# The Cox model: introduction and history

Bianca L. De Stavola
*Centre for Statistical Methodology*
LSHTM

March 8, 2013

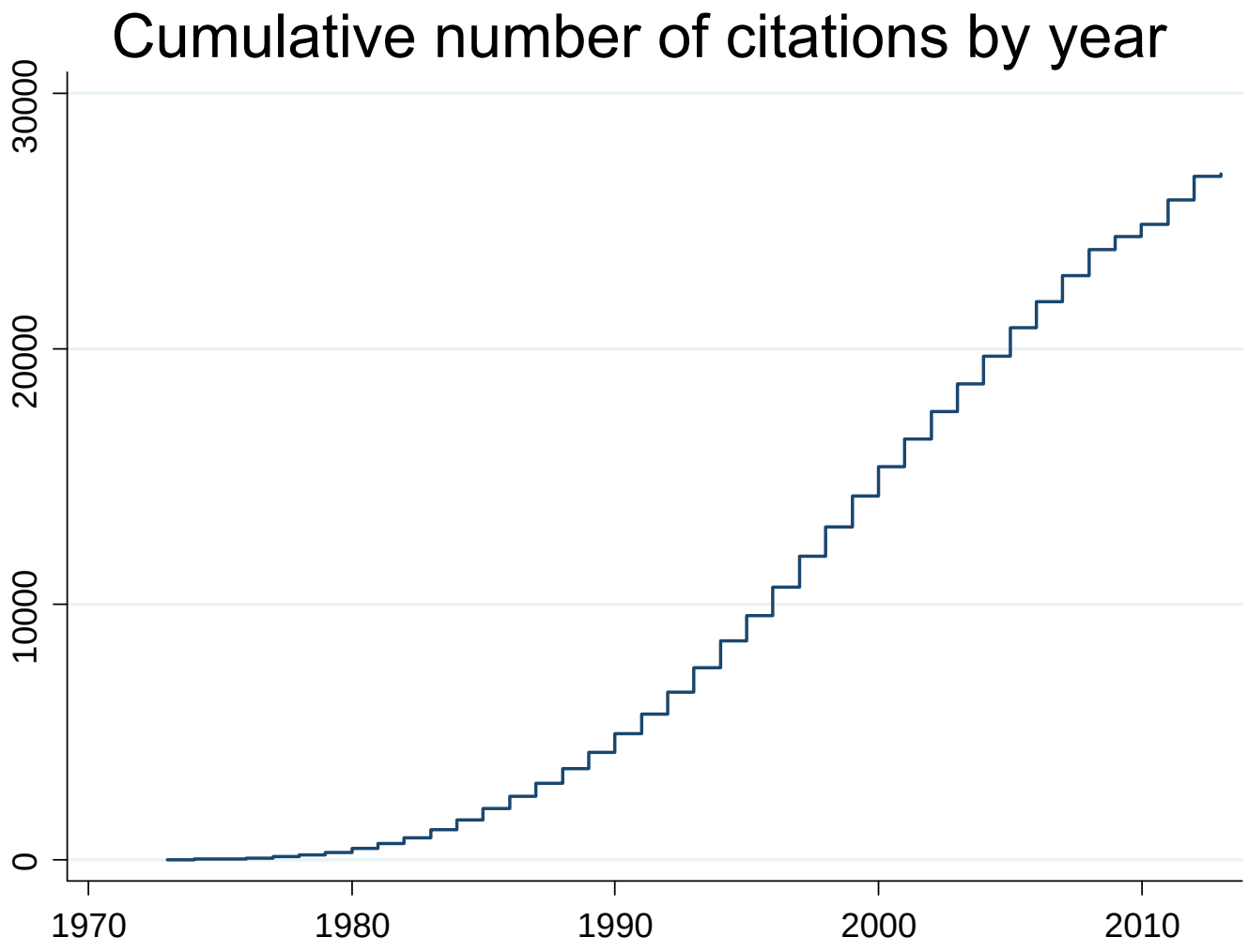
# Introduction

Today is a celebration of an incredibly influential paper:

- the most cited paper in the whole history of JRSS
- the third most cited paper in medical journals
- it has a total of nearly 30,000 citations (according to Web of Science)
- and this is still increasing

# Cumulative citations

## Web of Science



# Outline

# Outline

# Setting

We are concerned with studying:

- individuals at risk of experiencing a failure after time $T$
- measured from a relevant origin and according to a relevant measurement scale
- difficulty if some are not observed until failure occurs, *i.e.* are censored
- crucially, censoring must be independent of the failure process

# Definitions

If $T$ is a continuous positive random variable its probability distribution is equivalently specified by:

- the density function:

$$f(t) = \lim_{\Delta t \to 0^+} \frac{Pr\left(t \leq T < t + \Delta t\right)}{\Delta t}$$

- the survivor function:

$$S(t) = Pr\left(T \geq t\right)$$

- the hazard function:

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{Pr\left(t \leq T < t + \Delta t | t \leq T\right)}{\Delta t}$$

# Survivor and hazard function

By the product law of probability, $S(t)$ is related to $\lambda(t)$:

$$S(t) = \lim \prod_{k=0}^{r-1} \{1 - \lambda(\tau_k)(\tau_{k+1} - \tau_k)\}$$

where:

- the limit is for $(\tau_{k+1} - \tau_k) \to 0$
- $0 < \tau_1 < \tau_2 < \ldots < \tau_r = t$ the interval endpoints

# Survivor and hazard function

By the product law of probability, $S(t)$ is related to $\lambda(t)$:

$$S(t) = \lim \prod_{k=0}^{r-1} \{1 - \lambda(\tau_k)(\tau_{k+1} - \tau_k)\}$$

where:

- the limit is for $(\tau_{k+1} - \tau_k) \to 0$
- $0 < \tau_1 < \tau_2 < \ldots < \tau_r = t$ the interval endpoints

$S(t)$ is the product of the conditional survival probabilities for infinitesimal intervals up to $t$

# Survivor and hazard function

By the product law of probability, $S(t)$ is related to $\lambda(t)$:

$$S(t) = \lim \prod_{k=0}^{r-1} \{1 - \lambda(\tau_k)(\tau_{k+1} - \tau_k)\}$$

where:

- the limit is for $(\tau_{k+1} - \tau_k) \to 0$
- $0 < \tau_1 < \tau_2 < \ldots < \tau_r = t$ the interval endpoints

$S(t)$ is the product of the conditional survival probabilities for infinitesimal intervals up to $t$

construction of the likelihood depends on this

# Known at the time

Three main analytical approaches:

1. Non-parametric estimation of $S(t)$
2. Parametric estimation of $S(t)$
3. Comparison of survivor functions ("the two-sample problem")

# 1 – Non-parametric estimation of $S(t)$

- Actuarial (Life-Table) estimation:
  - long tradition in demography
  - assuming hazard function piecewise constant over pre-specified intervals $\{t_j, t_{j+1}\}$, $\hat{\lambda}_j =$ no. events / total follow-up time
  - survivor function estimated as the product of the conditional probabilities of surviving each interval: $\hat{S}(t_j) = \prod_{k<j} \left(1 - \hat{\lambda}_k\right)$

# 1 – Non-parametric estimation of $S(t)$

- Actuarial (Life-Table) estimation:
  - long tradition in demography
  - assuming hazard function piecewise constant over pre-specified intervals $\{t_j, t_{j+1}\}$, $\hat{\lambda}_j$ = no. events / total follow-up time
  - survivor function estimated as the product of the conditional probabilities of surviving each interval: $\hat{S}(t_j) = \prod_{k<j} \left(1 - \hat{\lambda}_k\right)$
- Product Limit estimation:
  - exactly the same but defined for vanishingly small intervals
  - hence $\hat{\lambda}_j$ = no. events / total no. persons at risk
  - derived by Kaplan & Meier (1958) as a non-parametric MLE of $S(t)$
- Although both derived from ML arguments, asym. properties not developed until later (Breslow and Crowley, 1974)

