Partial Likelihood

Nancy Reid
University of Toronto

A celebration ...

CONFERENCE

SERIES - CENTRE FOR STATISTICAL METHODOLOGY

A celebration of 50 Years of the Cox model in memory of Sir David Cox

→

Thu 10 Nov 2022

LSHTM, Keppel Street, London, United Kingdom

50 years...

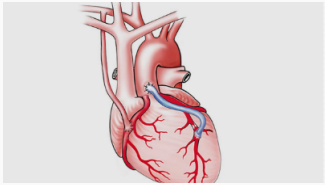
Revascularization of the Heart

Aortocoronary Bypass in Sixty-Three Patients

ROGER W. HALLIN, MD, Portland, Oregon
U. SCOTT PAGE, MD,* Portland, Oregon
JOHN C. BIGELOW, MD, Portland, Oregon
WILLIAM R. SWEETMAN, MD, Portland, Oregon

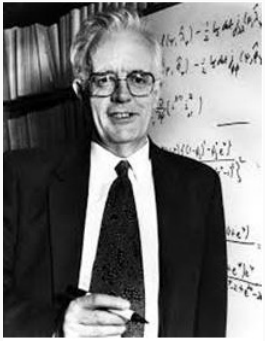
The ascending aorta-to-coronary artery bypass operation using autogenous saphenous vein is the

The primary indication for operation was (1) angina pectoris in fifty-six patients (six patients were

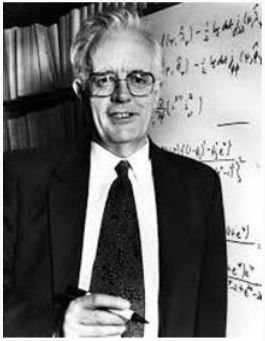


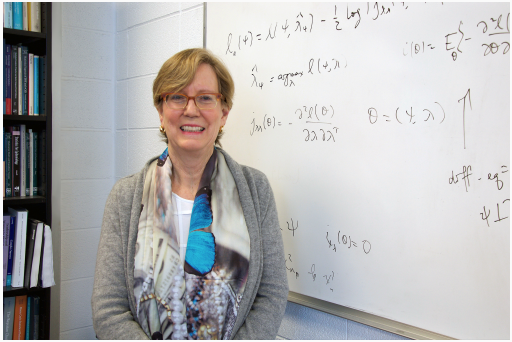
1974

A celebration ...



A celebration ...





Biometrika (1975), **62**, 2, *p.* 269
Printed in Great Britain

Partial likelihood

BY D. R. COX
Department of Mathematics, Imperial College, London

SUMMARY

A definition is given of partial likelihood generalizing the ideas of conditional and marginal likelihood. Applications include life tables and inference in stochastic processes. It is shown that the usual large-sample properties of maximum likelihood estimates and tests apply when partial likelihood is used.

Some key words: Asymptotic theory; Censoring; Conditional likelihood; Life table; Marginal likelihood; Regression; Stochastic process.

1. INTRODUCTION

Likelihood is central to much theoretical discussion of statistical inference, from what-ever viewpoint. In simple cases, the likelihood is just the joint density of the observed values

Biometrika (1975), **62**, 2, *p.* 269
Printed in Great Britain

Partial likelihood

BY D. R. COX
Department of Mathematics, Imperial College, London

SUMMARY

A definition is given of partial likelihood generalizing the ideas of conditional and marginal likelihood. Applications include life tables and inference in stochastic processes. It is shown that the usual large-sample properties of maximum likelihood estimates and tests apply when partial likelihood is used.

Some key words: Asymptotic theory; Censoring; Conditional likelihood; Life table; Marginal likelihood; Regression; Stochastic process.

1. INTRODUCTION

Likelihood is central to much theoretical discussion of statistical inference, from whatever viewpoint. In simple cases, the likelihood is just the joint density of the observed values

5. A Conditional Likelihood

Suppose then that $\lambda_0(t)$ is arbitrary. No information can be contributed about $\boldsymbol{\beta}$ by time intervals in which no failures occur because the component $\lambda_0(t)$ might conceivably be identically zero in such intervals. We therefore argue conditionally on the set $\{t_{(i)}\}$ of instants at which failures occur; in discrete time we shall condition

also on the observed multiplicities $\{m_{(i)}\}$. Once we require a method of analysis holding for all $\lambda_0(t)$, consideration of this conditional distribution seems inevitable.

For the particular failure at time $t_{(i)}$, conditionally on the risk set $\mathscr{R}(t_{(i)})$, the probability that the failure is on the individual as observed is

$$\exp\{\mathbf{z}_{(i)}\boldsymbol{\beta}\}\Big/\sum_{l\in\mathscr{R}(t_{(i)})}\exp\{\mathbf{z}_{(l)}\boldsymbol{\beta}\}. \tag{12}$$

Each failure contributes a factor of this nature and hence the required conditional log likelihood is

$$L(\boldsymbol{\beta})=\sum_{i=1}^{k}\mathbf{z}_{(i)}\boldsymbol{\beta}-\sum_{i=1}^{k}\log\Big[\sum_{l\in\mathscr{R}(t_{(i)})}\exp\{\mathbf{z}_{(l)}\boldsymbol{\beta}\}\Big]. \tag{13}$$

Drs JACK KALBFLEISCH and R. L. PRENTICE† (State University of New York at Buffalo): We would like to raise some questions concerning the conditional likelihood in Section 5 of this paper. Let us suppose a continuous hazard without censored observations. Expression (12) appears to be the conditional probability that individual i fails at $t_{(i)}$, given that a failure occurs at $t_{(i)}$ and given the risk at $R(t_{(i)})$. Thus if individuals 1, 2, 3 have associated covariate values z_1, z_2, z_3 and are observed to fail at t_1, t_2, t_3, with $t_1 < t_2 < t_3$, then expression (12) yields

