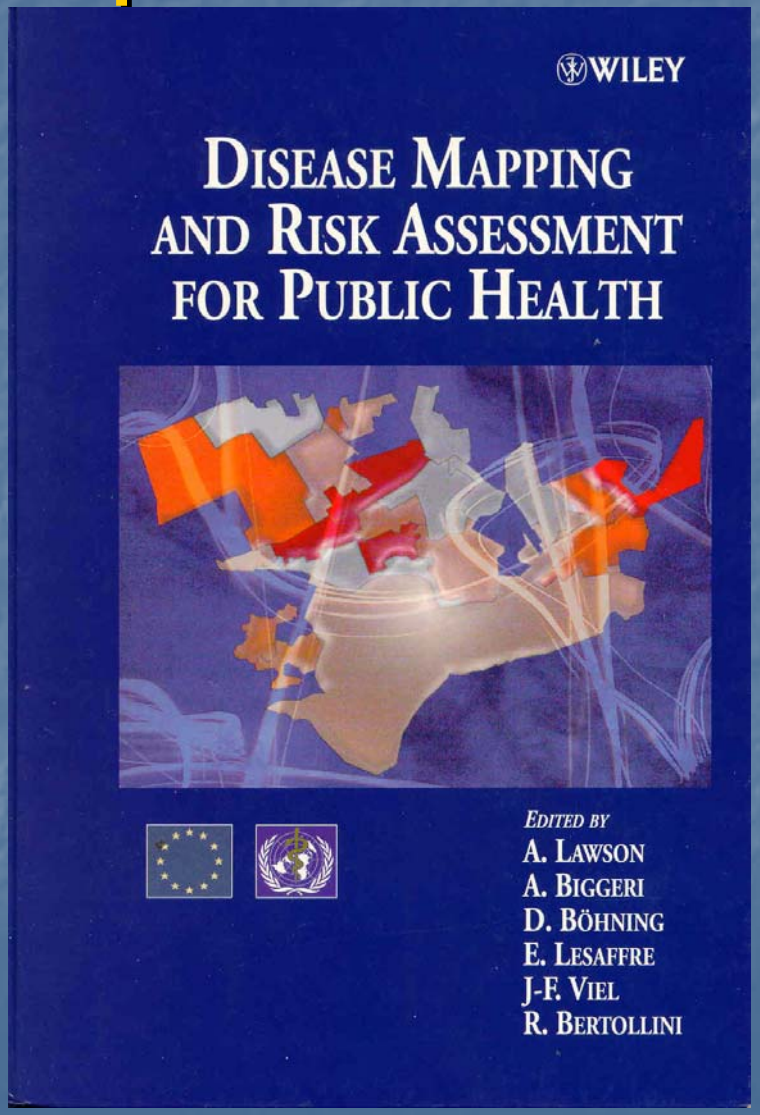
General Topics

- Mixture models
- Applications in Biometry and Epidemiology



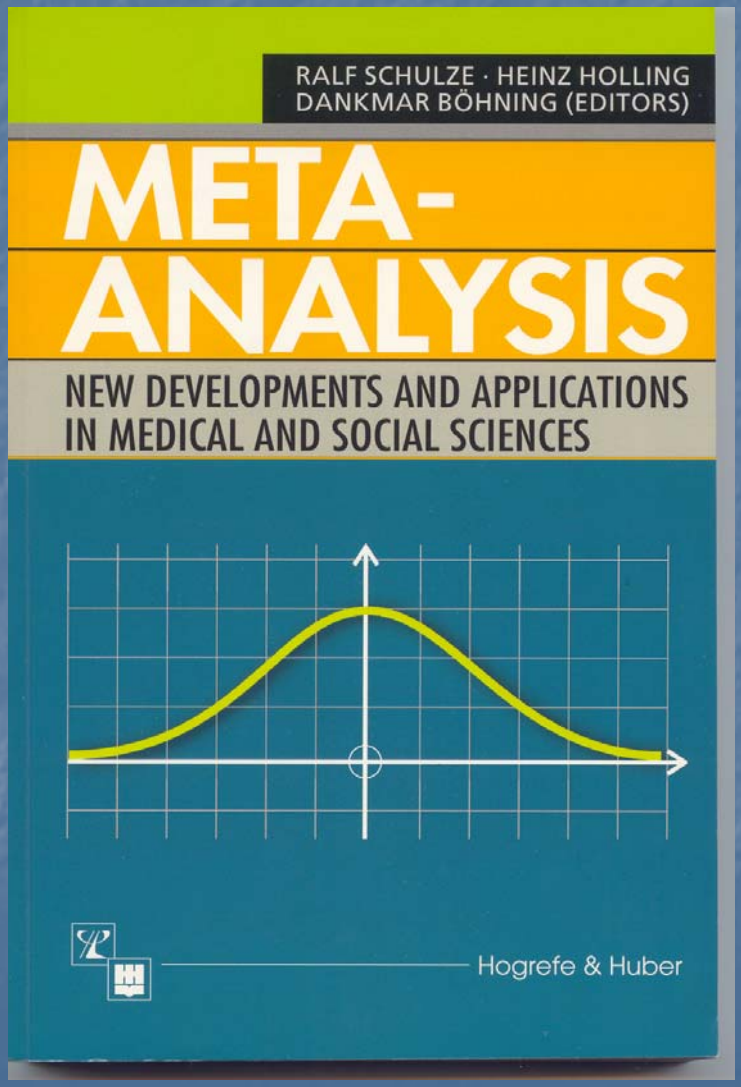
General Topics

- Disease Mapping and Geographical Epidemiology
- Smoothed Estimates of Geographical Risk



General Topics

- Systematic Reviews and Meta-Analysis
- Heterogeneity, Covariate and Publications Bias Modelling
- Unifying Concept



Personal Background: Editorial Board

- Biometrics (1992)
- Statistical Modelling (1999)
- Biometrical Journal (2004)

Overview

- History
- General Topics
- Current Areas of Interest and in Perspective
  - Profile Likelihood in Multicenter Studies
  - Capture-Recapture based on Counting Distibutions
  - Evaluation of Cumulative Evidence for Freedom of Disease with Application to BSE

Modelling Effect- and Nuisance Parameter in Multi-Center Studies

- Typical Setting: Treatment- and Control Arm
- For treatment arm:

$$x^T \text{ number of events}$$

$$P^T \text{ person-time}$$

$$\lambda^T \text{ event rate}$$

- For control arm:

$$x^C, P^C, \lambda^C$$

Modelling Effect in Multi-Center Studies: A Typical Example

Prophylactic Use of Lidocaine after Heart Attack (AMI) (after Normand 99)

Center	Treatment		Control		
	events x_i^T	person-time P_i^T	events x_i^C	person-time P_i^C	RR (95% CI)
1	2	39	1	43	2.21 (0.21-23.4)
2	4	44	4	44	1.00 (0.27-3.75)
3	6	107	4	110	1.54 (0.45-5.31)
4	7	103	5	100	1.36 (0.45-4.14)
5	7	110	3	106	2.25 (0.60-8.47)
6	11	154	4	146	2.61 (0.85-8.01)

Modelling Effect in Multi-Center Studies: Typical Issues

- Combination of Information
- Valid Combination: Homogeneity or Heterogeneity
- Modelling of Heterogeneity
  - Observed Heterogeneity: covariates
  - Unobserved Heterogeneity: mixtures

Modelling Effect- and Nuisance Parameter in Multi-Center Studies

- parameter of interest:

$$\text{risk ratio: } \theta = \lambda^T / \lambda^C$$

- nuisance parameter:

λ^C event rate in control arm

poisson log-likelihood (for one center):

$$-\lambda^T P^T + x^T \log(\lambda^T P^T) \quad -\lambda^C P^C + x^C \log(\lambda^C P^C)$$

Modelling Effect- and Nuisance Parameter in Multi-Center Studies

$$-\lambda^T P^T + x^T \log(\lambda^T P^T) \quad -\lambda^C P^C + x^C \log(\lambda^C P^C)$$

becomes using $\theta = \lambda^T / \lambda^C$ or $\lambda^T = \theta\lambda^C$

$$-\theta\lambda^C P^T + x^T \log(\theta\lambda^C P^T) \quad -\lambda^C P^C + x^C \log(\lambda^C P^C)$$

Keeping the parameter of interest fixed and maximizing for the nuisance parameter ...

$$\hat{\lambda}^C = \frac{x^C + x^T}{P^C + \theta P^T}$$

replacing λ^C by its estimate $\hat{\lambda}^C$

$$-\theta\hat{\lambda}^C P^T + x^T \log(\theta\hat{\lambda}^C P^T) \; - \; \hat{\lambda}^C P^C + x^C \log(\hat{\lambda}^C P^C)$$

leads to the beautiful simple Profile Log-likelihood ...

$$x^T \log(\theta) \; - (x^T + x^C) \log(P^C + \theta P^T)$$

... building the profile over all centers:

$$\sum_{i=1}^{k} x_i^T \log(\theta_i) \; - (x_i^T + x_i^C) \log(P_i^C + \theta_i P_i^T)$$