# 2 – Parametric estimation of $S(t)$

- Exponential and Weibull often used (simple formulæ for $S(t)$ and $\lambda(t)$)
- Approach attractive because of physical interpretation, *e.g.*
  - multi-hit carcinogenesis theories lead to Weibull models (Armitage and Doll, 1954, 1961)
- Estimation via ML mostly derived assuming fixed censoring time (*e.g.* Bartholomew 1963 for exponential, Pike 1966 for Weibull)
- Inclusion of explanatory variables rare and with no censoring

# 3 – Comparing survivor functions: the two-sample problem

Generalizations of the Savage–Wilcoxon rank test to settings with censored data:

- Mantel (1966):
  - test based on difference between observed and expected events at each failure time
  - expectations come from the hypergeometric distribution
  - results combined as in the Mantel–Haenszel test (1959)

# 3 – Comparing survivor functions: the two-sample problem

Generalizations of the Savage–Wilcoxon rank test to settings with censored data:

- Mantel (1966):
  - test based on difference between observed and expected events at each failure time
  - expectations come from the hypergeometric distribution
  - results combined as in the Mantel–Haenszel test (1959)
- Peto and Peto (1972):
  - different derivation of the same comparison
  - named log-rank test
  - Note: paper read $< 2$ months before "Regression models and life-tables"

# Summary of methods prevalent in 1972

- Statistical theory for non-parametric estimation of $S(t)$ not yet fully formalized
- Inference for parametric estimation of $S(t)$: complex even with simple censoring mechanisms
- Comparison of survivor curves dealt with via significance tests (and only possible for categorical variables)
- Extension of parametric models to include explanatory variables not generally available with censoring

# Outline

# Regression Models and Life-Tables

By D. R. Cox

*Imperial College, London*

[Read before the Royal Statistical Society, at a meeting organized by the Research Section, on Wednesday, March 8th, 1972, Mr M. J. R. Healy in the Chair]

Summary

The analysis of censored failure times is considered. It is assumed that on each individual are available values of one or more explanatory variables. The hazard function (age-specific failure rate) is taken to be a function of the explanatory variables and unknown regression coefficients multiplied by an arbitrary and unknown function of time. A conditional likelihood is obtained, leading to inferences about the unknown regression coefficients. Some generalizations are outlined.

Read here in the Goldsmiths Lecture Theatre

# Aims

- *"The present paper is largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables. . ."*
- *". . . and more generally to the incorporation of regression like arguments into life-table analysis"*
- it would be *"sensible to make a minimum of assumptions leading to a convenient analysis"*

# Aims

- *"The present paper is largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables. . . "*
- *". . . and more generally to the incorporation of regression like arguments into life-table analysis"*
- it would be *"sensible to make a minimum of assumptions leading to a convenient analysis"*

# Aims

- *"The present paper is largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables. . . "*
- *". . . and more generally to the incorporation of regression like arguments into life-table analysis"*
- it would be *"sensible to make a minimum of assumptions leading to a convenient analysis"*

# Outline

# The Cox model

- Let $\mathbf{z} = z_1, z_2, \ldots, z_p$ be explanatory variables of interest
- Proportional hazards (PH) model defined as

$$\lambda(t; \mathbf{z}) = \exp(\mathbf{z}\beta)\,\lambda_0(t)$$

  - $\beta$ vector of unknown parameters (of interest)
  - $\lambda_0(t)$ unknown arbitrary function (nuisance)
- $\lambda_0(t)$ describes the shape of the survival function
- $\exp(\mathbf{z}\beta)$ could be replaced by $h(\mathbf{z}, \beta)$

# The Cox model

- Let $\mathbf{z} = z_1, z_2, \ldots, z_p$ be explanatory variables of interest
- Proportional hazards (PH) model defined as

$$\lambda(t; \mathbf{z}) = \exp(\mathbf{z}\boldsymbol{\beta})\, \lambda_0(t)$$

  - $\boldsymbol{\beta}$ vector of unknown parameters (of interest)
  - $\lambda_0(t)$ unknown arbitrary function (nuisance)
- $\lambda_0(t)$ describes the shape of the survival function
- $\exp(\mathbf{z}\boldsymbol{\beta})$ could be replaced by $h(\mathbf{z}, \boldsymbol{\beta})$