(i) P (1 fails at t_1 | one failure at t_1 and $R(t_1) = \{1, 2, 3\}$)

Professor NORMAN BRESLOW (University of Washington): Like some of the other discussants I too was puzzled by the conditional likelihood of Section 2. I would like to suggest an alternative approach to the estimation of β and λ_0 which leads to equation (14) and also to a simpler estimate of the underlying survival distribution than is provided by equations (37) and (38). This approach is motivated in part by the discussion of Kalbfleisch and Prentice. However it differs from both their arguments and those of Cox in that

Statist. Sci. 1994

Drs JACK KALBFLEISCH and R. L. PRENTICE† (State University of New York at Buffalo): We would like to raise some questions concerning the conditional likelihood in Section 5 of this paper. Let us suppose a continuous hazard without censored observations. Expression (12) appears to be the conditional probability that individual i fails at $t_{(i)}$, given that a failure occurs at $t_{(i)}$ and given the risk at $R(t_{(i)})$. Thus if individuals 1, 2, 3 have associated covariate values z_1, z_2, z_3 and are observed to fail at t_1, t_2, t_3, with $t_1 < t_2 < t_3$, then expression (12) yields

(i) P (1 fails at t_1 | one failure at t_1 and $R(t_1) = \{1, 2, 3\}$)

Professor NORMAN BRESLOW (University of Washington): Like some of the other discussants I too was puzzled by the conditional likelihood of Section 2. I would like to suggest an alternative approach to the estimation of β and λ_0 which leads to equation (14) and also to a simpler estimate of the underlying survival distribution than is provided by equations (37) and (38). This approach is motivated in part by the discussion of Kalbfleisch and Prentice. However it differs from both their arguments and those of Cox in that

"...it really was *a* conditional likelihood; it was a form of conditional likelihood"

Statist. Sci. 1994

Nuisance parameters

- model $Y \sim f(y; \psi, \lambda), \quad \psi \in \mathbb{R}, \lambda \in \mathbb{R}^{d-1}, \quad \theta = (\psi, \lambda)$ $y = (y_1, \ldots, y_n)$
- or $Y \mid X \sim f(y \mid X; \psi, \lambda)$ $X_{n \times p}$, say

- log-likelihood function $\ell(\psi, \lambda; y) = \log f(y; \psi, \lambda) = \sum \log f(y_i; \psi, \lambda)$ if independent

- likelihood-based inference

Nuisance parameters

- model $Y \sim f(y; \psi, \lambda)$, $\quad \psi \in \mathbb{R}, \lambda \in \mathbb{R}^{d-1}, \quad \theta = (\psi, \lambda)$ $\quad y = (y_1, \ldots, y_n)$
- or $Y \mid X \sim f(y \mid X; \psi, \lambda)$ $\quad X_{n \times p}$, say

- log-likelihood function $\ell(\psi, \lambda; y) = \log f(y; \psi, \lambda) = \sum \log f(y_i; \psi, \lambda)$ if independent

- likelihood-based inference
  - profile log-likelihood maximize over λ

$$\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$$

  - maximum likelihood estimate $\quad j_p(\psi) = -\ell_p''(\psi)$

$$\hat{\psi} \dot{\sim} N\{\psi, j_p^{-1/2}(\psi)\}$$

  - likelihood ratio test

$$2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \dot{\sim} \chi^2_1$$

Nuisance parameters

- model $Y \sim f(y; \psi, \lambda)$, $\quad \psi \in \mathbb{R}, \lambda \in \mathbb{R}^{d-1}, \quad \theta = (\psi, \lambda)$ $\qquad y = (y_1, \ldots, y_n)$
- or $Y \mid X \sim f(y \mid X; \psi, \lambda)$ $\qquad X_{n \times p}$, say

- log-likelihood function $\ell(\psi, \lambda; y) = \log f(y; \psi, \lambda) = \sum \log f(y_i; \psi, \lambda)$ if independent

- likelihood-based inference
  - profile log-likelihood

    maximize over λ

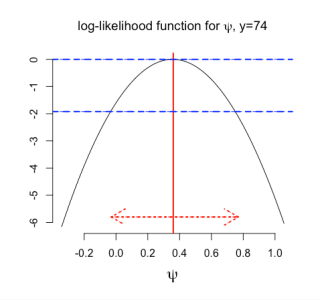
    $$\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$$

  - maximum likelihood estimate

    $$\hat{\psi} \dot{\sim} N\{\psi, j_p^{-1/2}(\psi)\}$$

  - likelihood ratio test

    $$2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \dot{\sim} \chi^2_1$$



Nuisance parameters

- model $Y \sim f(y; \psi, \lambda)$, $\quad \psi \in \mathbb{R}, \lambda \in \mathbb{R}^{d-1}$ $\qquad y = (y_1, \ldots, y_n)$
- or $Y \mid X \sim f(y \mid X; \psi, \lambda)$ $\qquad X_{n \times d}$, say

- log-likelihood function $\ell(\psi, \lambda; y) = \log f(y; \psi, \lambda) = \sum \log f(y_i; \psi, \lambda)$ if independent

- likelihood-based inference
  - profile log-likelihood
  - maximum likelihood estimate
  - likelihood ratio test

$\ell_p(\psi) = \ell(\psi$

$\hat{\psi} \dot{\sim} N\{\psi, j_p^{-1}$

$2\{\ell_p(\hat{\psi}) - \ell_p(\psi$

```
> summary(myglm)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.079      0.987   -3.12   0.0018 **
aged1         -0.292      0.754   -0.39   0.6988
stage1         1.373      0.784    1.75   0.0799 .
grade1         0.872      0.816    1.07   0.2850
xray1          1.801      0.810    2.22   0.0263 *
acid1          1.684      0.791    2.13   0.0334 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 40.710  on 22  degrees of freedom
Residual deviance: 18.069  on 17  degrees of freedom

> confint(myglm, variable-name)
Waiting for profiling to be done...
    2.5 %   97.5 %
0.266908 3.523458
```