Advantages

- nuisance parameter eliminated
- Profile likelihood is simple (in this case):

$$\sum_{i=1}^{k} x_i^T \log(\theta_i) \; - (x_i^T + x_i^C) \log(P_i^C + \theta_i P_i^T)$$

- beneficial not only for effect structures but also for covariance structures (simplification of Fisher information)

Problems looked at: homogenous case

$\theta_i = \theta$ for all centers $i = 1, \ldots, k$:

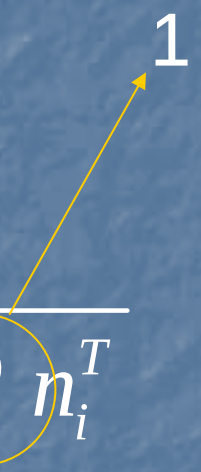
$$\sum_{i=1}^{k} x_i^T \log(\theta) \; - (x_i^T + x_i^C) \log(P_i^C + \theta P_i^T)$$

Score equation for profile MLE gives:

$$\hat{\theta} = \frac{\sum_{i=1}^{k} x_i^T n_i^C w_i(\hat{\theta})}{\sum_{i=1}^{k} x_i^C n_i^T w_i(\hat{\theta})}, \quad w_i(\hat{\theta}) = \frac{1}{n_i^C + \hat{\theta}\, n_i^T}$$

Problems looked at: homogenous case

$$\hat{\theta} = \frac{\sum_{i=1}^{k} x_i^T n_i^C w_i(\hat{\theta})}{\sum_{i=1}^{k} x_i^C n_i^T w_i(\hat{\theta})}, \quad w_i(\hat{\theta}) = \frac{1}{n_i^C + \hat{\theta}\, n_i^T}$$



- Close connection to Mantel-Haenszel:
  - arms balanced then: PMLE = MH
  - Non-sparsity: PMLE and MH close
  - Sparsity: PMLE more efficient

Overview

- History
- General Topics
- Current Areas of Interest and in Perspective
  - Profile Likelihood in Multicenter Studies
    - Research Areas in Preparation
  - Capture-Recapture based on Counting Distibutions
  - Evaluation of Cumulative Evidence for Freedom of Disease with Application to BSE

Modelling Effect-Heterogeneity in Multi-Center Studies: Unobserved Heterogeneity

- Allowing for unobserved heterogeneity leads to mixtures of profile log-likelihoods

$$\sum_{i=1}^{k} \log \int_{\theta} [\theta^{x_i^T} / (P_i^C + \theta P_i^T)^{x_i^T + x_i^C}] \, Q(d\theta)$$

- where mixing distribution can be parametric
- or non-parametric
  - strong results on NPMLE possible using convex theory
  - estimation with EM or global ascent algorithms

Modelling Effect-Heterogeneity in Multi-Center Studies: Unobserved Heterogeneity

- Comparison with other appraoches such as
  - approximating normal (problem: use empirical estimate of trial variance)

$$\sum_{i=1}^{k} \log \int_{\lambda^C} \phi(\ (z_i - \log\theta)/\sigma_i)\ Q(d\log\theta)$$

where z_i obs. log-rate ratio and $\sigma_i^2 = 1/x_i^T + 1/x_i^C$

  - multi-level approach (a la Murray Aitkin)

$$\sum_{i=1}^{k} \log \int_{\lambda^C} [\exp(-\theta\lambda^C P_i^T)(\theta\lambda^C P_i^T)^{x_i^T}\ \exp(-\lambda^C P_i^C)(\lambda^C P_i^C)^{x_i^C}]\ Q(d\lambda^C)$$

Modelling Effect-Heterogeneity in Multi-Center Studies: Observed Heterogeneity-Covariate Information

- Often additonal trial information is available s.a. study date, treatment modifications, patient characteristics
- Suppose information is captured in a covariate vector

z_i for center i: (GLM-type formulation)

$$\theta_i = \exp(\beta_0 + \beta' z_i)$$

Modelling Effect-Heterogeneity in Multi-Center Studies: Observed Heterogeneity-Covariate Information

Log-likelihood becomes

$$\sum_{i=1}^{k} x_i^T \log\theta_i - (x_i^T + x_i^C)\log(P_i^C + \theta_i P_i^T) \quad \text{using } \theta_i = \exp(\beta_0 + \beta' z_i)$$

$$= \sum_{i=1}^{k} x_i^T (\beta_0 + \beta' z_i) - (x_i^T + x_i^C)\log[P_i^C + \exp(\beta_0 + \beta' z_i) P_i^T]$$

- Strong results possible:
  - Hessian has simple structure
  - Hessian has lower bound (lower bound algorithm possible)
  - Guaranteed convergence to MLE

Overview

- History
- General Topics
- Current Areas of Interest and in Perspective
  - Profile Likelihood in Multicenter Studies
  - Capture-Recapture based on Counting Distibutions
  - Evaluation of Cumulative Evidence for Freedom of Disease with Application to BSE

Capture-Recapture Procedures based upon Counting Distributions

- Basic objective of CR: estimate population size
- In particular of interest in areas where direct counting is difficult such as
  - a wildlife population (historic genesis)
  - how many people drive a car without license?
  - how many practicing physicians are alcohol dep.?
  - how may cases of a disease remain undetected?
- Adjustment for undercount


N
Chiang Rai
Mae Hong Son
Chiang Mai
Laos
Nong Khai
Loei
Sukhothai
Mae Sot
Burma
Khon Kaen
Thailand
Ubon
Ratchatani
Ayuthaya
Kanchanaburi
BANGKOK
Cambodia
Pattaya
Cha Am
Hua Hin
Rayong
Koh Samet
Koh Chang
Chumpon
Gulf of Thailand
Koh Phangan
Koh Samui
Surat Thani
Phuket
Krabi
Andaman
Sea
Songkhla
Had Yai
Koh Turatao
Malaysia

Capture-Recapture Procedures based upon Counting Distributions

- Basic objective of CR: estimate population size
- In particular of interest in areas where direct counting is difficult such as
  - a wildlife population (historic genesis)
  - how many people drive a car without license?
  - how many practicing physicians are alcohol dep.?
  - how may cases of a disease remain undetected?
- Adjustment for undercount

How many cases *N* in a population?

- Some mechanism identifies n cases
- p_0 probability of being **not** identified by the mechanism
- **Then:**

$N = N p_0 + (1 - p_0) N$

= unobserved + observed cases

$= N p_0 + n$

$$\widehat{N} = n / (1 - p_0)$$

(Horwitz-Thompson)