- *explore the consequences of allowing $\lambda_0(t)$ to be arbitrary*
- *method to have sensible properties, whatever $\lambda_0(t)$*
- *plausible loss of information about $\boldsymbol{\beta}$ is usually slight*

# Estimation (1)

- Observations:
  - $n$ individuals, $k$ fail
  - independent censoring
  - failure times: $0 < t_{(1)} < t_{(2)} < \ldots < t_{(k)} < \infty$
  - $\Re(t)$ the set of individuals at risk at time $t$

# Estimation (1)

- Observations:
  - $n$ individuals, $k$ fail
  - independent censoring
  - failure times: $0 < t_{(1)} < t_{(2)} < \ldots < t_{(k)} < \infty$
  - $\Re(t)$ the set of individuals at risk at time $t$
- Originally estimation derived from a 'conditional likelihood'
- This is the product of factors, one per event time $t_{(i)}$:

$$\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\left\{\mathbf{z}_{\ell}\boldsymbol{\beta}\right\}}$$

# Estimation (1)

- Observations:
  - $n$ individuals, $k$ fail
  - independent censoring
  - failure times: $0 < t_{(1)} < t_{(2)} < \ldots < t_{(k)} < \infty$
  - $\Re(t)$ the set of individuals at risk at time $t$

- Originally estimation derived from a 'conditional likelihood'
- This is the product of factors, one per event time $t_{(i)}$:

$$\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\left\{\mathbf{z}_{\ell}\boldsymbol{\beta}\right\}}$$

conditional probabilities that the failure is on the individual as observed

# Estimation (2)

- 'Conditional log-likelihood' then is:

$$l(\beta) = \Sigma_{i=1}^{k} \mathbf{z}_{(i)}\beta - \Sigma_{i=1}^{k} \log \left[ \Sigma_{\ell \in \Re\left(t_{(i)}\right)} \exp\left\{\mathbf{z}_{\ell}\beta\right\} \right]$$

# Estimation (2)

- 'Conditional log-likelihood' then is:

$$l(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \mathbf{z}_{(i)}\boldsymbol{\beta} - \Sigma_{i=1}^{k} \log\left[\Sigma_{\ell \in \Re\left(t_{(i)}\right)} \exp\left\{\mathbf{z}_{\ell}\boldsymbol{\beta}\right\}\right]$$

- The score function:

$$U_{\xi}(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \left\{z_{\xi i} - A_{\xi i}(\boldsymbol{\beta})\right\}$$

where $A_{\xi i}(\boldsymbol{\beta})$ is the weighted average of $z_{\xi}$ in $\Re$

# Estimation (2)

- 'Conditional log-likelihood' then is:

$$l(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \mathbf{z}_{(i)}\boldsymbol{\beta} - \Sigma_{i=1}^{k} \log\left[\Sigma_{\ell \in \Re\left(t_{(i)}\right)} \exp\left\{\mathbf{z}_{\ell}\boldsymbol{\beta}\right\}\right]$$

- The score function:

$$U_{\xi}(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \left\{z_{\xi i} - A_{\xi i}(\boldsymbol{\beta})\right\}$$

where $A_{\xi i}(\boldsymbol{\beta})$ is the weighted average of $z_{\xi}$ in $\Re$

- The information matrix the sum of weighted covariance matrices

# Estimation (2)

- 'Conditional log-likelihood' then is:

$$l(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \mathbf{z}_{(i)}\boldsymbol{\beta} - \Sigma_{i=1}^{k} \log\left[\Sigma_{\ell \in \Re\left(t_{(i)}\right)} \exp\left\{\mathbf{z}_{\ell}\boldsymbol{\beta}\right\}\right]$$

- The score function:

$$U_{\xi}(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \left\{z_{\xi i} - A_{\xi i}(\boldsymbol{\beta})\right\}$$

where $A_{\xi i}(\boldsymbol{\beta})$ is the weighted average of $z_{\xi}$ in $\Re$