... Nuisance parameters

- inference based on profile log-likelihood may be inaccurate — if p large, relative to n
- if the parameter of interest can be isolated in a conditional or marginal distribution, this makes inference much easier

$$f(y;\psi,\lambda) \propto f_m(t_1;\psi,\lambda)\, f_c(t_2 \mid t_1;\psi)$$
$$f(y;\psi,\lambda) \propto f_m(t_1;\psi)\, f_c(t_2 \mid t_1;\psi,\lambda)$$

... Nuisance parameters

- inference based on profile log-likelihood may be inaccurate — if p large, relative to n
- if the parameter of interest can be isolated in a conditional or marginal distribution, this makes inference much easier

$$f(y;\psi,\lambda) \propto f_m(t_1;\psi,\lambda)\, f_c(t_2 \mid t_1;\psi)$$
$$f(y;\psi,\lambda) \propto f_m(t_1;\psi)\, f_c(t_2 \mid t_1;\psi,\lambda)$$

- e.g. inference for common odds ratio in several 2×2 tables — conditional
- e.g. REML estimation for variance components — marginal

... Nuisance parameters

- inference based on profile log-likelihood may be inaccurate — if p large, relative to n

- if the parameter of interest can be isolated in a conditional or marginal distribution, this makes inference much easier

$$f(y;\psi,\lambda) \propto f_m(t_1;\psi,\lambda)\, f_c(t_2 \mid t_1;\psi)$$
$$f(y;\psi,\lambda) \propto f_m(t_1;\psi)\, f_c(t_2 \mid t_1;\psi,\lambda)$$

- e.g. inference for common odds ratio in several 2×2 tables — conditional
- e.g. REML estimation for variance components — marginal

- in the proportional hazards model, there are regression parameters, of interest, which can be specified in familiar forms
- as well as the failure and censoring processes, which operate in continuous time

Partial likelihood

Cox 1975

- data $(X_1, S_1, X_2, S_2, \ldots, X_j, S_j, \ldots, X_n, S_n)$

- successive densities conditional on the past: X_j, given $X_{(j-1)}, S_{(j-1)}$; S_j, given $X_{(j)}, S_{(j-1)}$
- likelihood function

joint density

$$L(\psi, \lambda; x, s) \propto \prod_{j=1}^{n} f(x_j \mid x_{(j-1)}, s_{(j-1)}; \psi, \lambda) \prod_{j=1}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi, \lambda)$$

Partial likelihood

Cox 1975

- data $(X_1, S_1, X_2, S_2, \ldots, X_j, S_j, \ldots, X_n, S_n)$

- successive densities conditional on the past: X_j, given $X_{(j-1)}, S_{(j-1)}$; S_j, given $X_{(j)}, S_{(j-1)}$
- likelihood function

joint density

$$L(\psi, \lambda; x, s) \propto \prod_{j=1}^{n} f(x_j \mid x_{(j-1)}, s_{(j-1)}; \psi, \lambda) \prod_{j=1}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi, \lambda)$$

- partial likelihood function

$$L_{\text{part}}(\psi, \lambda; x, s) \propto \prod_{j=2}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi, \lambda)$$

Partial likelihood

Cox 1975

- data $(X_1, S_1, X_2, S_2, \ldots, X_j, S_j, \ldots, X_n, S_n)$

- successive densities conditional on the past: X_j, given $X_{(j-1)}, S_{(j-1)}$; S_j, given $X_{(j)}, S_{(j-1)}$
- likelihood function

joint density

$$L(\psi, \lambda; x, s) \propto \prod_{j=1}^{n} f(x_j \mid x_{(j-1)}, s_{(j-1)}; \psi, \lambda) \prod_{j=1}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi, \lambda)$$

- partial likelihood function

$$L_{\text{part}}(\psi, \lambda; x, s) \propto \prod_{j=2}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi, \lambda)$$

- ideally, parameters of interest appear in L_{part} and not in the other bit
- e.g. regression parameters affecting relative hazards, parameters determining baseline hazards

Partial likelihood

Cox 1975

- data $(S_1, X_1, S_2, X_2, \ldots, X_j, S_j, \ldots, S_n, X_n)$

- successive densities conditional on the past: X_j, given $X_{(j-1)}, S_{(j-1)}$; S_j, given $X_{(j)}, S_{(j-1)}$
- likelihood function

joint density

$$L(\psi, \lambda; x, s) \propto \prod_{j=2}^{n} f(x_j \mid x_{(j-1)}, s_{(j-1)}; \quad \lambda) \prod_{j=2}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi \quad)$$

- partial likelihood function

$$L_{\text{part}}(\psi \quad ; x, s) \propto \prod_{j=2}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi \quad)$$

- ideally, parameters of interest appear in L_{part} and not in the other bit
- e.g. regression parameters affecting relative hazards, parameters determining baseline hazards

Partial likelihood

Cox 1975

- data $(S_1, X_1, S_2, X_2, \ldots, X_j, S_j, \ldots, S_n, X_n)$

- successive densities conditional on the past: X_j, given $X_{(j-1)}, S_{(j-1)}$; S_j, given $X_{(j)}, S_{(j-1)}$
- likelihood function

joint density

$$L(\psi, \lambda; x, s) \propto \prod_{j=2}^{n} f(x_j \mid x_{(j-1)}, s_{(j-1)}; \psi, \lambda) \prod_{j=2}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi \quad)$$

- partial likelihood function

$$L_{\text{part}}(\psi \quad ; x, s) \propto \prod_{j=2}^{n} f(s_j \mid x_{(j)}, s_{(j-1)}; \psi \quad)$$