Horwitz-Thompson-Approach seems easy, but …

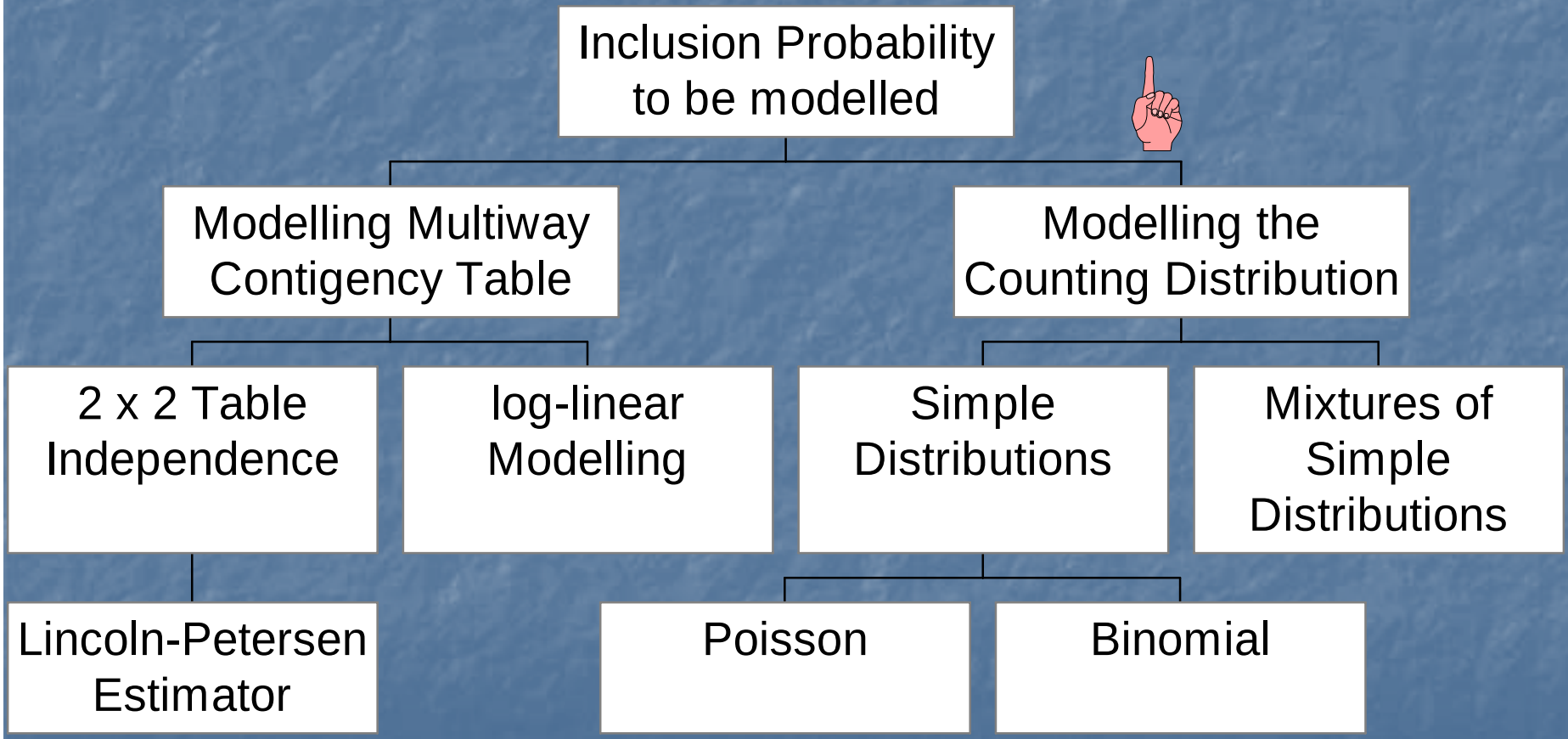
inclusion probability often unknown

and consequently,

approaches differ in the way they estimate the inclusion probability,

or in other words, how they model p_0 

Developments



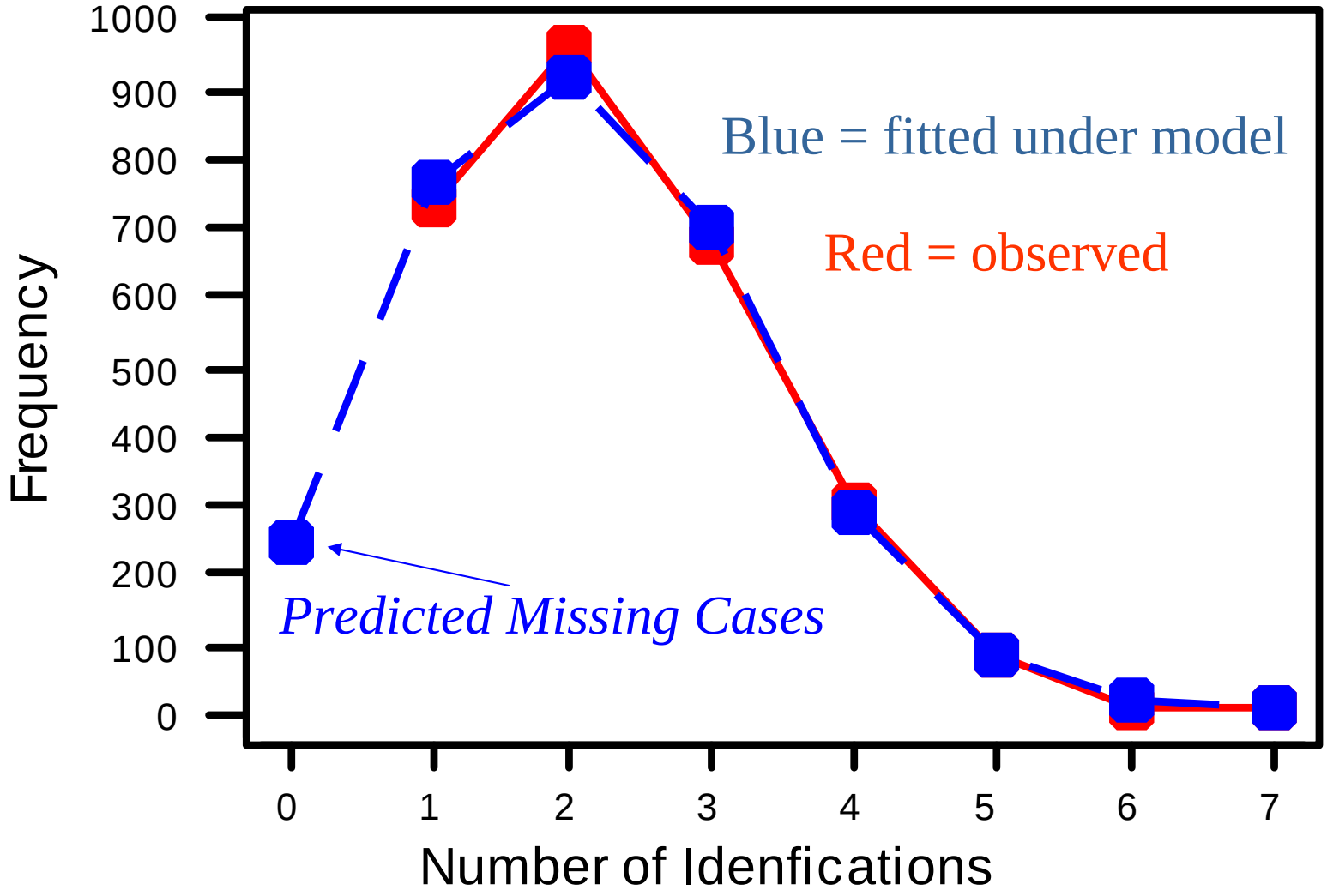
The Counting Distribution

... occurs when the mechanism can catch multiple identifications (s.a. police identifies and expells an illegal immigrant several times)

Count of identifications *i*	**Frequency of counts with *i* identifications**	**observed**
0	n_0	no
1	n_1	yes
2	n_2	yes
3	n_3	yes
4	n_4	yes
…	…	…

Distribution of Observed and Predicted Counts of Sources

for fictional data of multiple identifications



The Counting Distribution: A historic Example

- McKendrick´s cholera data
- Village in India had households with cholera cases $n_1=32$, $n_2=16$, $n_3=6$, $n_4=1$
- McKendrick ignored the houses with no cases
- Constructed an estimate (moment) based upon a Poisson assumption for the counts

Cholera Epidemic in an Indian Village (1915-1920)


House not affected, no cases


House affected, no cases

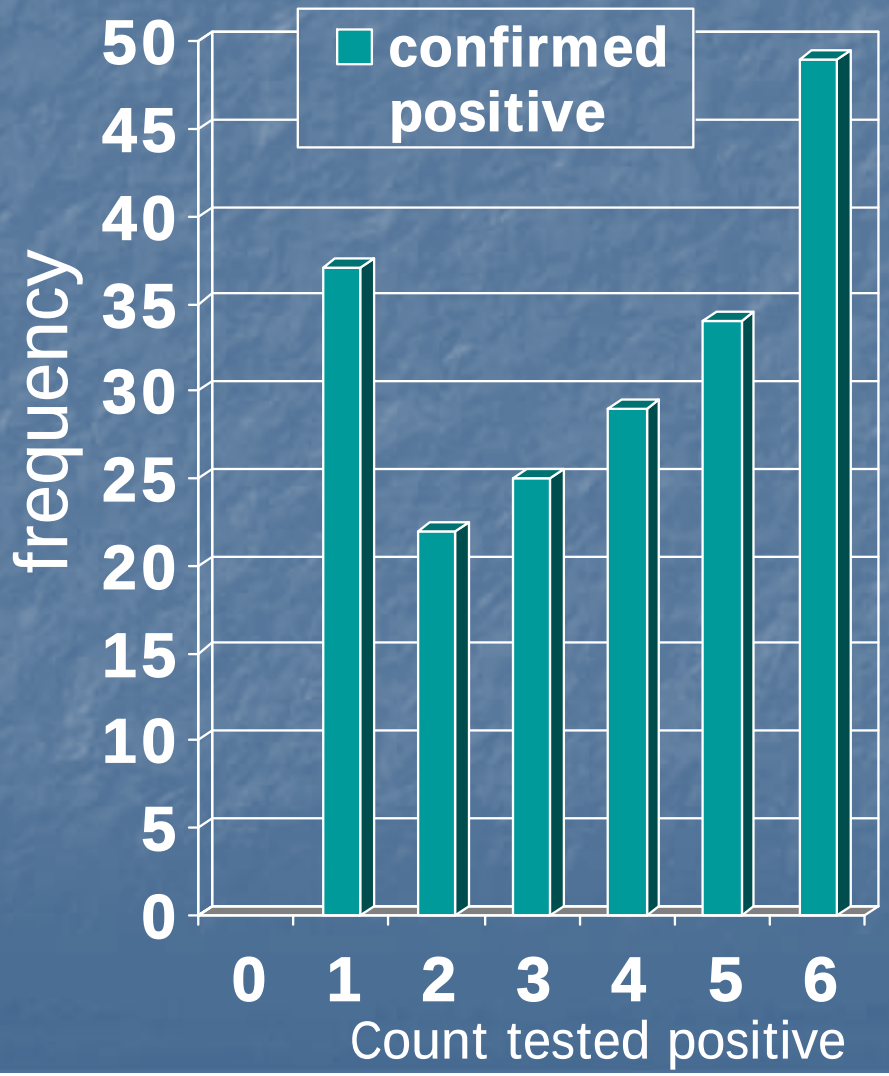

House affected, *m* cases

The counting distribution: a recent example from screening

- Lloyd & Frommer (2004, Applied Statistics) screening for bowel cancer
- 38,000 men screened in Sidney at 6 consecutive days by means of self-tesing for blood in stools
- 3,000 tested positively at least once and cancer status evaluated
- 196 were confirmed positive to have bowel cancer
- How many of 35,000 unconfirmed negative have bowel cancer?

