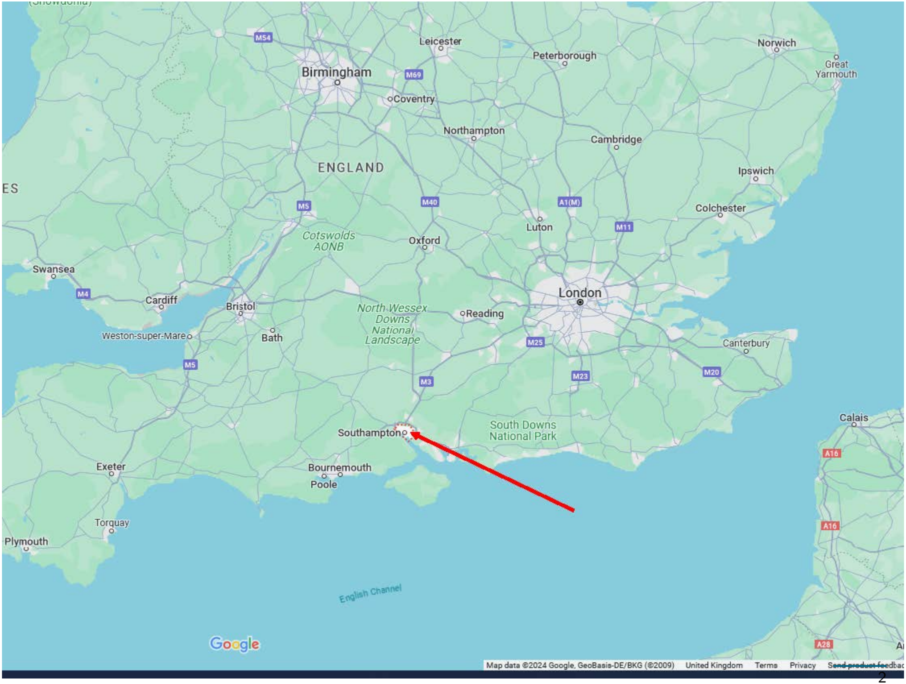
Applications of Uni-List Capture-Recapture Methods in Meta-Analysis

Dankmar Böhning
Southampton Statistical Sciences Research Institute
and Mathematical Sciences, University of Southampton, UK


Leicester
Peterborough
Norwich
Great Yarmouth
Birmingham
Coventry
Northampton
Cambridge
ENGLAND
Ipswich
Colchester
Luton
Oxford
Cotswolds AONB
Swansea
Cardiff
Bristol
Bath
Weston-super-Mare
North Wessex Downs National Landscape
Reading
London
Canterbury
Southampton
South Downs National Park
Calais
Exeter
Bournemouth
Poole
Torquay
Plymouth
English Channel
M54
M69
M5
M40
A1(M)
M11
M4
M25
M3
M23
M20
A16
A28
Google
Map data ©2024 Google, GeoBasis-DE/BKG (©2009) United Kingdom Terms Privacy

Southampton Statistical Sciences Research Institute (S3RI)

S3RI brings together staff from across the University for research in methods and applications of statistics.

Key facts

Unless otherwise stated

English language: IELTS 6.5, with minimum of 5.5 in each component, or an equivalent standard in other qualifications approved by the University, achieved within the past two years

Duration: up to four years (full-time); up to seven years (part-time); dependent on funding route: Integrated PhD, PhD, 1+3

Start date: September (Integrated PhD, PhD, 1+3); sometimes possible throughout the year (PhD only)

Applying: University application form with transcripts, research proposal and two references

Fees: www.southampton.ac.uk/pgfeesandfunding

Find out more

T: +44 (0)23 8059 7385
E: pgrapply.fshs@southampton.ac.uk
www.southampton.ac.uk/s3ri/pgp

"Being a part of S3RI is one of the most significant milestones in my career. The courses, delivered by excellent professionals from the University of Southampton and abroad, provided good insights into statistics. I have no doubt that S3RI will continue to grow and enrich with its highly qualified and professional academic staff. It has been a pleasure to be part of this prestigious group."

Carla Azevedo
S3RI PhD student

PhD Statistics

We have a lively and thriving community of postgraduate students engaged in research across a range of areas and we support them extensively. Supervisors, who are international experts in their field, provide in-depth training. You will be given a personal computer, a desk in a shared office and a conference attendance allowance. We offer a number of competitive studentships to cover fees and cost of living. The type of funding depends on the eligibility of the candidate.

Key facts: additional information

Entry requirements: first- or upper second-class bachelor degree in a relevant mathematical subject (for four year PhD). Masters in a relevant mathematical subject or first- or upper-second class degree (for PhD) in a relevant mathematical subject at MMath or MPhys level or equivalent, or satisfactory performance at interview
Assessment: progression from year one to year two of Integrated PhD by taught courses; annual reports, confirmation (for PhD award), thesis and viva*
Closing date: none, but funding decisions will be made from mid-March
Funding: www.southampton.ac.uk/maths/postgraduate/fees_and_funding.page

PhD Social Statistics

Social Statistics at Southampton has been awarded Doctoral Training Centre status by the ESRC. Full funding is available for strong applicants wishing to undertake frontier research.