- ideally, parameters of interest appear in L_{part} and not in the other bit
- e.g. regression parameters affecting relative hazards, parameters determining baseline hazards
- has the flavour of a conditional likelihood as above

but it's not

Partial likelihood and proportional hazards

Cox 1972; 1975

- S_j is jth individual observed to fail; $X_{(j)}$ is everything else — censoring, $\exists$ failure at t_j
- hazard for failure at t is $\lambda(t) = f(t)/\{1 - F(t)\}$ — density; survival
- proportional hazards has

$$\lambda(t; x) = \lambda_0(t) \exp(x^T \beta)$$

- data $t_1 < \cdots < t_n$ observed times — failure or censoring

Partial likelihood and proportional hazards

Cox 1972; 1975

- S_j is jth individual observed to fail; $X_{(j)}$ is everything else — censoring, $\exists$ failure at t_j
- hazard for failure at t is $\lambda(t) = f(t)/\{1 - F(t)\}$ — density; survival
- proportional hazards has

$$\lambda(t; x) = \lambda_0(t) \exp(x^T \beta)$$

- data $t_1 < \cdots < t_n$ observed times — failure or censoring

$$\begin{aligned} L(\beta, \lambda_0(\cdot); t, x) &= \prod_{j=1}^{n} \{\lambda(t_j; x_j)\{1 - F(t_j; x_j)\}^{\delta_j} \{1 - F(t_j; x_j)\}^{1-\delta_j} = \prod_{j=1}^{n} \{\lambda(t_j; x_j)\}^{\delta_j} \{1 - F(t_j; x_j)\} \\ &= \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta) \Lambda_0(t_j)\} \end{aligned}$$

- S_j is jth individual observed to fail; $X_{(j)}$ is everything else — censoring, $\exists$ failure at t_j
- hazard for failure at t is $\lambda(t) = f(t)/\{1 - F(t)\}$ — density; survival
- proportional hazards has

$$\lambda(t; x) = \lambda_0(t) \exp(x^T \beta)$$

- data $t_1 < \cdots < t_n$ observed times — failure or censoring

$$\begin{aligned} L(\beta, \lambda_0(\cdot); t, x) &= \prod_{j=1}^{n} \{\lambda(t_j; x_j)\{1 - F(t_j; x_j)\}^{\delta_j} \{1 - F(t_j; x_j)\}^{1-\delta_j} = \prod_{j=1}^{n} \{\lambda(t_j; x_j)\}^{\delta_j} \{1 - F(t_j; x_j)\} \\ &= \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta) \Lambda_0(t_j)\} \end{aligned}$$

$$L_{\text{part}}(\beta; t, x) = \prod_{\text{failures}} \frac{\exp(x_j^T \beta)}{\sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)}$$

jth individual fails, given there is a failure at t_j

x_j matches ordered times t_j

- full likelihood

$$L(\beta, \lambda_0(\cdot); t, x) = \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta)\Lambda_0(t_j)\}$$

- partial likelihood

$$L_{\text{part}}(\beta; t, x) = \prod_{failures} \frac{\exp(x_j^T \beta)}{\sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)}$$

- full likelihood

$$L(\beta, \lambda_0(\cdot); t, x) = \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta)\Lambda_0(t_j)\}$$

- partial likelihood

$$L_{\text{part}}(\beta; t, x) = \prod_{\textit{failures}} \frac{\exp(x_j^T \beta)}{\sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)}$$

- inference $\ell_{\text{part}}(\beta) = \log L_{\text{part}}(\beta)$

$$\ell'_{\text{part}}(\hat{\beta}) = 0; \quad -\ell''_{\text{part}}(\hat{\beta}) \doteq \{\widehat{var}(\hat{\beta})\}^{-1}$$

Cox 1972; 1975

- full likelihood

$$L(\beta, \lambda_0(\cdot); t, x) = \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta) \Lambda_0(t_j)\}$$

- partial likelihood

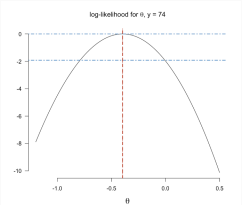
$$L_{\text{part}}(\beta; t, x) = \prod_{\text{failures}} \frac{\exp(x_j^T \beta)}{\sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)}$$

- inference $\ell_{\text{part}}(\beta) = \log L_{\text{part}}(\beta)$

$$\ell'_{\text{part}}(\hat{\beta}) = 0; \quad -\ell''_{\text{part}}(\hat{\beta}) \doteq \{\widehat{var}(\hat{\beta})\}^{-1}$$

$$\hat{\beta} - \beta \dot{\sim} N(0, \widehat{var}(\hat{\beta}))$$

$$2\{\ell_{\text{part}}(\hat{\beta}) - \ell_{\text{part}}(\beta_0)\} \dot{\sim} \chi^2_p$$



Pseudo-likelihood

Besag 1974; Renard et al. 2004

- modelling of spatial data
- analogue to auto-regression in time series
- condition on nearest neighbours of a given point

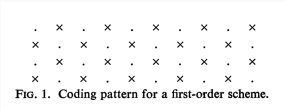


FIG. 1. Coding pattern for a first-order scheme.

$$L_{\text{pseudo}}(\theta) = \prod_{r=1}^{m} f(y_r \mid y_s;\ \text{site } s \text{ is a neighbour of site } r)$$

Pseudo-likelihood

Besag 1974; Renard et al. 2004

- modelling of spatial data
- analogue to auto-regression in time series
- condition on nearest neighbours of a given point

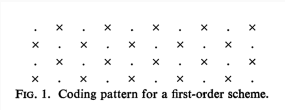


FIG. 1. Coding pattern for a first-order scheme.

$$L_{\text{pseudo}}(\theta) = \prod_{r=1}^{m} f(y_r \mid y_s;\ \text{site } s \text{ is a neighbour of site } r)$$