The counting distribution: a recent example from screening

- frequency n_0 of those tested negative at all 6 times with bowel cancer is unknown
- an estimate of n_0 might be constructed from the distribution $n_1, n_2, n_3 \ldots$ of counts



Simple Distributional Count Models

Poisson (for unobservable counts)

$$f(y,\theta) = e^{-\theta}\theta^{y} / y! \ , \ y = 0,1,2 \ ...$$

truncated Poisson (for observable counts)

$$f(y,\theta) = \frac{1}{1-e^{-\theta}} \ e^{-\theta}\theta^{y} / y! \ , \ y = 1,2 \ ...$$

Predicted Probability of a Zero:

$$p_0 = f(y,\theta) = e^{-\theta}$$

Simple Distributional Count Models

after θ is identified ...

.... probability of a zero count:

$$p_0 = f(y=0,\theta) = e^{-\theta}$$

$$\Rightarrow \widehat{N} = \frac{n}{1-p_0} = \frac{n}{1-e^{-\theta}}$$

ML-Estimation in Zero-Truncated Poisson Models

Step 1: suppose $\hat{n}_0$ would be available

$$\hat{\theta} = \frac{1}{n + \hat{n}_0} \sum_{i=1}^{m} i\, n_i$$

Step 2: suppose $\hat{\theta}$ would be available

$$\hat{N} = \frac{n}{1 - p_0} = \frac{n}{1 - e^{-\hat{\theta}}} \Rightarrow \hat{n}_0 = \hat{N} - n = n \frac{e^{-\hat{\theta}}}{1 - e^{-\hat{\theta}}}$$

EM-Algorithm

Step 1 (M-Step): suppose $\hat{n}_0$ would be available

$$\hat{\theta} = \frac{1}{n + \hat{n}_0} \sum_{i=1}^{m} i \; n_i$$

Step 2 (E-Step): suppose $\hat{\theta}$ would be available

$$\hat{n}_0 = E(n_0 \mid \hat{\theta}; n_1, n_2, \ldots) = n \frac{p_0}{1 - p_0} = n \frac{e^{-\hat{\theta}}}{1 - e^{-\hat{\theta}}}$$

ML-Estimation in Zero-Truncated Count Models

general count distribution

$$f(y,\theta)\ ,\ y=0,1,2,\ldots$$

assoc. zero-truncated distribution

$$\frac{1}{1-f(0,\theta)}\,f(y,\theta)\ ,\ y=1,2,\ldots$$

EM-Algorithm

Step 1 (M-Step): suppose $\hat{n}_0$ is given:

$$\hat{\theta} = MLE, \text{ based upon } \hat{n}_0, n_1, n_2, \ldots$$

Step 2 (E-Step): suppose $\hat{\theta}$ is given:

$$\hat{n}_0 = E(n_0 \mid \hat{\theta}; n_1, n_2, \ldots) = n\frac{p_0}{1-p_0} = n\frac{f(0,\hat{\theta})}{1-f(0,\hat{\theta})}$$

More flexible and robust approach through mixtures

- Simple counting sources distributions such as Binomial and Poisson require assumptions such as homogeneity of identification probabilities that are seldom met in reality
- allowing the identification probability to vary in unobserved sub-populations will be more realistic

The mixture approach in a nutshell

mixture density:

$$f(y,\theta) = f(y,\lambda_1)q_1 + \ldots + f(y,\lambda_k)q_k$$

$f(y,\lambda)$ is component density

(Example: $f(y,\lambda) = e^{-\lambda}\lambda^y / y!$)

$$\theta = \begin{pmatrix} \lambda_1 & \ldots & \lambda_k \\ q_1 & \ldots & q_k \end{pmatrix}$$ is mixing distribution

Nested EM-Algorithm

Step 1 (M-Step): suppose $\hat{n}_0$ is given:

$\hat{\theta} = MLE$ of mixing distribution $\theta = \begin{pmatrix} \lambda_1 & \dots & \lambda_k \\ q_1 & \dots & q_k \end{pmatrix}$

provided by EM algorithm for mixtures

Step 2 (E-Step): suppose $\hat{\theta}$ is given:

$$\hat{n}_0 = E(n_0 \mid \hat{\theta}; n_1, n_2, \dots) = n \frac{p_0}{1 - p_0}$$

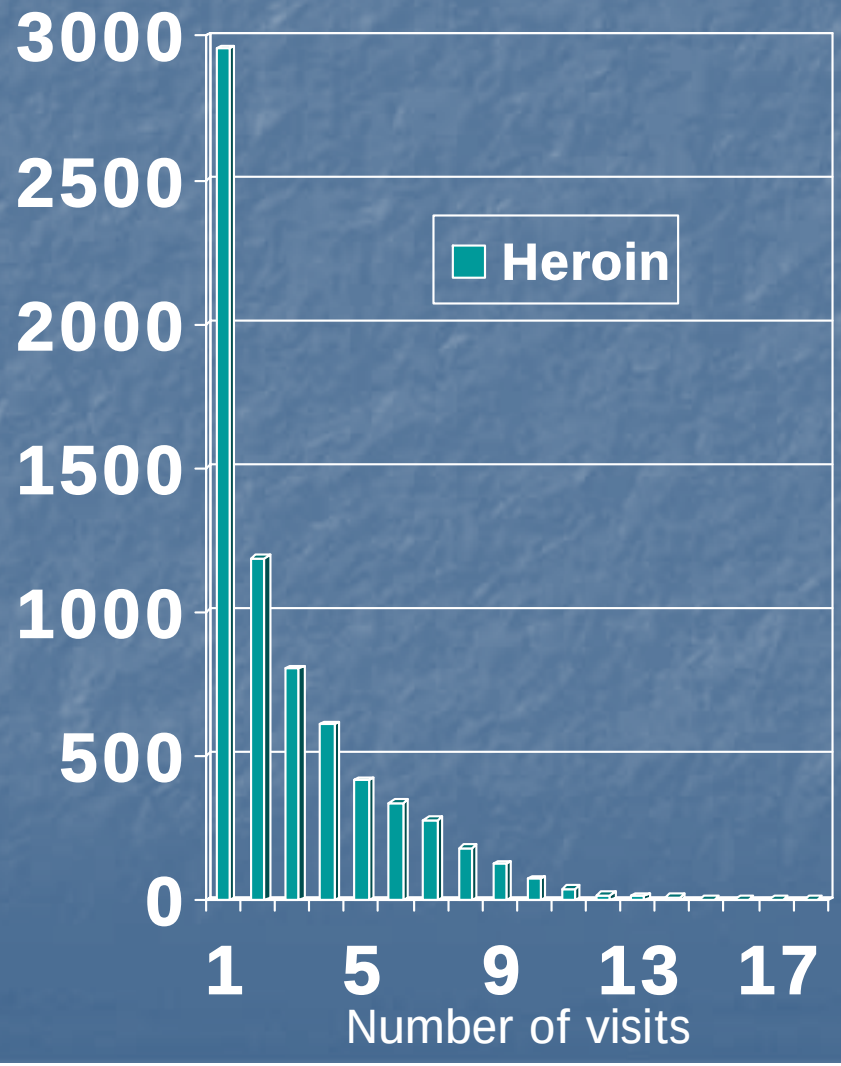
$$= n \frac{f(0, \hat{\theta})}{1 - f(0, \hat{\theta})} = n \frac{\hat{q}_1 e^{-\hat{\lambda}_1} + \dots + \hat{q}_k e^{-\hat{\lambda}_k}}{1 - (\hat{q}_1 e^{-\hat{\lambda}_1} + \dots + \hat{q}_k e^{-\hat{\lambda}_k})}$$

Application: surveillance study on drug use in Thailand

- Ministry of Public Health (Th) collects routinely data on drug use via the ONCB on drug users visiting treatment institutions
- In a pilot study (Böhning, Busaba, Chukiat et al. 2004 *EUJE)* CR-Poisson mixture model applied to data from 2002 (last quarter)
- Major emphasis on heroin and metamphetamin users

Application: surveillance study on drug use in Thailand

- Count distribution (counting number of visits) for heroin users
- n = 7,048 observed heroin users (2001, 4)