Key facts: additional information

Entry requirements: first- or upper second-class degree (1+3 route). First- or upper second-class degree plus a masters at merit level (+3) in a relevant subject, or equivalent qualifications plus satisfactory performance at interview
Assessment: progression from year one on 1+3 by examination taught courses; annual reports, confirmation (for PhD award), thesis and viva*
Closing date: none, but early application advised
Funding: may be available through University's Vice-Chancellor's Scholarship programme
Additional costs: fieldwork, printing and photocopying, etc.; some help may be provided

* For more information on continued assessment throughout your research programme, see page 19

Carla Azevedo

Research themes

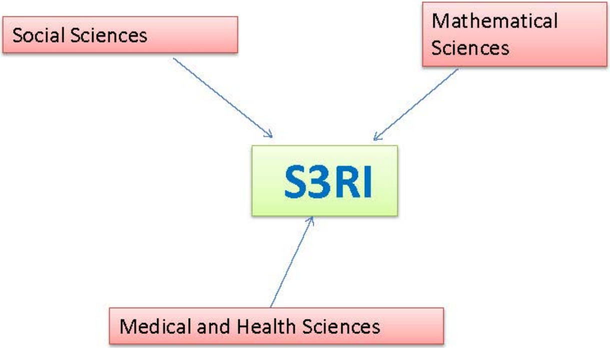
Biostatistics
www.southampton.ac.uk/s3ri/biostatistics

Design of experiments
www.southampton.ac.uk/s3ri/experiments

Policy and evaluation
www.southampton.ac.uk/s3ri/policyandevaluation

Statistical modelling
www.southampton.ac.uk/s3ri/modelling

Survey methods
www.southampton.ac.uk/s3ri/surveymethods


Social Sciences
Mathematical Sciences
S3RI
Medical and Health Sciences

obesity reviews

doi: 10.1111/obr.12014

Obesity Treatment

Risk of completed suicide after bariatric surgery: a systematic review

C. Peterhänsel[1,2], D. Petroff[3,4], G. Klinitzke[1,2], A. Kersting[1] and B. Wagner[1,2]

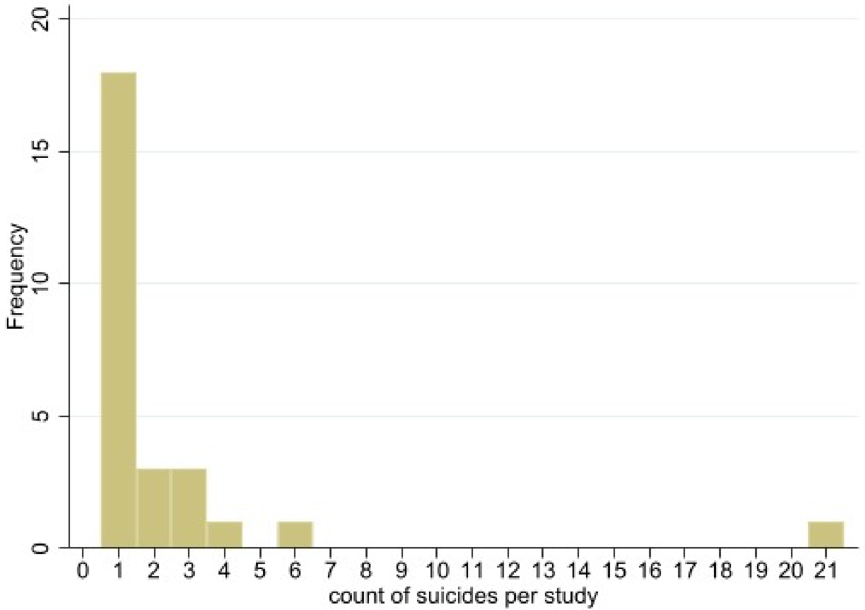
Case-study: Obesity Treatment
Risk of completed suicide after bariatric surgery: a systematic review

- bariatric surgery is one of the most effective treatments for morbid obesity, indicating a significant long-term weight loss
- while overall mortality decreases in patients who received bariatric surgery, risk of suicide is still an issue
- Peterhänsel et al. (2013) undertake a meta-analysis on completed suicide after bariatric surgery
- 27 studies are included in the analysis

Table 2 List of papers included for the estimate of the suicide rate in decreasing order of person-years

	Person-years	Weight	# of patients	# of women	# of suicides	Country
Adams	77,602	0.5397	9,949	8,556	21	USA
Marceau	10,388	0.0722	1,423	1,025	6	Canada
Marsk	8,877	0.0617	1,216	0	4	Sweden
Pories	8,316	0.0578	594	494	3	USA
Carelli	6,057	0.0421	2,909	1,989	1	USA
Busetto	4,598	0.0320	821	618	1	Italy
Smith 1995 (51)	3,882	0.0270	1,762	1,567	2	USA
Peeters	3,478	0.0242	966	744	1	Australia
Christou	2,599	0.0181	228	187	2	Canada
Günther	2,244	0.0156	98	82	1	Germany
Capella	2,237	0.0156	888	730	3	USA
Suter 2011 (31)	2,152	0.0150	379	282	3	Switzerland
Suter 2006 (32)	1,639	0.0114	311	269	1	Switzerland
Van de Weijgert	1,634	0.0114	200	174	1	Netherlands
Cadière	1,362	0.0095	470	392	1	Belgium
Mitchell	1,121	0.0078	85	72	1	USA
Himpens	1,066	0.0074	82	74	1	Belgium
Näslund 1994 (38)	799	0.0056	85	69	2	Sweden
Forsell	761	0.0053	326	248	1	Sweden
Powers 1997 (55)	747	0.0052	131	111	1	USA
Kral	477	0.0033	69	56	1	USA/Sweden
Näslund 1995 (35)	457	0.0032	142	84	1	Sweden
Powers 1992 (52)	395	0.0027	100	85	1	USA
Smith 2004 (50)	354	0.0025	779		1	USA
Nocca	228	0.0016	133	90	1	France
Svenheden	166	0.0012	91	72	1	Sweden
Pekkarinen	146	0.0010	27	19	1	Finland

The column entitled 'weight' is the fraction of the total number of person-years and is used in the analysis for comparing the estimated suicide rate for patients after a bariatric operation with the rate for an equivalent general population.