- multi-level/longitudinal binary data
- e.g. $\text{pr}(y_{ij} = 1 \mid b_i) = \Phi(x_{ij}^T\beta + z_{ij}^T b_i), j = 1, \ldots, q; i = 1, \ldots, m; \quad b_i \sim N(0, \Sigma_b)$

Pseudo-likelihood

Besag 1974; Renard et al. 2004

- modelling of spatial data
- analogue to auto-regression in time series
- condition on nearest neighbours of a given point

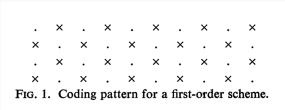


FIG. 1. Coding pattern for a first-order scheme.

$$L_{\text{pseudo}}(\theta) = \prod_{r=1}^{m} f(y_r \mid y_s;\ \text{site } s \text{ is a neighbour of site } r)$$

- multi-level/longitudinal binary data
- e.g. $\text{pr}(y_{ij} = 1 \mid b_i) = \Phi(x_{ij}^T\beta + z_{ij}^T b_i), j = 1, \ldots, q; i = 1, \ldots, m; \quad b_i \sim N(0, \Sigma_b)$
- likelihood function

$$L(\beta, \Sigma_b) = \prod_{i=1}^{n} \int \prod_{j=1}^{q} \{\Phi(x_{ij}^T\beta + z_{ij}^T b_i)\}^{y_{ij}} \{1 - \Phi(x_{ij}^T\beta + z_{ij}^T b_i)\}^{(1-y_{ij})} \phi(b_i; \Sigma_b) db_i$$

Pseudo-likelihood

Besag 1974; Renard et al. 2004

- modelling of spatial data
- analogue to auto-regression in time series
- condition on nearest neighbours of a given point

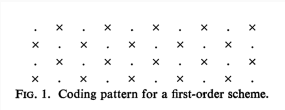


FIG. 1. Coding pattern for a first-order scheme.

$$L_{\text{pseudo}}(\theta) = \prod_{r=1}^{m} f(y_r \mid y_s;\ \text{site } s \text{ is a neighbour of site } r)$$

- multi-level/longitudinal binary data
- e.g. $\text{pr}(y_{ij} = 1 \mid b_i) = \Phi(x_{ij}^T\beta + z_{ij}^T b_i), j = 1, \ldots, q; i = 1, \ldots, m; \quad b_i \sim N(0, \Sigma_b)$
- likelihood function

$$L(\beta, \Sigma_b) = \prod_{i=1}^{n} \int \prod_{j=1}^{q} \{\Phi(x_{ij}^T\beta + z_{ij}^T b_i)\}^{y_{ij}} \{1 - \Phi(x_{ij}^T\beta + z_{ij}^T b_i)\}^{(1-y_{ij})} \phi(b_i; \Sigma_b) db_i$$

$$L_{\text{pseudo}}(\beta, \Sigma_b) = \prod_{i=1}^{n} \prod_{r<s} p_{11}^{y_{ir}y_{is}} p_{10}^{y_{ir}(1-y_{is})} p_{01}^{(1-y_{ir})y_{is}} p_{00}^{(1-y_{ir})(1-y_{is})}$$

each p_{ij} from bivariate normal probabilities

Pairwise likelihood

Cox & R 2004; Lindsay 1988

- random vector of responses $y_i = (y_{i1}, \ldots, y_{iq})$; joint density $f(y_i; \theta)$
- likelihood function $L(\theta; y) = \prod_{i=1}^{n} f(y_i; \theta)$
- pairwise likelihood function

$$L_{\text{pair}}(\theta; y) = \prod_{i=1}^{n} \prod_{s<t} f_2(y_{is}, y_{it}; \theta), \quad \text{or possibly} \quad \prod_{i=1}^{n} \prod_{s<t} \{f_2(y_{is}, y_{it}; \theta)\}^{w_i}$$

Pairwise likelihood

Cox & R 2004; Lindsay 1988

- random vector of responses $y_i = (y_{i1}, \ldots, y_{iq})$; joint density $f(y_i; \theta)$
- likelihood function $L(\theta; y) = \prod_{i=1}^{n} f(y_i; \theta)$
- pairwise likelihood function

$$L_{\text{pair}}(\theta; y) = \prod_{i=1}^{n} \prod_{s<t} f_2(y_{is}, y_{it}; \theta), \quad \text{or possibly} \quad \prod_{i=1}^{n} \prod_{s<t} \{f_2(y_{is}, y_{it}; \theta)\}^{w_i}$$

- or even

$q \to \infty$?

$$\ell_{\text{pair}}(\theta) = \sum_{i=1}^{n} \sum_{s<t} \log\{f_2(y_{is}, y_{it})\} - aq \sum_{i=1}^{n} \log\{f_1(y_{is})\}$$

- random vector of responses $y_i = (y_{i1}, \ldots, y_{iq})$; joint density $f(y_i; \theta)$
- likelihood function $L(\theta; y) = \prod_{i=1}^{n} f(y_i; \theta)$
- pairwise likelihood function

$$L_{\text{pair}}(\theta; y) = \prod_{i=1}^{n} \prod_{s<t} f_2(y_{is}, y_{it}; \theta), \quad \text{or possibly} \quad \prod_{i=1}^{n} \prod_{s<t} \{f_2(y_{is}, y_{it}; \theta)\}^{w_i}$$

- or even $q \to \infty$?

$$\ell_{\text{pair}}(\theta) = \sum_{i=1}^{n} \sum_{s<t} \log\{f_2(y_{is}, y_{it})\} - aq \sum_{i=1}^{n} \log\{f_1(y_{is})\}$$

- partial, pseudo-, pairwise, ... all examples of composite likelihood Lindsay 1988
- inference via maximum "likelihood" estimate and "likelihood" ratio test

with corrections for misspecification

PH partial likelihood is special

Andersen & Gill 1982

$$L_{\text{part}}(\beta; t, x) = \prod_{\text{failures}} \frac{\exp(x_j^T \beta)}{\sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)}, \quad \ell_{\text{part}}(\beta; t, x) = \sum_{\text{failures}} \{x_j^T \beta - \log \sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)\}$$