Counting contacts to treatment institutions not uncommon

- Previous modelling done primarily by practitioners with publications in
  - Addiction, Addiction Research & Theory, Journal of Drug Issues, Journal of Quantitative Cirminology
- Modelling uses primarily simple Poisson
  - simple to understand, to apply and use, and to communicate
  - however: often not appropriate
- better: semi-parametric models for counts such as Poisson mixtures

Some results

- n=7,048 (observed)
- N=17,278
- N-n=10,230 (hidden)
- Ratio: observed/hidden=0.69

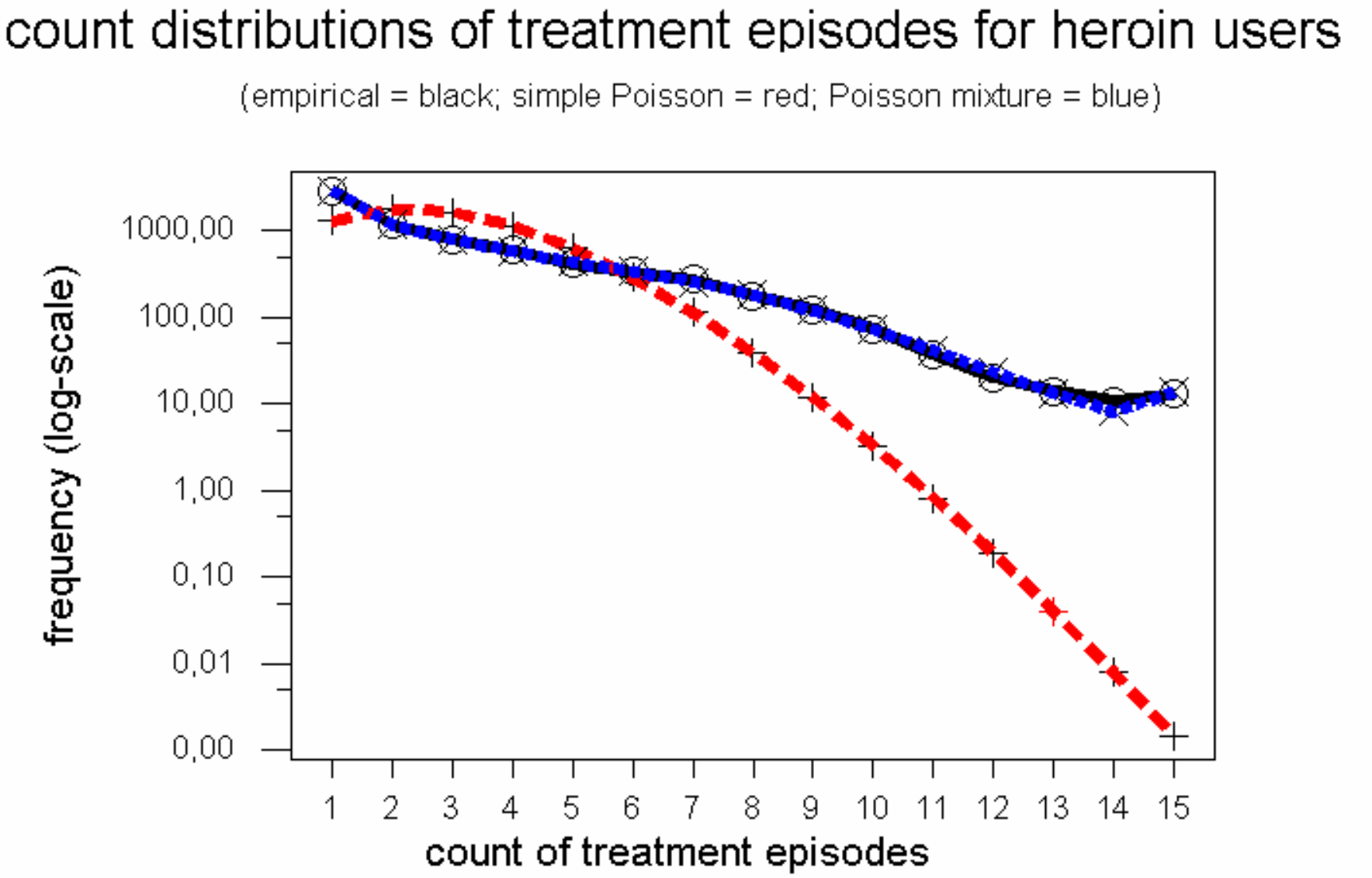
Estimating the Number of Heroin Users:

k	$\hat{\lambda}_j$	$\hat{q}_j$	log-likelih.	AIC	BIC	$\hat{N}$
1	2,75	1,00	-15462	-30927	-30934	7543
2	0,88	0,75	-13214	-26434	-26455	10226
	5,40	0,25				
3	0,41	0,69	-13134	-26279	-26313	13350
	2,97	0,22				
	6,80	0,09				
4	0,21	0,70	-13120	**-26255**	**-26303**	17278
	2,13	0,19				
	5,84	0,10				
	12,20	0,01				

$$AIC = 2 \times \text{log-likelihood} - (2k - 1)2$$

$$BIC = 2 \times \text{log-likelihood} - (2k - 1)\log(n)$$

Pilot study for Bangkok, 2001 (4)



take another look

Estimating the Number of Heroin Users:

k	$\hat{\lambda}_j$	$\hat{q}_j$	log-likelih.	AIC	BIC	$\hat{N}$
1	2,75	1,00	-15462	-30927	-30934	7543
2	0,88	0,75	-13214	-26434	-26455	10226
	5,40	0,25				
3	0,41	0,69	-13134	-26279	-26313	13350
	2,97	0,22				
	6,80	0,09				
4	0,21	0,70	-13120	**-26255**	**-26303**	17278
	2,13	0,19				
	5,84	0,10				
	12,20	0,01				

$$AIC = 2 \times \text{log-likelihood} - (2k - 1)2$$

$$BIC = 2 \times \text{log-likelihood} - (2k - 1)\log(n)$$

proof uses:

- a) Jensen's inequality
- b) mean of the MLE of the mixing distribution = sample mean

V. A Monotonicity Property for the Population Size Estimator

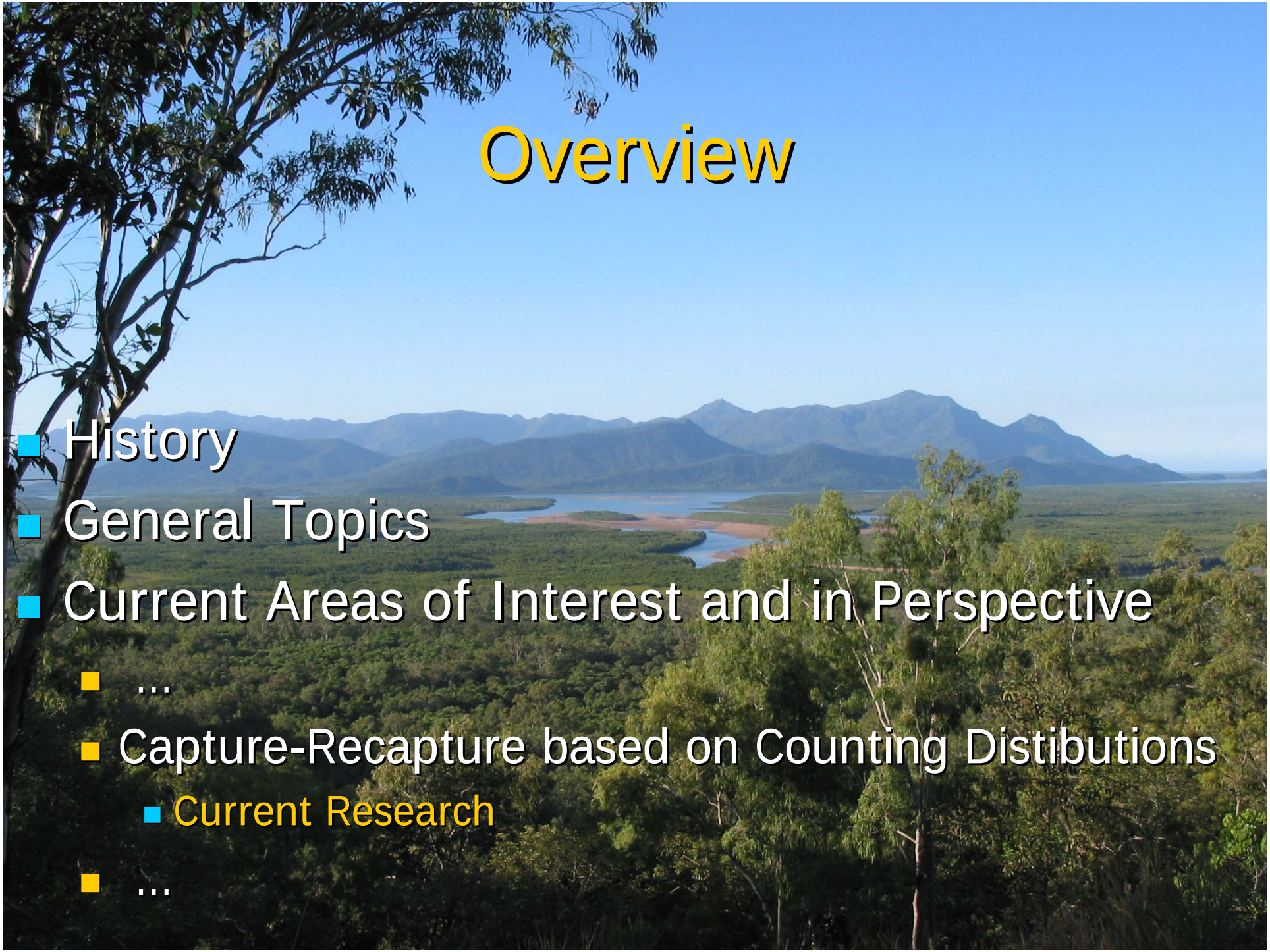
Result: Böhning and Schön (*JRSS C* 2004)

$\hat{N}_k$ MLE of population size w.r.t. a truncated Poisson mixture with k components, $k = 1, 2, \ldots$ Then:

$$\hat{N}_k \geq \hat{N}_1$$

likely, the **more general statement** is also true:

$$\hat{N}_{k+1} \geq \hat{N}_k$$


Overview
History
General Topics
Current Areas of Interest and in Perspective
...
Capture-Recapture based on Counting Distibutions
Current Research
...