Frequency
0
5
10
15
20
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
count of suicides per study

Case-study: Obesity Treatment
Risk of completed suicide after bariatric surgery: a systematic review

- selection bias issue: only studies *with* completed suicide are included
- Peterhänsel et al. (2013):

  > *The most crucial point in the analysis was the proper treatment of the selection bias because of the method of finding papers.*

- hence, suicide rate will be *overestimated* (potentially substantially)

conventional meta-analysis

- in a nutshell, the conventional approach for a meta-analytic analysis (Cooper and Hedges 1994, Egger *et al.* 1995, Stangl and Berry 2000, Borenstein *et al.* 2009:311) proceed as follows:
- let X_i denote the observed count of suicides in study i and $E(X_i) = \mu_i$ its corresponding expected value
- also, let P_i denote the person-years in study i
- Then, in meta-analysis a summary measure as a weighted average of the study-specific rates on log-scale is used:

$$\sum_{i=1}^{n} w_i \log(X_i/P_i) / \sum_{i=1}^{n} w_i$$

where w_i is a proxy estimate of the inverse variance, here $w_i = Y_i$ leading to

$$\sum_{i=1}^{n} Y_i \log(X_i/P_i) / \sum_{i=1}^{n} Y_i$$

conventional meta-analysis

- another approach (Barendregt *et al.* 2013) works on the rate scale
- an attractive choice for w_i in

$$\sum_{i=1}^{n} w_i(X_i/P_i)/\sum_{i=1}^{n} w_i$$

is

$$w_i = P_i$$

- this is in the Mantel-Haenszel philosophy weighting with the denominator (here the person-years) leading to

$$\hat{\lambda} = \sum_{i=1}^{n} X_i/\sum_{i=1}^{n} P_i$$

as a summary estimate of the overall rate λ

conventional meta-analysis

- a benefit of the Mantel-Haenszel approach here is that the variance of $\hat{\lambda}$ is easy to calculate:

$$Var(\hat{\lambda}) = Var(\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} P_i)$$

$$= \sum_{i=1}^{n} \lambda P_i / (\sum_{i=1}^{n} P_i)^2$$

- which is estimated as

$$\hat{\lambda} / \sum_{i=1}^{n} P_i$$

- using this technique we find an overall rate of 44.51 suicides per $100,000$ person years with a 95% CI of $33.60 - 55.42$

problem with the conventional approach

- any of these conventional approaches cope with zero-event studies missing
- hence we need to turn to other ideas

the idea of capture-recapture

- objective is to determine the size N of an elusive target population
- some mechanism (life trapping, register, surveillance system) identifies a unit repeatingly
- there is a count X informing about the number of identifications of each unit in the target population

sample

available: sample

$$X_1, X_2, ..., X_N$$

leading to

Table: Frequency distribution of count X of repeated identifications

x	0	1	2	3	4	...	population size
f_x	f_0	f_1	f_2	f_3	f_4	...	N

problem

if $X_i = 0$ unit is not observed leading to a reduced observable sample

$$X_1, X_2, ..., X_n$$

where – w.l.g. – we assume that

$$X_{n+1} = X_{n+2} = ... = X_N = 0$$

Table: Frequency distribution of count X of repeated identifications

x	0	1	2	3	4	...	observed size
f_x	-	f_1	f_2	f_3	f_4	...	n

hence

$$f_0 = N - n \text{ is unknown}$$

why does data set fit into the capture-recapture setting?

- target population: *studies* on bariatric surgery with or without completed suicide
- identifying mechanism: online web-search including databases PubMed (PM), Web of Knowledge (WK), PsychInfo (PI), ScienceDirect (SD) and Google Scholar (GS)
- X_i number of completed suicides in study i: can be viewed as the count of repeated identifications for study i

modelling

- to cope with missing zeros we need to involve modelling
- $p_x = P(X = x)$ for $x = 0, 1, 2, \cdots$ base model
- for example *Poisson* :

$$p_x = \exp(-\mu)\mu^x/x! = \exp(-\lambda P)(\lambda P)^x/x!$$

λ suicide rate, P person-time, $\mu = \lambda P$

Table: Frequency distribution of count X of repeated identifications

x	0	1	2	3	4	...	m
f_x	-	f_1	f_2	f_3	f_4	...	f_m
p_x	p_0	p_1	p_2	p_3	p_4	...	p_m

modelling

- need to incorporate study-specific person-times
- $p_{ix} = P(X_i = x|P_i)$ probab. for x events in study with person-time P_i
- for example *Poisson* :

$$p_{ix} = \exp(-\lambda P_i)(\lambda P_i)^x / x!$$

  λ suicide rate, P_i person-time in study i, $\mu = \lambda P$

- complete data likelihood

$$\prod_{i=1}^{n} \prod_{x=0}^{m} p_{ix}^{f_{ix}}$$

  where f_{ix} is the frequency of studies with person-time P_i and event count x

- in our case, for given P_i the frequency f_{ix} is zero except for one value of x where it is one

EM philosophy: E-step

f_{i0} is unknown and needs to be replaced by its expected value: $E-step$

there is a general solution for the E-step:

$$e_{i0} := E(f_{i0}|f_{i1}, \cdots, f_{in}; P_i) = N_i p_{i0}$$

where N_i is the population size of studies with person-time P_i

it follows that

$$e_{i0} = N_i p_{i0} = (n_i + e_i) p_{i0}$$

where $n_i = f_{i1} + \cdots + f_{in}$ ($= 1$ in our case)

it follows further that

$$e_{i0} = n_i \frac{p_{i0}}{1 - p_{i0}}$$

which *replaces* f_{i0} in the complete, unobserved likelihood leading to the complete, expected likelihood

EM philosophy: E-step

note the relationship to the $Horvitz - Thompson$ estimator:

$$\hat{N}_i = n_i + e_{i0} = n_i + n_i \frac{p_{i0}}{1 - p_{i0}} = \frac{n_i}{1 - p_{i0}}$$

and

$$\hat{N} = \sum_{i=1}^{n} \hat{N}_i = \sum_{i=1}^{n} \frac{n_i}{1 - p_{i0}}$$

in the case study we have that $n_i = 1$ for $i = 1, \cdots, n$

the E-step provides as $by - product$ the item we are most interested in: the count of studies with no suicides, alternatively, the total number of studies