- score function $\ell'_{\text{part}}(\beta; t, x)$ is a martingale
- information function $-\ell''_{\text{part}}(\beta; t, x)$ estimates asymptotic variance of $\hat{\beta}_{\text{part}}$

PH partial likelihood is special

Andersen & Gill 1982

$$L_{\text{part}}(\beta; t, x) = \prod_{\text{failures}} \frac{\exp(x_j^T \beta)}{\sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)}, \quad \ell_{\text{part}}(\beta; t, x) = \sum_{\text{failures}} \{x_j^T \beta - \log \sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)\}$$

- score function $\ell'_{\text{part}}(\beta; t, x)$ is a martingale
- information function $-\ell''_{\text{part}}(\beta; t, x)$ estimates asymptotic variance of $\hat{\beta}_{\text{part}}$

$$\hat{\beta} - \beta \dot{\sim} N[0, \{-\ell''_{\text{part}}(\hat{\beta}_{\text{part}})\}^{-1}]$$

PH partial likelihood is special

Andersen & Gill 1982

$$L_{\text{part}}(\beta; t, x) = \prod_{\text{failures}} \frac{\exp(x_j^T \beta)}{\sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)}, \quad \ell_{\text{part}}(\beta; t, x) = \sum_{\text{failures}} \{x_j^T \beta - \log \sum_{k \in \mathcal{R}_j} \exp(x_k^T \beta)\}$$

- score function $\ell'_{\text{part}}(\beta; t, x)$ is a martingale
- information function $-\ell''_{\text{part}}(\beta; t, x)$ estimates asymptotic variance of $\hat{\beta}_{\text{part}}$

$$\hat{\beta} - \beta \dot{\sim} N[0, \{-\ell''_{\text{part}}(\hat{\beta}_{\text{part}})\}^{-1}]$$

- and weak convergence of the estimated cumulative hazard function

... PH partial likelihood is special

Murphy & van der Vaart 2001

$$L(\beta, \lambda_0(\cdot); t, x) = \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta)\Lambda_0(t_j)\}$$

- assume hazard function is an arbitrary constant between successive failure times

Breslow 1972

- a type of semi-parametric model
- we end up with n nuisance parameters, which is too many

Cox 1972

... PH partial likelihood is special

Murphy & van der Vaart 2001

$$L(\beta, \lambda_0(\cdot); t, x) = \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta) \Lambda_0(t_j)\}$$

- assume hazard function is an arbitrary constant between successive failure times

  Breslow 1972
- a type of semi-parametric model
- we end up with n nuisance parameters, which is too many — Cox 1972
- but, L_{part} is the profile likelihood, after maximizing over these n nuisance parameters

  M & vdV 2001; Davison 2003 §10.8

Murphy & van der Vaart 2001

$$L(\beta, \lambda_0(\cdot); t, x) = \prod_{j=1}^{n} \{\lambda_0(t_j) \exp(x_j^T \beta)\}^{\delta_j} \exp\{-\exp(x_j^T \beta) \Lambda_0(t_j)\}$$

- assume hazard function is an arbitrary constant between successive failure times

  Breslow 1972
- a type of semi-parametric model
- we end up with n nuisance parameters, which is too many

  Cox 1972
- but, L_{part} is the profile likelihood, after maximizing over these n nuisance parameters

  M & vdV 2001; Davison 2003 §10.8
- equivalently, $L\{\beta, \lambda_0(\cdot)\}$ is an empirical likelihood, with baseline hazard function a point mass at the observed failure times
- leads to proof that $\hat{\beta}_{\text{part}}$ is asymptotically normal and efficient
- likelihood ratio test asymptotically χ^2

  Murphy & vdV 2001; Sorensen 1983

Aside: Composite likelihood is less special

Lindsay 1988

- each component is a density
- e.g.

marginal or conditional or ...

$$L_{\text{pair}}(\theta; y) = \prod_{i=1}^{n} \prod_{s<t} f_2(y_{is}, y_{it}; \theta)$$

- estimating equation based on score function is unbiased for 0

$$\ell'_{\text{pair}}(\tilde{\theta}; y) = 0; \quad \mathrm{E}_\theta\{\ell'_{\text{pair}}(\theta; Y)\} = 0$$

- leads to proof that $\tilde{\theta}$ is consistent for θ

Aside: Composite likelihood is less special

Lindsay 1988

- each component is a density
- e.g.

marginal or conditional or ...

$$L_{\text{pair}}(\theta; y) = \prod_{i=1}^{n} \prod_{s<t} f_2(y_{is}, y_{it}; \theta)$$

- estimating equation based on score function is unbiased for 0

$$\ell'_{\text{pair}}(\tilde{\theta}; y) = 0; \quad \mathrm{E}_\theta\{\ell'_{\text{pair}}(\theta; Y)\} = 0$$

- leads to proof that $\tilde{\theta}$ is consistent for θ

- but $-\ell''_{\text{comp}}(\tilde{\theta})$ doesn't estimate a.var$(\tilde{\theta})$
- estimate is consistent but not asymptotically efficient
- correction needed for asymptotic variance and for likelihood ratio statistic

$$\mathrm{E}_\theta\{\ell'_{\text{comp}}(\tilde{\theta}; y)\}^2 \neq \mathrm{E}_\theta\{-\ell''_{\text{comp}}(\tilde{\theta})\}$$

The ANDROMEDA Trial

Hernandez et al., 2019

- randomized clinical trial to compare two treatments for septic shock
- 28-day mortality as response; analysed with Cox proportional hazards model
- estimated hazard ratio 0.75 [0.55, 1.02] after adjusting for confounders
- 2-sided p-value 0.06
- survival proportions (unadjusted for covariates) 34.9% vs 43.4% 8% reduction