Concluding Remarks

Open Problems and Research Questions

- Standard errors and confidence intervals
- Suitable modification of resampling techniques
- Validation studies
- Comparison to other approaches (Pollock-Norris or Zelterman)
- ... Mixtures of binomials

very recent work in perspective

- truncated mixture of Poisson distributions
- or ...
- mixture of truncated Poisson distributions

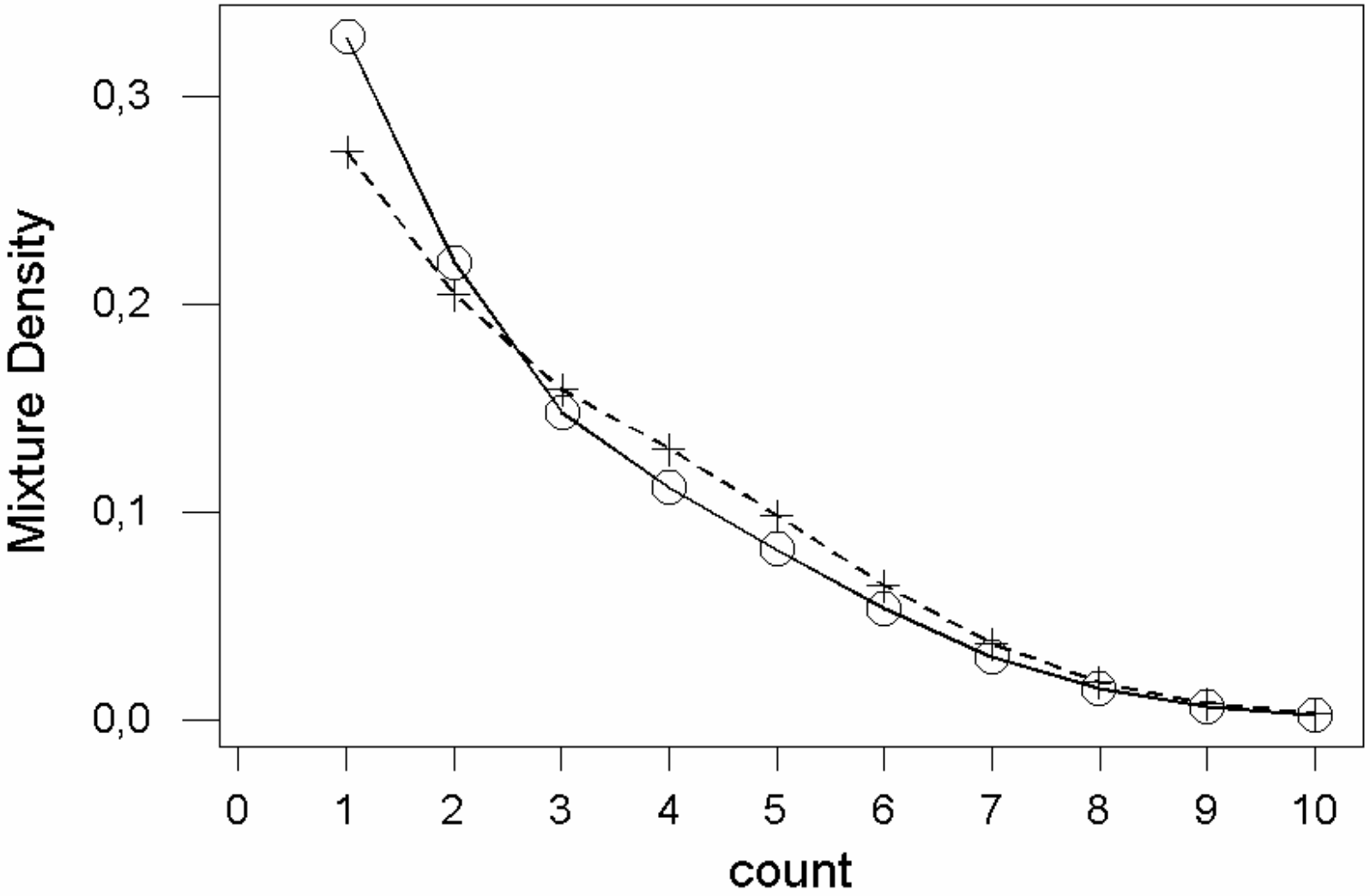
truncated Poisson mixture (dual model)

$$\frac{\sum_{j=1}^{k} q_j \; Po(y, \lambda_j)}{1 - \sum_{j=1}^{k} q_j \; Po(0, \lambda_j)}$$

mixture of truncated Poissons (primal model)

$$\sum_{j=1}^{k} q_j' \frac{Po(y, \lambda_j')}{1 - Po(0, \lambda_j')}$$

Illustration: dual model (ring) and primal model (+) use equal weights and component means 1 and 4



truncated Poisson mixture (dual model)

$$\frac{\sum_{j=1}^{k} q_j \; Po(y, \lambda_j)}{1 - \sum_{j=1}^{k} q_j \; Po(0, \lambda_j)}$$

- close to the original problem, easy to understand and to communicate
- But technical difficult, because of non-linearity

mixture of truncated Poissons (primal model)

$$\sum_{j=1}^{k} q_j' \frac{Po(y, \lambda_j')}{1 - Po(0, \lambda_j')}$$

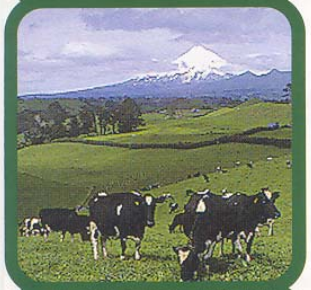
- less close to the original problem
- but convex problem with strong results available on NPMLE and global ML estimation

How are dual and primal model related?

- Böhning and Kuhnert (2005, JASA)
- Both share the same likelihood surfaces
- MLEs can be explicitly transformed into each other
- $\widehat{N} = \widehat{N}'$

Another area of interest

Help protect New Zealand.



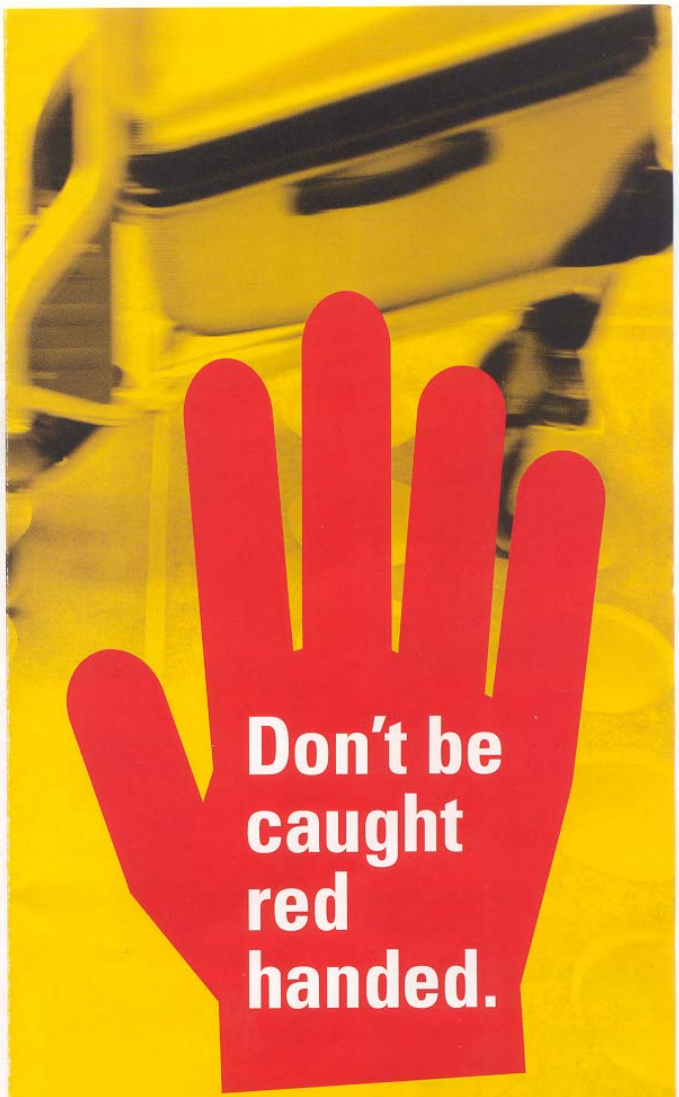
Protect New Zealand
Tiakina Aotearoa
www.protectnz.org.nz

New Zealand's remoteness and small population mean that less than 10 percent of the world's pests and diseases exist here. The future of New Zealand's environment, economy and way of life depends on keeping these pests and diseases out.