EM philosophy: M-step

we need to maximize the *complete*, *expected* data likelihood

$$\prod_{i=1}^{n}\prod_{x=1}^{m} p_{ix}^{f_{ix}} p_{i0}^{e_{i0}}$$

the solution will *depend* on the model used: in the *Poisson* case the complete data log-likelihood is

$$\sum_{i=1}^{n}\sum_{x=1}^{m} f_{ix}[-\mu_i + x\log\mu_i] - e_{i0}\mu_i$$

with $\mu_i = \lambda P_i$ which is maximized for

$$\hat{\lambda} = \frac{\sum_{i=1}^{n}\sum_{x=1}^{m} x\ f_{ix}}{\sum_{i=1}^{n}(\sum_{x=1}^{m} P_i f_{ix} + P_i e_{i0})}$$

EM philosophy

now, the EM algorithm toggles between E- and M-step until convergence

$$\text{E-step} \longleftrightarrow \text{M-step}$$

```
start rate MH: 0.0004451183

step:  1 rate:  0.000353999 size:  121.9951
step:  2 rate:  0.000329974 size:  129.6188
step:  3 rate:  0.000321995 size:  132.4051
step:  4 rate:  0.000319157 size:  133.4304
step:  5 rate:  0.000318122 size:  133.8086
...
step:  14 rate:  0.0003175201 size:  134.03
step:  15 rate:  0.0003175201 size:  134.03
```

Table 2 List of papers included for the estimate of the suicide rate in decreasing order of person-years

	Person-years	Weight	# of patients	# of women	# of suicides	Country
Adams	77,602	0.5397	9,949	8,556	21	USA
Marceau	10,388	0.0722	1,423	1,025	6	Canada
Marsk	8,877	0.0617	1,216	0	4	Sweden
Pories	8,316	0.0578	594	494	3	USA
Carelli	6,057	0.0421	2,909	1,989	1	USA
Busetto	4,598	0.0320	821	618	1	Italy
Smith 1995 (51)	3,882	0.0270	1,762	1,567	2	USA
Peeters	3,478	0.0242	966	744	1	Australia
Christou	2,599	0.0181	228	187	2	Canada
Günther	2,244	0.0156	98	82	1	Germany
Capella	2,237	0.0156	888	730	3	USA
Suter 2011 (31)	2,152	0.0150	379	282	3	Switzerland
Suter 2006 (32)	1,639	0.0114	311	269	1	Switzerland
Van de Weijgert	1,634	0.0114	200	174	1	Netherlands
Cadière	1,362	0.0095	470	392	1	Belgium
Mitchell	1,121	0.0078	85	72	1	USA
Himpens	1,066	0.0074	82	74	1	Belgium
Näslund 1994 (38)	799	0.0056	85	69	2	Sweden
Forsell	761	0.0053	326	248	1	Sweden
Powers 1997 (55)	747	0.0052	131	111	1	USA
Kral	477	0.0033	69	56	1	USA/Sweden
Näslund 1995 (35)	457	0.0032	142	84	1	Sweden
Powers 1992 (52)	395	0.0027	100	85	1	USA
Smith 2004 (50)	354	0.0025	779		1	USA
Nocca	228	0.0016	133	90	1	France
Svenheden	166	0.0012	91	72	1	Sweden
Pekkarinen	146	0.0010	27	19	1	Finland

The column entitled 'weight' is the fraction of the total number of person-years and is used in the analysis for comparing the estimated suicide rate for patients after a bariatric operation with the rate for an equivalent general population.

EM philosophy: full set of covariates

here an illustration in the Poisson case

$$p_{ix} = P(X_i = x|\beta; \mathbf{z_i}) = \exp(-\mu_i)\mu_i^x/x!$$

and

$$\log \mu_i = \beta^T \mathbf{z_i}$$

if there are *only* person-times

$$\log \mu_i = \log \lambda + \log P_i$$

EM philosophy

complete data likelihood – *with* covariates

$$\prod_{i=1}^{n} \prod_{x=0}^{m} p_{ix}^{f_{ix}}$$

where

- $p_{ix} = P(X_i = x | \beta; \mathbf{z_i})$
- $\mathbf{z_i}$ represents the i-th covariate combination for $i = 1, \cdots, n$
- f_{ix} is the frequency of observed counts equal to x for the i-th covariate combination
- f_{i0} remains unknown

E-step

we have

$$e_{i0} = n_i \frac{p_{i0}}{1 - p_{i0}}$$

with $p_{i0} = P(X_i = 0|\beta; \mathbf{z_i})$

M-step

to maximize

$$\prod_{i=1}^{n} \prod_{x=1}^{m} p_{ix}^{f_{ix}} p_{i0}^{e_{i0}}$$

this is model dependent; in the Poisson case with log-link

$$p_{ix} = P(X_i = x|\beta; \mathbf{z_i}) = \exp(-\mu_i)\mu_i^x / x!,$$

with $\log \mu_i = \beta^T \mathbf{z_i}$

M-step for the Poisson case with only person-times

$$p_{ij} = P(X_i = j|\beta; \mathbf{z_i}) = \exp(-\mu_i)\mu_i^j/j!$$

and

$$\mu_i = \exp(\eta + \underbrace{\log P_i}_{\text{log-person-times become offset}})$$

so, here simply

$$\mu_i = \exp(\beta^T \mathbf{z_i}) = \exp(\eta + \log P_i)$$

where η is the log-rate

alternatives to the EM philosophy

- use the observed, zero-truncated likelihood directly:

$$\prod_{i=1}^{n} \prod_{x=1}^{m} \left(\frac{p_{ix}}{1 - p_{i0}}\right)^{f_{ix}}$$

  where $p_{ix} = P(X_i = x | \beta; \mathbf{z_i})$ as before
- depends on the chosen model (Poisson, geometric, binomial, negative-binomial,...)
- use favorite algorithm such as NR, FS, or GN
- retrieve effect estimate $\hat{\beta}$