The ANDROMEDA Trial

Hernandez et al., 2019

- randomized clinical trial to compare two treatments for septic shock
- 28-day mortality as response; analysed with Cox proportional hazards model
- estimated hazard ratio 0.75 [0.55, 1.02] after adjusting for confounders
- 2-sided p-value 0.06
- survival proportions (unadjusted for covariates) 34.9% vs 43.4% 8% reduction

- Discussion: " a peripheral perfusion-targeted resuscitation strategy did not result in a significantly lower 28-day mortality when compared with a lactate level-targeted strategy"

The ANDROMEDA Trial

Hernandez et al., 2019

- randomized clinical trial to compare two treatments for septic shock
- 28-day mortality as response; analysed with Cox proportional hazards model
- estimated hazard ratio 0.75 [0.55, 1.02] after adjusting for confounders
- 2-sided p-value 0.06
- survival proportions (unadjusted for covariates) 34.9% vs 43.4% 8% reduction

- Discussion: " a peripheral perfusion-targeted resuscitation strategy did not result in a significantly lower 28-day mortality when compared with a lactate level-targeted strategy"

- Abstract: "Among patients with septic shock, a resuscitation strategy targeting normalization of capillary refill time, compared with a strategy targeting serum lactate levels, did not reduce all-cause 28-day mortality."

Spiegelhalter, 2019

- Bayesian re-analysis to focus on posterior probability $\beta < 0$
- equivalently P(hazard ratio $< 1 \mid$ data)

ANDROMEDA, revisited

Zampieri et al 2020

- Bayesian re-analysis to focus on posterior probability $\beta < 0$
- equivalently P(hazard ratio < 1 | data)

- added random effect for center, used default priors for covariates, changed analysis to logistic regression
- with several different normal priors for the log odds-ratio
- the posterior probability that the odds-ratio is less than 1 — treatment is beneficial
- ranged from 0.94 to 0.99 — most pessimistic to most optimistic prior

ANDROMEDA, revisited

Zampieri et al 2020

- Bayesian re-analysis to focus on posterior probability $\beta < 0$
- equivalently P(hazard ratio < 1 | data)

- added random effect for center, used default priors for covariates, changed analysis to logistic regression
- with several different normal priors for the log odds-ratio
- the posterior probability that the odds-ratio is less than 1 — treatment is beneficial
- ranged from 0.94 to 0.99 — most pessimistic to most optimistic prior

ANDROMEDA, revisited

Zampieri et al 2020

- with several different normal priors for the log odds-ratio
- the posterior probability that the odds-ratio is less than 1 — treatment is beneficial
- ranged from 0.94 to 0.99 — most pessimistic to most optimistic prior

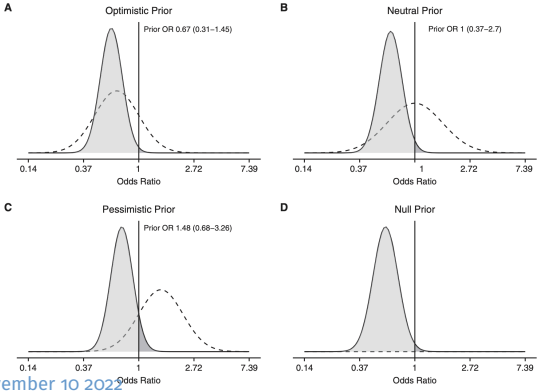


Figure 1. (A–D) Prior distributions for the odds ratio (OR) of the intervention (dashed lines). Posterior distributions of the ORs are shown by the solid lines. The light gray areas indicate the areas associated with benefit for peripheral perfusion-targeted resuscitation (i.e., OR < 1) and the dark gray areas the areas associated with harm (i.e., OR > 1). The text inside each frame reports the median and lower and upper 95% credible limits for the priors of the effect

see also van Zwet et al. 2021
used empirical prior
posterior prob 0.91

Table 1. Odds Ratio, 95% Credible Interval, Probability That the Odds Ratio Is below Given Thresholds, and Absolute Difference between Groups

Prior	28-d Outcome			90-d Outcome			Reason for Prior Use
	OR (95% Credible Interval)	Probability OR < 1 (Probability OR < 0.8)	Absolute Difference (95% Credible Interval)*	OR (95% Credible Interval)	Probability OR < 1 (Probability OR < 0.8)	Absolute Difference (95% Credible Interval)*	
Optimistic	0.61 (0.41 to 0.90)	99% (92%)	−9% (−17% to −1%)	0.69 (0.47 to 1.01)	97% (79%)	−7% (−16% to 2%)	Considers an OR of 0.67 for the intervention (slightly more conservative than the effect size ANDROMEDA-SHOCK was powered to detect), while considering that there is still a 15% probability that the intervention was harmful
Neutral	0.65 (0.43 to 0.96)	98% (85%)	−7% (−16% to 1%)	0.74 (0.50 to 1.08)	94% (66%)	−5% (−14% to 4%)	Has a mean OR of 1 (i.e., absence of effect) and 50% probability of benefit and 50% of harm from the intervention
Pessimistic	0.74 (0.50 to 1.09)	94% (66%)	−5% (−13% to 3%)	0.83 (0.57 to 1.21)	83% (42%)	−3% (−11% to 6%)	Opposite values of the optimistic prior; considers a very pessimistic scenario in which the intervention is harmful but still acknowledges a 15% chance that the intervention might be beneficial
Null	0.59 (0.38 to 0.92)	98% (91%)	−8% (−17% to 1%)	0.69 (0.45 to 1.07)	95% (74%)	−6% (−15% to 4%)	No prior information is considered

Definition of abbreviation: OR = odds ratio.
*Refers to a simple model adjusted only for study arm and not for all predictors.