- The information matrix the sum of weighted covariance matrices
- $\hat{\boldsymbol{\beta}}$ found by iteration

# Estimation (2)

- 'Conditional log-likelihood' then is:

$$l(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \mathbf{z}_{(i)}\boldsymbol{\beta} - \Sigma_{i=1}^{k} \log\left[\Sigma_{\ell \in \Re\left(t_{(i)}\right)} \exp\left\{\mathbf{z}_{\ell}\boldsymbol{\beta}\right\}\right]$$

- The score function:

$$U_{\xi}(\boldsymbol{\beta}) = \Sigma_{i=1}^{k} \left\{z_{\xi i} - A_{\xi i}(\boldsymbol{\beta})\right\}$$

where $A_{\xi i}(\boldsymbol{\beta})$ is the weighted average of $z_{\xi}$ in $\Re$

- The information matrix the sum of weighted covariance matrices
- $\hat{\boldsymbol{\beta}}$ found by iteration

Derivation was controversial

# Is it a conditional likelihood?

- $\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\{\mathbf{z}_\ell \boldsymbol{\beta}\}}$: interpreted as cond. prob. that individual $(i)$ is the one failing at time $t_{(i)}$, given that a failure occurs at $t_{(i)}$
- but it is given $\Re(t_{(i)})$
- equivalent to conditioning on the history of the process up to $t$
- independent of times $\Rightarrow$ conditional probabilities for the ranks: $l(\boldsymbol{\beta})$ is marginal log lik. of the ranks (Kalbfleisch and Prentice, 1973)
- Cox (1975) calls it Partial Likelihood (PL) and shows max PLE consistent and asym. normal, with asym. covariance matrix estimated consistently (ordering of $t_{(i)}$ defines a nesting of conditioning events: $U$ and $\mathfrak{I}$ derived cond. but hold uncond.)
- Tsiatis (1981) shows this more formally using empirical processes

# Is it a conditional likelihood?

- $\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\{\mathbf{z}_\ell \boldsymbol{\beta}\}}$: interpreted as cond. prob. that individual $(i)$ is the one failing at time $t_{(i)}$, given that a failure occurs at $t_{(i)}$
- but it is given $\Re(t_{(i)})$
- equivalent to conditioning on the history of the process up to $t$
- independent of times $\Rightarrow$ conditional probabilities for the ranks: $l(\boldsymbol{\beta})$ is marginal log lik. of the ranks (Kalbfleisch and Prentice, 1973)
- Cox (1975) calls it Partial Likelihood (PL) and shows max PLE consistent and asym. normal, with asym. covariance matrix estimated consistently (ordering of $t_{(i)}$ defines a nesting of conditioning events: $U$ and $\mathfrak{I}$ derived cond. but hold uncond.)
- Tsiatis (1981) shows this more formally using empirical processes

# Is it a conditional likelihood?

- $\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\{\mathbf{z}_\ell \boldsymbol{\beta}\}}$: interpreted as cond. prob. that individual $(i)$ is the one failing at time $t_{(i)}$, given that a failure occurs at $t_{(i)}$
- but it is given $\Re(t_{(i)})$
- equivalent to conditioning on the history of the process up to $t$
- independent of times $\Rightarrow$ conditional probabilities for the ranks: $l(\boldsymbol{\beta})$ is marginal log lik. of the ranks (Kalbfleisch and Prentice, 1973)
- Cox (1975) calls it Partial Likelihood (PL) and shows max PLE consistent and asym. normal, with asym. covariance matrix estimated consistently (ordering of $t_{(i)}$ defines a nesting of conditioning events: $U$ and $\mathfrak{I}$ derived cond. but hold uncond.)
- Tsiatis (1981) shows this more formally using empirical processes