population size estimation with Horvitz-Thompson

$Horvitz - Thompson\ estimator$

$$\hat{N} = \sum_{i=1}^{N} I_i / w_i$$

where

- I_i is an indicator if the i-th study of the population of target studies is observed
- $w_i = P(I_i = 1) = 1 - P(I_i = 0) = 1 - p_{i0} = 1 - P(X_i = 0|\hat{\beta}; \mathbf{z_i})$
- under Poisson: $w_i = 1 - \exp(-\mu_i)$ and $\hat{\mu}_i = \exp(\hat{\beta}^T \mathbf{z_i})$

so that

$$\hat{N} = \sum_{i=1}^{n} 1/[1 - \exp(\hat{\beta}^T \mathbf{z_i})]$$

study population size estimation

so, in case we have use only person-times as offset

$$\hat{N} = \sum_{i=1}^{n} 1/[1 - \exp(-\exp(\hat{\eta} + \log PT_i))]$$

for the data

$$\hat{N} = \sum_{i=1}^{n} 1/[1 - \exp(\exp(\hat{\eta} + \log PT_i)] = 134$$

total studies with and *without* completed suicide after bariatric surggery

Table 2 List of papers included for the estimate of the suicide rate in decreasing order of person-years

	Person-years	Weight	# of patients	# of women	# of suicides	Country
Adams	77,602	0.5397	9,949	8,556	21	USA
Marceau	10,388	0.0722	1,423	1,025	6	Canada
Marsk	8,877	0.0617	1,216	0	4	Sweden
Pories	8,316	0.0578	594	494	3	USA
Carelli	6,057	0.0421	2,909	1,989	1	USA
Busetto	4,598	0.0320	821	618	1	Italy
Smith 1995 (51)	3,882	0.0270	1,762	1,567	2	USA
Peeters	3,478	0.0242	966	744	1	Australia
Christou	2,599	0.0181	228	187	2	Canada
Günther	2,244	0.0156	98	82	1	Germany
Capella	2,237	0.0156	888	730	3	USA
Suter 2011 (31)	2,152	0.0150	379	282	3	Switzerland
Suter 2006 (32)	1,639	0.0114	311	269	1	Switzerland
Van de Weijgert	1,634	0.0114	200	174	1	Netherlands
Cadière	1,362	0.0095	470	392	1	Belgium
Mitchell	1,121	0.0078	85	72	1	USA
Himpens	1,066	0.0074	82	74	1	Belgium
Näslund 1994 (38)	799	0.0056	85	69	2	Sweden
Forsell	761	0.0053	326	248	1	Sweden
Powers 1997 (55)	747	0.0052	131	111	1	USA
Kral	477	0.0033	69	56	1	USA/Sweden
Näslund 1995 (35)	457	0.0032	142	84	1	Sweden
Powers 1992 (52)	395	0.0027	100	85	1	USA
Smith 2004 (50)	354	0.0025	779		1	USA
Nocca	228	0.0016	133	90	1	France
Svenheden	166	0.0012	91	72	1	Sweden
Pekkarinen	146	0.0010	27	19	1	Finland

The column entitled 'weight' is the fraction of the total number of person-years and is used in the analysis for comparing the estimated suicide rate for patients after a bariatric operation with the rate for an equivalent general population.

practical modelling

Table: Linear predictors considered

Linear predictor	Proportion of women	Country of origin	Interaction	log-person-time as offset
0	No	No	No	No
1	No	No	No	Yes
2	Yes	No	No	Yes
3	No	Yes	No	Yes
4	Yes	Yes	No	Yes
5	Yes	Yes	Yes	Yes

Table: Values of the maximised log-likelihood, number of parameters, and BIC statistic s for models under consideration.

Distribution	LP	Maximised log-likelihood	Number of parameters	BIC
Poisson	5	-22.7	4	58.6
	4	-23.0	3	55.9
	3	-23.0	2	52.6
	2	-23.4	2	53.4
	1	-23.7	1	**50.7**
	0	-68.7	1	139.9
Negative-binomial	5	-22.7	5	61.9
	4	-23.0	4	59.2
	3	-23.0	3	55.9
	2	-23.4	3	56.7
	1	-23.7	2	54.0
	0	-38.7	2	84.0

uncertainty assessment with the bootstrap

- in principle, we have a population of size N
- for each element i we have an indicator I_i telling us if element i has been sampled or not

$$I_i = \begin{cases} 1, & \text{if sampled} \\ 0, & \text{otherwise} \end{cases}$$

  where $i = 1, ..., N$

- the classical nonparametric bootstrap would then consider random samples with replacement from $I_1, ..., I_N$
- problem is that we have *only* observed n out of N
- using the observed sample $I_1, ..., I_n$ for the bootstrap would *underestimate* the variability of $\hat{N}$
- the idea is to impute N using $\hat{N}$