Our aquaculture, agriculture, horticulture, forestry and tourism, even our health could be seriously affected by the introduction of just one new pest or disease.

So please read this guide carefully. It is a quick reference list of goods you must not bring or must declare when entering New Zealand.

For details of certificates needed and Import Health Standard requirements to be met before any restricted item can be brought into New Zealand, contact MAF Quarantine Service on 0800 222 009. You can also visit www.protectnz.org.nz or www.maf.govt.nz



Don't be caught red handed.

Your quick guide to items you must not bring in to New Zealand.

Overview

- History
- General Topics
- Current Areas of Interest and in Perspective
  - Profile Likelihood in Multicenter Studies
  - Capture-Recapture based on Counting Distibutions
  - Evaluation of Cumulative Evidence for Freedom of Disease with Application to BSE

Idea of Project

- birth cohorts of animals (in different surveillance streams) are monitored for occurence of BSE
- in particular, prevalence is small, potentially cohort is disease-free
- in contrast to estimating prevalence, this project wants to answer the question:
- When can a particular cohort considered to be disease free?

π

Idea of Project
Basic Principle of the Sequential Trial

interest is in a prevalence parameter π
and associated null hypothesis

$$H_0 : \pi = 0$$

(implying, birth cohort is disease-free)

sequential trial (ST): animals are tested in discrete calendar or sequential time

Y_t result of testing animal t

$(y_t = 1$ test positive, $y_t = 0$ test negative$)$:

$H_0: \; Y_t = 0$ for all times $t = 1, 2, 3, \ldots$

clearly, $\Pr(Y_t > 0 \mid H_0) = 0$, for all times t

in other words, there is no type-I error

Y_1, Y_2, Y_3... series of BSE-tests:

waiting time T for first animal testing positive:

$$\Pr(T = t \mid \pi) = \pi(1-\pi)^{t-1}$$

has geometric distribution

T	sequence of tests	probability
1	1	π
2	01	$(1-\pi)\pi$
3	001	$(1-\pi)^2\pi$
4	0001	$(1-\pi)^3\pi$
....	.....	.....

Rationale of the ST:

since

$$\Pr(T > 0 \mid \pi) = \sum_{t=1}^{\infty} \pi(1-\pi)^{t-1} = 1$$

unless $\pi = 0$, there exists some positive time waiting time $s > 0$ such that

$$\Pr(0 < T \leq s \mid \pi) = 1 - \beta$$

for given arbitrary small $\beta > 0$

Rationale of the ST:

instead of waiting for all times ($T = \infty$)

to conclude with $\pi = 0$,

we wait until time $s < \infty$ such that

$$\Pr(0 < T \leq s \mid \pi) = \sum_{t=1}^{s} \pi(1-\pi)^{t-1} = 1-\beta$$

to conclude with $\pi = 0$, necessarily.

now,

$$\Pr(0 < T \leq s \mid \pi) = \sum_{t=1}^{s} \pi(1-\pi)^{t-1} = 1-(1-\pi)^{s}$$

and equating

$$1-(1-\pi)^{s} = 1-\beta$$

leads to

$$(1-\pi)^{s} = \beta$$

Idea of Project: Solution

$$(1-\pi)^s = \beta$$

from where the stopping time s

$$s = \frac{\log(\beta)}{\log(1-\pi)}$$

is deduced

Preliminary Results

project has focus on

power function:

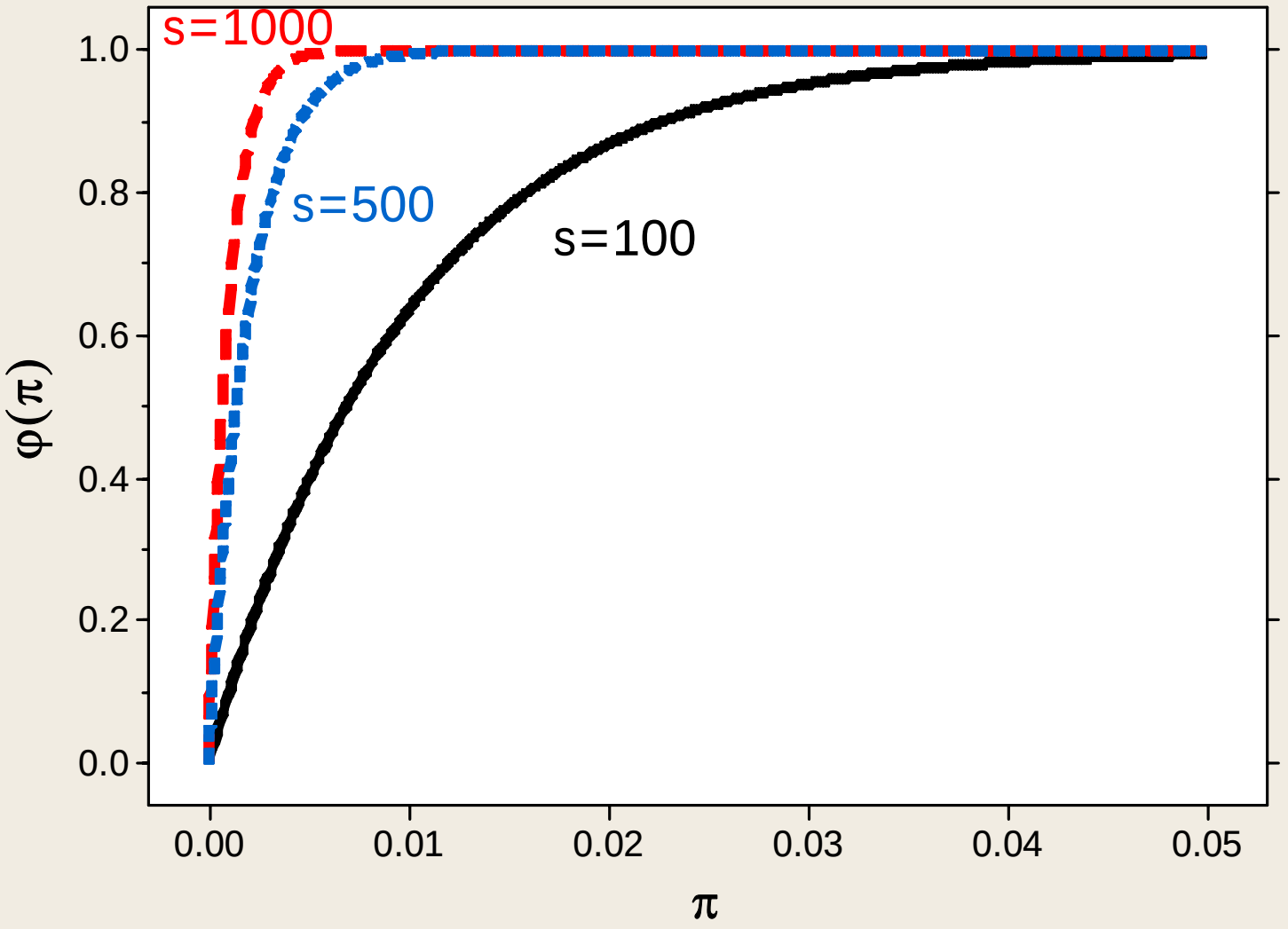
$$\varphi(\pi) = 1 - (1 - \pi)^s$$

Result: power function is monotone increasing

$$\pi_1 \leq \pi_2$$

$$\Rightarrow \varphi(\pi_1) \leq \varphi(\pi_2)$$

Monotonicity of power function



Important consequence

since true prevalence π is unknown, only minimum detectable prevalence (design prevalence) π_0 needs to be specified: it follows