Table 1. Odds Ratio, 95% Credible Interval, Probability That the Odds Ratio Is below Given Thresholds, and Absolute Difference between Groups

Prior	28-d Outcome			90-d Outcome			Reason for Prior Use
	OR (95% Credible Interval)	Probability OR < 1 (Probability OR < 0.8)	Absolute Difference (95% Credible Interval)*	OR (95% Credible Interval)	Probability OR < 1 (Probability OR < 0.8)	Absolute Difference (95% Credible Interval)*	
Optimistic	0.61 (0.41 to 0.90)	99% (92%)	−9% (−17% to −1%)	0.69 (0.47 to 1.01)	97% (79%)	−7% (−16% to 2%)	Considers an OR of 0.67 for the intervention (slightly more conservative than the effect size ANDROMEDA-SHOCK was powered to detect), while considering that there is still a 15% probability that the intervention was harmful
Neutral	0.65 (0.43 to 0.96)	98% (85%)	−7% (−16% to 1%)	0.74 (0.50 to 1.08)	94% (66%)	−5% (−14% to 4%)	Has a mean OR of 1 (i.e., absence of effect) and 50% probability of benefit and 50% of harm from the intervention
Pessimistic	0.74 (0.50 to 1.09)	94% (66%)	−5% (−13% to 3%)	0.83 (0.57 to 1.21)	83% (42%)	−3% (−11% to 6%)	Opposite values of the optimistic prior; considers a very pessimistic scenario in which the intervention is harmful but still acknowledges a 15% chance that the intervention might be beneficial
Null	0.59 (0.38 to 0.92)	98% (91%)	−8% (−17% to 1%)	0.69 (0.45 to 1.07)	95% (74%)	−6% (−15% to 4%)	No prior information is considered

Definition of abbreviation: OR = odds ratio.
*Refers to a simple model adjusted only for study arm and not for all predictors.

OR 0.59 (0.38,0.92); risk diff -8% (-17% to 1%)

- initial analysis: “Observed hazard ratio of 0.75 was not statistically significantly different from 1 at level 0.05”
- $p = 0.06$, 95% confidence interval $(0.55, 1.02)$

- translation: “new therapy has no benefit”

- initial analysis: “Observed hazard ratio of 0.75 was not statistically significantly different from 1 at level 0.05”
- $p = 0.06$, 95% confidence interval $(0.55, 1.02)$

- translation: “new therapy has no benefit”

- second analysis: “Posterior probability that odds ratio is less than one is 0.98”
- posterior credible interval $(0.38, 0.92)$ van Zwet et al. 2020

- translation: “new therapy is better” logistic regression

- initial analysis: “Observed hazard ratio of 0.75 was not statistically significantly different from 1 at level 0.05”
- $p = 0.06$, 95% confidence interval $(0.55, 1.02)$
- translation: “new therapy has no benefit”
- second analysis: “Posterior probability that odds ratio is less than one is 0.98”
- posterior credible interval $(0.38, 0.92)$ van Zwet et al. 2020
- translation: “new therapy is better” logistic regression
- is more study needed?

Research

JAMA | Original Investigation

Effect of Ivermectin vs Placebo on Time to Sustained Recovery in Outpatients With Mild to Moderate COVID-19

A Randomized Clinical Trial

Susanna Naggie, MD, MHS; David R. Boulware, MD, MPH; Christopher J. Lindsell, PhD; Thomas G. Stewart, PhD; Nina Gentile, MD; Sean Collins, MD, MSci; Matthew William McCarthy, MD; Dushyantha Jayaweera, MD; Mario Castro, MD, MPH; Mark Sulkowski, MD; Kathleen McTigue, MD, MPH, MS; Florence Thicklin; G. Michael Felker, MD, MHS; Adit A. Ginde, MD, MPH; Carolyn T. Bramante, MD, MPH; Alex J. Slandzicki, MD; Ahab Gabriel, MD; Nirav S. Shah, MD, MPH; Leslie A. Lenert, MD, MS; Sarah E. Dunsmore, PhD; Stacey J. Adam, PhD; Allison DeLong, BS; George Hanna, MD; April Remaly, BA; Rhonda Wilder, MS; Sybil Wilson, RN; Elizabeth Shenkman, PhD; Adrian F. Hernandez, MD, MHS; for the Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV-6) Study Group and Investigators

Key Points

Question Does ivermectin, 400 µg/kg, daily for 3 days, compared with placebo, shorten symptom duration among adult (≥30 years) outpatients in the US with symptomatic mild to moderate COVID-19?

Findings In this double-blinded, randomized, placebo-controlled platform trial conducted in the US during a period of Delta and Omicron variant predominance, and that included 1591 adult outpatients with COVID-19, the posterior probability of improvement in time to recovery in those treated with ivermectin vs placebo had a hazard ratio of 1.07, with a posterior probability of benefit of .91. This did not meet the prespecified threshold of posterior probability greater than .95.

Meaning These findings do not support the use of ivermectin in outpatients with mild to moderate COVID-19.

"Conclusions Among outpatients with mild to moderate COVID-19, treatment with ivermectin, compared with placebo, did not significantly improve time to recovery. "

Key Points

Question Does ivermectin, 400 µg/kg, daily for 3 days, compared with placebo, shorten symptom duration among adult (≥30 years) outpatients in the US with symptomatic mild to moderate COVID-19?

Findings In this double-blinded, randomized, placebo-controlled platform trial conducted in the US during a period of Delta and Omicron variant predominance, and that included 1591 adult outpatients with COVID-19, the posterior probability of improvement in time to recovery in those treated with ivermectin vs placebo had a hazard ratio of 1.07, with a posterior probability of benefit of .91. This did not meet the prespecified threshold of posterior probability greater than .95.

Meaning These findings do not support the use of ivermectin in outpatients with mild to moderate COVID-19.

"Conclusions Among outpatients with mild to moderate COVID-19, treatment with ivermectin, compared with placebo, did not significantly improve time to recovery. "

Thank you!