uncertainty assessment with the bootstrap

Horvitz − Thompson estimator

$$\hat{N} = \sum_{i=1}^{N} I_i / \hat{w}_i$$

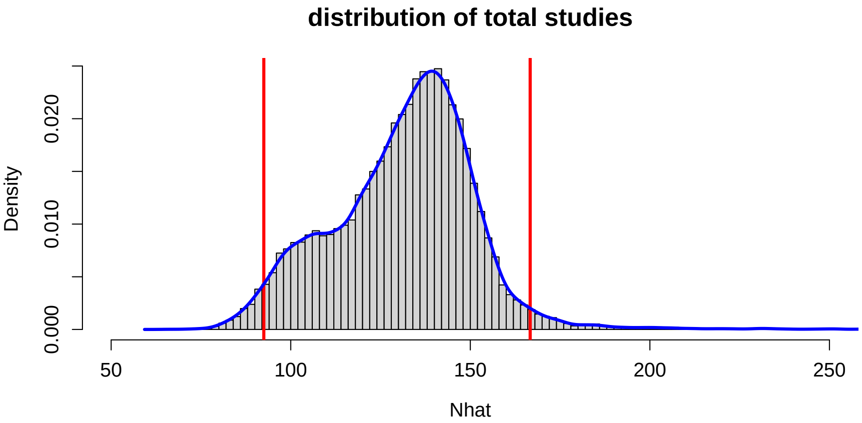
where

- $\hat{w}_i = \hat{P}(I_i = 1) = 1 - \hat{P}(I_i = 0)$
- under Poisson: $\hat{w}_i = 1 - \exp(-\hat{\mu}_i)$ and $\hat{\mu}_i = \exp(\hat{\beta}^T \mathbf{z_i})$
- or $\hat{N} = \sum_{i=1}^{n} 1/[1 - \exp(-\exp(\hat{\beta}^T \mathbf{z_i})]$
- this gives our imputed sample $I_1, ... I_n, ... I_{\hat{N}}$
- note that $I_{n+1}, ... I_{\hat{N}}$ are all zero ($\hat{N}$ needs to be rounded)

uncertainty assessment with the bootstrap

finally

- we can consider bootstrap samples $I_1^*, \ldots I_{\hat{N}}^*$
- note that there is now variability in the observed sample size n
- as all elements in the bootstrap sample with zero counts are truncated, it does not matter that we have *no* covariate information on the truncated counts
- using the zero-truncated bootstrap sample we estimate $\hat{N}^*$
- this process is repeated B times ($B = 25,000$ for example)



- *median* $= 133$ studies on bariatic surgery with or without completed suicide
- 95% *percentile confidence interval*: $93 - 167$ (red vertical bars)

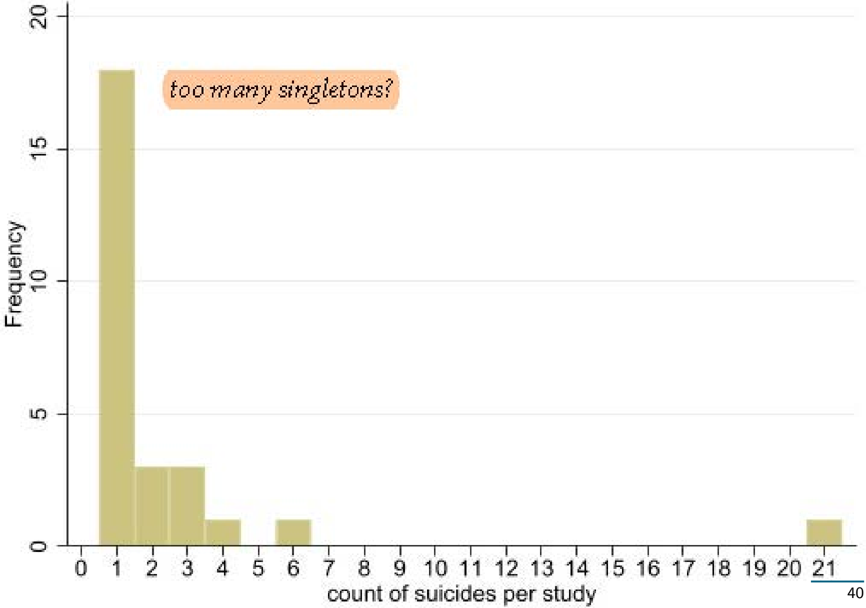
uncertainty assessment with the bootstrap

- in a similar way a 95% percentile confidence interval for the suicide rate is computed
- $24.84 - 49.39$ per $100,000$ person years
- with median rate of 31.86 per $100,000$ person years
- for comparison: the unadjusted rate is 44.51 per $100,000$ person years

acknowledgments

- joint work with Layna Dennett and Antony Overstall (University of Southampton)
- a paper version is available at:
- Layna Charlie Dennett, Antony Overstall, Dankmar Böhning (2023): Zero-Truncated Modelling Meta-Analysis for When Studies with No Events Are Systematically Excluded: Estimating Completed Suicide After Bariatric Surgery. https://arxiv.org/abs/2305.01277

further issues: one-inflation



further issues: one-inflation

Figure: The Guardian 30 Dec 2016: "Thousands of drink-drivers offend again"

drink-driving in Britain

- drink-driving (DD) relates to driving (or attempting to drive) while being above the legal alcohol limit
- according to the Guardian (30/12/16): 219,000 motorist were caught once, 8,068 twice, etc. (see Table below)

Table: Frequency distribution of the count (per person) of DVLA reported drink-driving (DD) in the UK between 2011 and 2015 (figures are based on DR10 endorsements)

count of DD	f_0	f_1	f_2	f_3	f_4	f_5	f_6	n
frequency		219,008	8,068	449	46	5	2	227,578