$$\varphi(\pi_0) \leq \varphi(\pi)$$

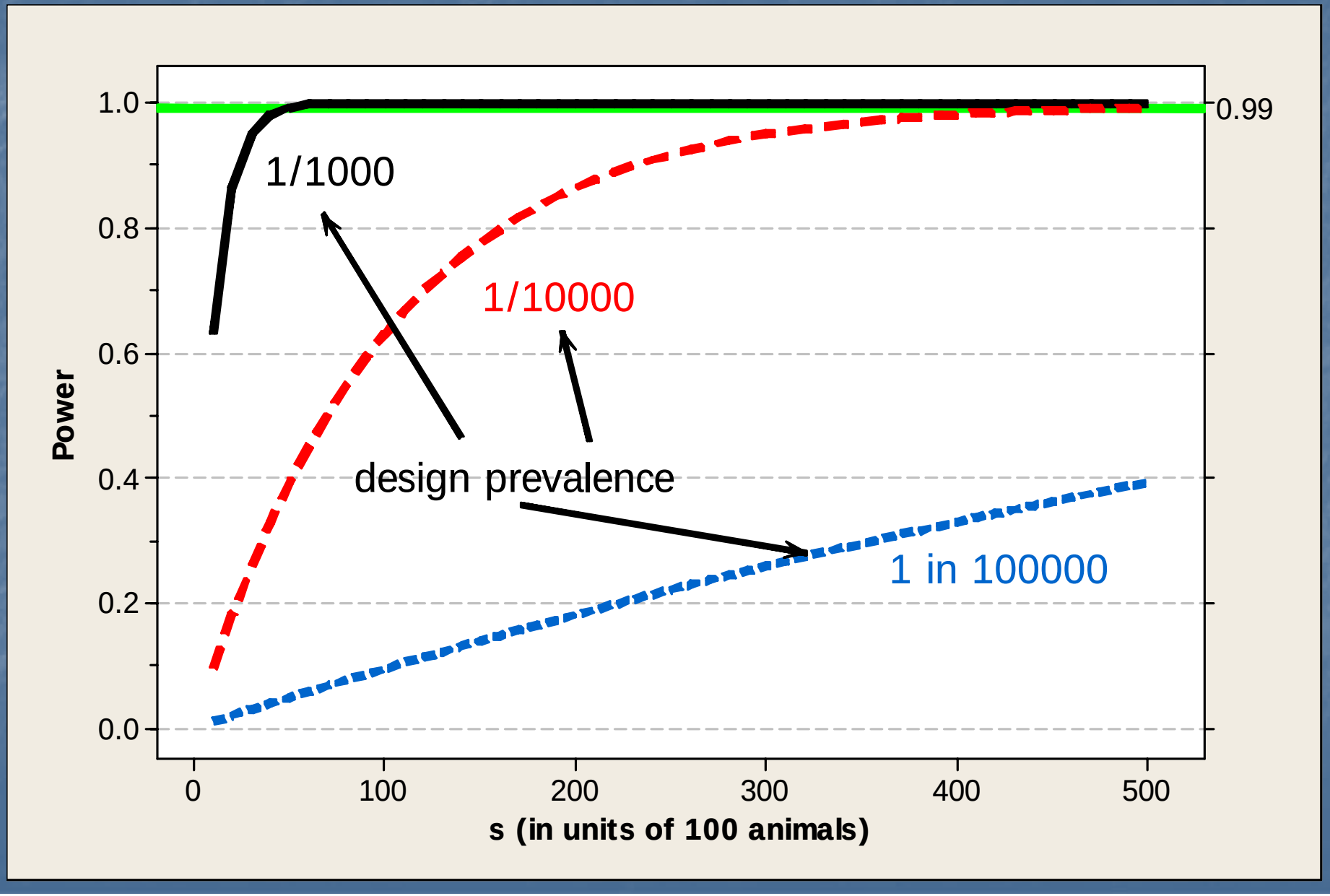
Power is also monotone in the waiting time s

power function

$$\varphi(s) = 1 - (1 - \pi)^s$$

(now as function of s)

Power as function of waiting time



Two Questions

What is the waiting time s to reach power of ...

$$1-\beta = 1-(1-\pi)^{s} \quad ?$$

from where the stopping time solution

$$s = \frac{\log(\beta)}{\log(1-\pi)}$$

is found

What is the waiting time *s* to reach power of ...

Design prevalence: 1 in	Power=0.99	Power=0.999
1000	4603	6904
10000*	46049	69074
100000*	460515	690772

* EC: Opinion in requirements for BSE/TSE Surveys, 2001

Which power have we reached given waiting time s ?

power =

$$\varphi(\pi) = 1 - (1 - \pi)^s$$

What power is reached given waiting time *s* ?

Design prevalence: 1 in	*s*=10000	*s*=100000
1000	0.999955	1.00000
10000*	0.632139	0.99995
100000*	0.095163	0.63212

* EC: Opinion in requirements for BSE/TSE Surveys, 2001

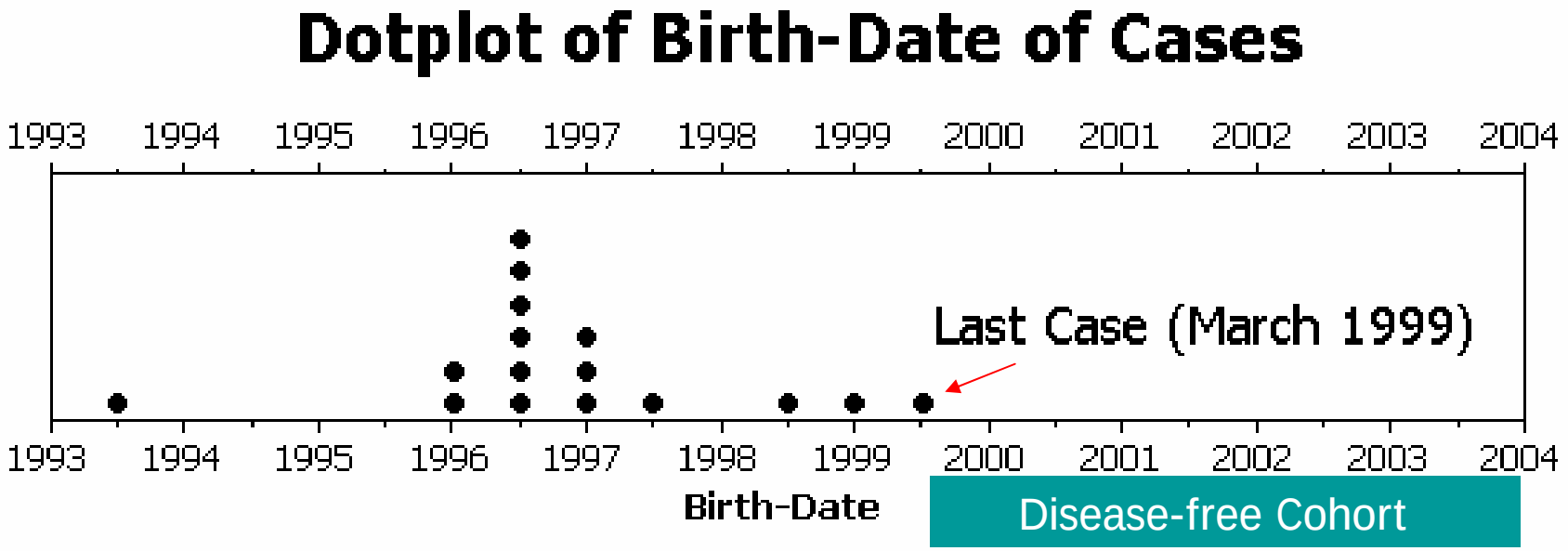
Situation in Denmark

- TSE Database: public register for BSE-testing
  - Controlled by the Danish Veterinary and Food Administration
  - Development, service and maintenance done by private company
  - Information on all animals tested for BSE since 01-Jan-2001

Situation in Denmark

- From TSE Database the following variables were made available for project:
  - Animal Identification Number
  - Age (at death)
  - Birth- and death-date
  - Cause of submission like clinical suspect, emergency slaughter, healthy slaughter,...
  - Result of BSE-testing (+/-)

Situation in Denmark: Identification of Positive Cases



Situation in Denmark

Rows: BIRTHMONTH Columns: BIRTHYEAR

	1999	2000	2001	2002	All
1	0	11154	5936	988	18078
2	0	11235	5636	692	17563
3	0	13852	6808	356	21016
4	17012	11285	6016	152	34465
5	14821	9766	4744	76	29407
6	12748	8292	3745	21	24806
7	14380	9131	3732	11	27254
8	14285	9078	3167	3	26533
9	13397	8342	2646	0	24385
10	12441	8112	2212	0	22765
11	11660	7236	1791	0	20687
12	11654	6781	1348	0	19783
All	122398	114264	47781	2299	286742

Situation in Denmark: achieved power given waiting time *s=286742*

Prevalence 1 in	Power	Prevalence 1 in	Power
10000*	1.0000	60000	0.9916
20000	1.0000	70000	0.9834
30000	0.9999	80000	0.9722
40000	0.9992	90000	0.9587
50000	0.9968	100000	0.9432

* EC: Opinion in requirements for BSE/TSE Surveys, 2001

Current Research

- Incorporating Non-Perfect Testing
  - Loss of power
- Incorporating population heterogeneity
  - Increase of power

Thank you!