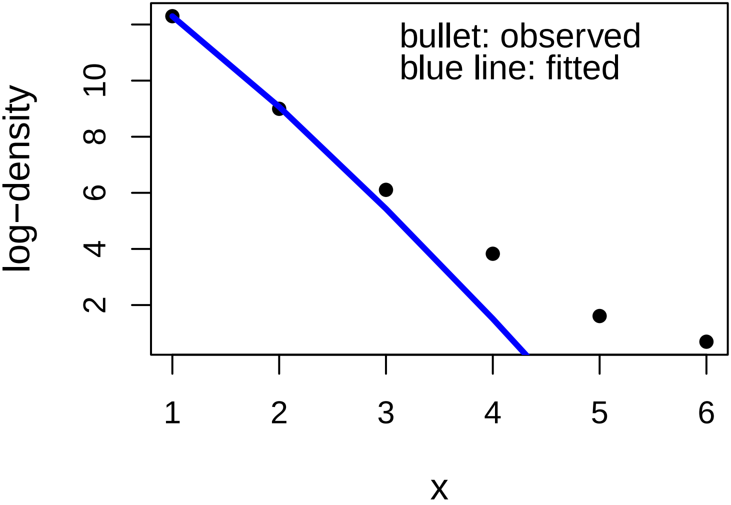


Figure: One-inflation distorts the Poisson fit

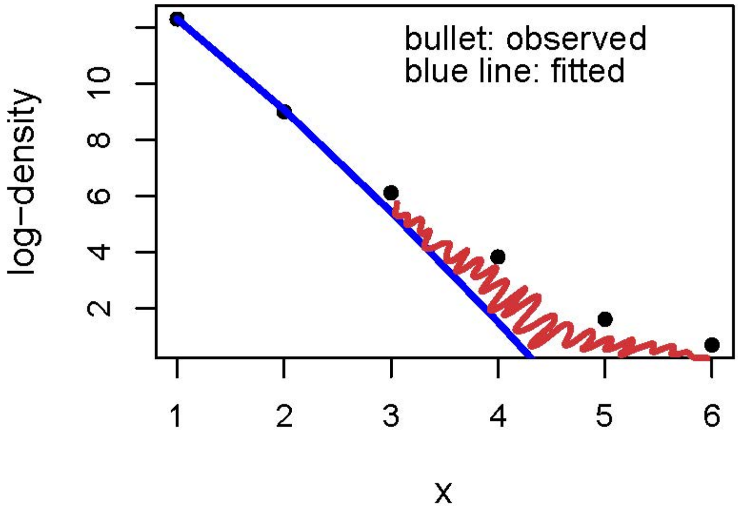


Figure: One-inflation distorts the Poisson fit

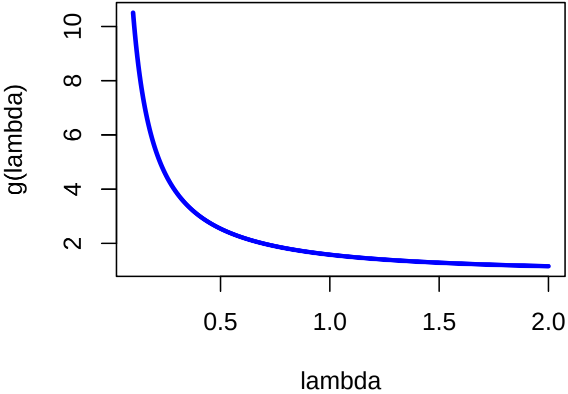
a synthetic example

- 500 counts sampled from $Po(1)$
- 500 extra-counts of 1 so that $N = 1,000$
- $\hat{\lambda} = 0.4091$ and

$$\text{HTE} = \frac{n}{1 - \exp(-\hat{\lambda})} = \frac{824}{1 - \exp(-0.4091)} = 2454$$

Table: one-inflated Poisson data

f_0	f_1	f_2	f_3	f_{4+}	n
176	690	95	32	7	824



- one-inflation leads to $\hat{\lambda} << \lambda$
- Horvitz-Thompson estimator $n \frac{1}{1-\exp(-\hat{\lambda})} >> N$
- as $g(\lambda) = \frac{1}{1-\exp(-\lambda)}$ strictly decreasing

two processes

- do not know the size: $zero - truncation$
- many counts of ones (singletons): $one - inflation$

this can be modelled as

$$(1-w)l_1(x) + \frac{w}{1-\ p(0;\theta)}p(x;\theta)$$

The Annals of Applied Statistics
2019, Vol. 13, No. 2, 1198–1211
https://doi.org/10.1214/18-AOAS1232

THE IDENTITY OF THE ZERO-TRUNCATED, ONE-INFLATED LIKELIHOOD AND THE ZERO-ONE-TRUNCATED LIKELIHOOD FOR GENERAL COUNT DENSITIES WITH AN APPLICATION TO DRINK-DRIVING IN BRITAIN

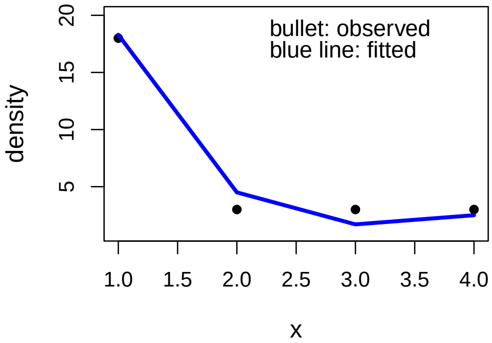
BY DANKMAR BÖHNING AND PETER G. M. VAN DER HEIJDEN

University of Southampton and University of Utrecht

GOF in the case study

Table: Frequency distribution for observed and fitted count of completed suicide under zero-truncated Poisson with offset for person-times; $\chi^2_{(2)} = 1.59$ and $p - \text{value} = 0.45$

count of completed suicide	0	1	2	3	4+
observed frequency f_x	-	18	3	3	3
fitted frequency $\hat{f}_x$	-	18.3	4.5	1.7	2.5



how to present fitted frequency for complex model

suppose a model (here for a Poisson with log-link) the has been fitted leading to

$$\hat{\mu}_i = \exp(\hat{\beta}^T \mathbf{z_i})$$

for unit i in the sample, then:

$$\hat{f}_x = \sum_{i=1}^{n} \exp(-\hat{\mu}_i)\hat{\mu}_i^x / x!$$

Statistical Methods in Medical Research
Volume 25, Issue 2, April 2016, Pages 902-916
© The Author(s) 2013, Article Reuse Guidelines
https://doi.org/10.1177/0962280212473386

SAGE journals

Article

The covariate-adjusted frequency plot

Heinz Holling[1], Walailuck Böhning[1], Dankmar Böhning[2], and Anton K Formann[3,†]

alternative: Bayes

- posterior $\propto$ likelihood $\times$ prior
- in our case

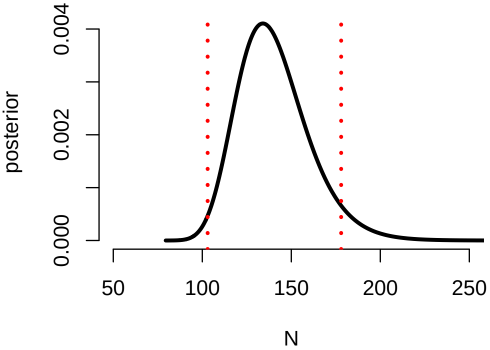
$$\pi(\lambda|x_1,\cdots,x_n) \propto \underbrace{\prod_i \frac{\exp(-\lambda P_i)}{1-\exp(-\lambda P_i)}(\lambda P_i)^{x_i}}_{ZT-Poisson-likelihood} \times \underbrace{\pi(\lambda)}_{prior}$$

- or

$$\pi(\lambda|x_1,\cdots,x_n) = \frac{\prod_i \frac{(\lambda P_i)^{x_i}}{\exp(-\lambda P_i)-1} \times \pi(\lambda)}{\int_\lambda \prod_i \frac{(\lambda P_i)^{x_i}}{\exp(-\lambda P_i)-1} \times \pi(\lambda)\ d\lambda}$$

priors

- non-informative $\pi(\lambda) = 1$
- 95% CI: $23.14 - 43.20$ per $100,000$ person years
- posterior median 31.75 per $100,000$ person years
- more interesting are the population sizes
- 95% CI: $103 - 178$ with posterior median of 134 studies

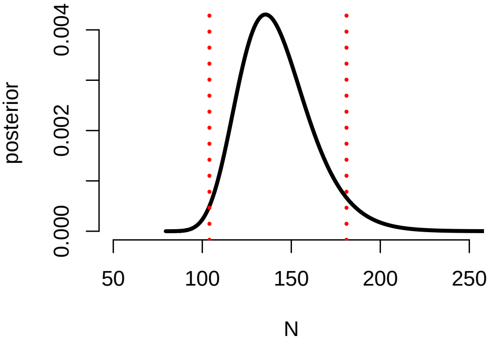


priors

- non-informative but proper $\log \lambda \sim N(0, 1000^2)$
- 95% CI: $23.47 - 43.17$ per $100,000$ person years
- posterior median 31.66 per $100,000$ person years
- more interesting the population sizes
- 95% CI: $103 - 175$ with posterior median of 134 studies
-
- for comparison with $\pi(\lambda) = 1$:
- 95% CI: $103 - \mathbf{178}$ with posterior median of 134 studies

priors

- Jeffreys invariance prior $\pi(\lambda) \propto \sqrt{\text{Fisher information}} = \sqrt{(\sum_i P_i)/\lambda}$
- 95% CI: $104 - 181$ with posterior median of 133 studies
- for comparison with $\pi(\lambda) = 1$:
- 95% CI: $103 - 178$ with posterior median of 134 studies



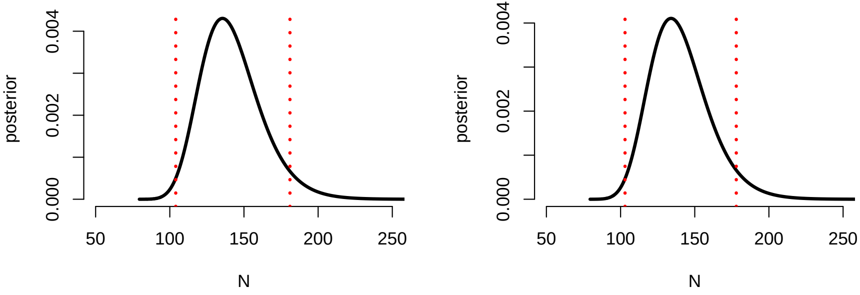


Figure: *left*: Jeffreys invariance prior *right*: non-informative improper prior

overview

Table: **all methods for estimating the total size of studies in a nutshell**

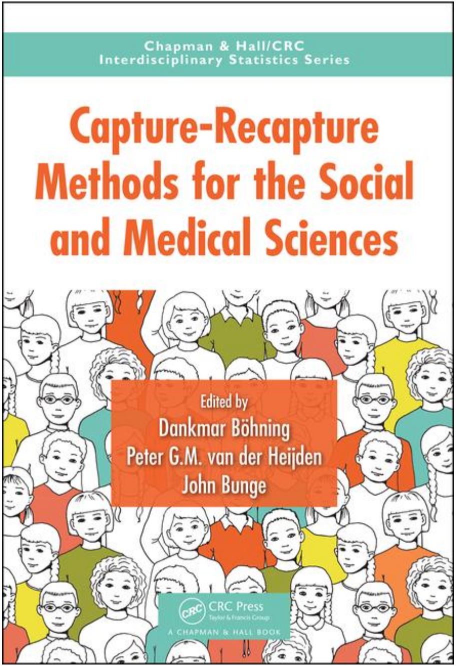
method	median	95% CI
MLE with bootstrap	133	93 – 167
Bayes prior:		
improper non-informative	134	103 – 178
log-normal	134	103 – 175
Jeffreys	133	104 – 181

Table: **a final point: model (likelihood) assessment is essential**

Distribution	LP	BIC	pop size
Poisson	5	58.6	125
	4	55.9	119
	3	52.6	118
	2	53.4	134
	1	50.7	134
	0	139.9	**31**

Table: recall: linear predictors considered

Linear predictor	Proportion of women	Country of origin	Interaction	log-person-time as offset
0	No	No	No	No
1	No	No	No	Yes
2	Yes	No	No	Yes
3	No	Yes	No	Yes
4	Yes	Yes	No	Yes
5	Yes	Yes	Yes	Yes


Chapman & Hall/CRC
Interdisciplinary Statistics Series
Capture-Recapture Methods for the Social and Medical Sciences
Edited by
Dankmar Böhning
Peter G.M. van der Heijden
John Bunge
CRC Press
A CHAPMAN & HALL BOOK