# Is it a conditional likelihood?

- $\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\{\mathbf{z}_\ell \boldsymbol{\beta}\}}$: interpreted as cond. prob. that individual $(i)$ is the one failing at time $t_{(i)}$, given that a failure occurs at $t_{(i)}$
- but it is given $\Re(t_{(i)})$
- equivalent to conditioning on the history of the process up to $t$
- independent of times $\Rightarrow$ conditional probabilities for the ranks: $l(\boldsymbol{\beta})$ is marginal log lik. of the ranks (Kalbfleisch and Prentice, 1973)
- Cox (1975) calls it Partial Likelihood (PL) and shows max PLE consistent and asym. normal, with asym. covariance matrix estimated consistently (ordering of $t_{(i)}$ defines a nesting of conditioning events: $U$ and $\mathfrak{I}$ derived cond. but hold uncond.)
- Tsiatis (1981) shows this more formally using empirical processes

# Is it a conditional likelihood?

- $\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\{\mathbf{z}_\ell \boldsymbol{\beta}\}}$: interpreted as cond. prob. that individual $(i)$ is the one failing at time $t_{(i)}$, given that a failure occurs at $t_{(i)}$
- but it is given $\Re(t_{(i)})$
- equivalent to conditioning on the history of the process up to $t$
- independent of times $\Rightarrow$ conditional probabilities for the ranks: $l(\boldsymbol{\beta})$ is marginal log lik. of the ranks (Kalbfleisch and Prentice, 1973)
- Cox (1975) calls it Partial Likelihood (PL) and shows max PLE consistent and asym. normal, with asym. covariance matrix estimated consistently (ordering of $t_{(i)}$ defines a nesting of conditioning events: $U$ and $\mathfrak{I}$ derived cond. but hold uncond.)
- Tsiatis (1981) shows this more formally using empirical processes

# Is it a conditional likelihood?

- $\frac{\exp \mathbf{z}_{(i)}\boldsymbol{\beta}}{\sum_{\ell \in \Re\left(t_{(i)}\right)} \exp\{\mathbf{z}_\ell \boldsymbol{\beta}\}}$: interpreted as cond. prob. that individual $(i)$ is the one failing at time $t_{(i)}$, given that a failure occurs at $t_{(i)}$
- but it is given $\Re(t_{(i)})$
- equivalent to conditioning on the history of the process up to $t$
- independent of times $\Rightarrow$ conditional probabilities for the ranks: $l(\boldsymbol{\beta})$ is marginal log lik. of the ranks (Kalbfleisch and Prentice, 1973)
- Cox (1975) calls it Partial Likelihood (PL) and shows max PLE consistent and asym. normal, with asym. covariance matrix estimated consistently (ordering of $t_{(i)}$ defines a nesting of conditioning events: $U$ and $\mathfrak{I}$ derived cond. but hold uncond.)
- Tsiatis (1981) shows this more formally using empirical processes

# Outline

# Treatment of ties

If there are tied events the PL is not appropriate.
To deal with this the paper proposes two strategies:

# Treatment of ties

If there are tied events the PL is not appropriate.
To deal with this the paper proposes two strategies:

1. If few, a correction of the term contributing to the partial likelihood
2. If several, PH model replaced by a proportional odds (PO) model:

$$\frac{\lambda(t;\mathbf{z})\,dt}{1-\lambda(t;\mathbf{z})\,dt} = \frac{\lambda_0(t)\,dt}{1-\lambda_0(t)\,dt}\exp(\mathbf{z}\boldsymbol{\beta})$$

where $\lambda_0(t) = Pr(T \leq t+1 | T > t)$ arbitrary

# Treatment of ties

If there are tied events the PL is not appropriate.
To deal with this the paper proposes two strategies:

1. If few, a correction of the term contributing to the partial likelihood
2. If several, PH model replaced by a proportional odds (PO) model:

$$\frac{\lambda(t;\mathbf{z})\,dt}{1-\lambda(t;\mathbf{z})\,dt} = \frac{\lambda_0(t)\,dt}{1-\lambda_0(t)\,dt}\exp(\mathbf{z}\boldsymbol{\beta})$$

where $\lambda_0(t) = Pr(T \leq t+1 | T > t)$ arbitrary

Similar arguments lead to PLE

PO model $\Rightarrow$ PH model as the intervals become infinitesimal

# Additional insights

Other insights, both methodological and relevant for applications:

1. Dealing with the two sample problem:
   - with PH model this becomes a comparison of $\lambda_0(t)$ and $e^{\beta_1}\lambda_0(t)$
   - score test from the PL for discrete times equivalent to Mantel's test and asym. equivalent to log-rank test
   - novelty: it can be applied to continuous exposures

# Additional insights

Other insights, both methodological and relevant for applications:

1. Dealing with the two sample problem:
   - with PH model this becomes a comparison of $\lambda_0(t)$ and $e^{\beta_1}\lambda_0(t)$
   - score test from the PL for discrete times equivalent to Mantel's test and asym. equivalent to log-rank test
   - novelty: it can be applied to continuous exposures
2. Departures from proportionality:
   - formulation of both PH and PO allows for time varying explanatory variables
   - special case: covariate generated from the interaction between a time fixed variable and time
   - This allows testing the proportional assumption

# Additional insights

Other insights, both methodological and relevant for applications:

1. Dealing with the two sample problem:
   - with PH model this becomes a comparison of $\lambda_0(t)$ and $e^{\beta_1}\lambda_0(t)$
   - score test from the PL for discrete times equivalent to Mantel's test and asym. equivalent to log-rank test
   - novelty: it can be applied to continuous exposures
2. Departures from proportionality:
   - formulation of both PH and PO allows for time varying explanatory variables
   - special case: covariate generated from the interaction between a time fixed variable and time
   - This allows testing the proportional assumption
3. Estimating failure time dsn: generalization of the product limit estimation of $\lambda(t)$

# Outline

# What was new

The paper is well known although maybe not so well read

- Aspects that are well known:
  - semi-parametric PH model and its estimation approach
  - ability to perform score tests for continuous exposures

# What was new

The paper is well known although maybe not so well read

- Aspects that are well known:
  - semi-parametric PH model and its estimation approach
  - ability to perform score tests for continuous exposures
- Aspects that are not so well known:
  - checking of PH assumption
  - solutions for tied event times, including semi-parametric PO model for discrete times
  - estimation of cumulative failure time distribution
  - extension to multivariate $T$ and links to the accelerated failure time models
  - physical versus empirical interpretation of the model

# Outline

# Methodological influences

Extremely influential paper methodologically on three 'time scales':

1. at the time origin: the Society discussion
2. in the next 10 years
3. in the following 20 years and beyond

# Outline

# 1 – The Society Discussion

Highlights of that discussion:

1. Richard Peto: proposed an alternative approach to dealing with tied events
2. Jack Kalbfleish and Ross Prentice: raised questions regarding the 'conditional' likelihood
3. Norman Breslow: showed how the baseline cumulative hazard function could be estimated in a more natural way
4. Susannah Howard: showed how easily max. PL estimation could be performed

# Outline

# 2 – The next 10 years

(a) The development of the theory of partial likelihood

(b) The incorporation within counting processes theory:

- Andersen and Gill (1982) simplified and generalized the results on asym. properties of PLE using martingale theory for counting processes (from Aalen 1975)
- Indeed this viewpoint is required for an elegant derivation of these properties
- PH model played a key role for this powerful methodological development

# 2 – The next 10 years

(a) The development of the theory of partial likelihood

(b) The incorporation within counting processes theory:

- Andersen and Gill (1982) simplified and generalized the results on asym. properties of PLE using martingale theory for counting processes (from Aalen 1975)
- Indeed this viewpoint is required for an elegant derivation of these properties
- PH model played a key role for this powerful methodological development

# Outline

# 3 – The next 20 years and beyond

- Interest in semi-parametric models exploded following PH model, prompting huge developments in theory (Bickel et al, 1998)
- These developments are increasingly important in causal inference, missing data, *etc.* (Tsiatis, 2006)

# 3 – The next 20 years and beyond

- Interest in semi-parametric models exploded following PH model, prompting huge developments in theory (Bickel et al, 1998)
- These developments are increasingly important in causal inference, missing data, *etc.* (Tsiatis, 2006)

> Would we be using doubly robust methods, efficient g-estimation, targeted ML *etc.* had "Regression Models and Life-Tables" not been published?

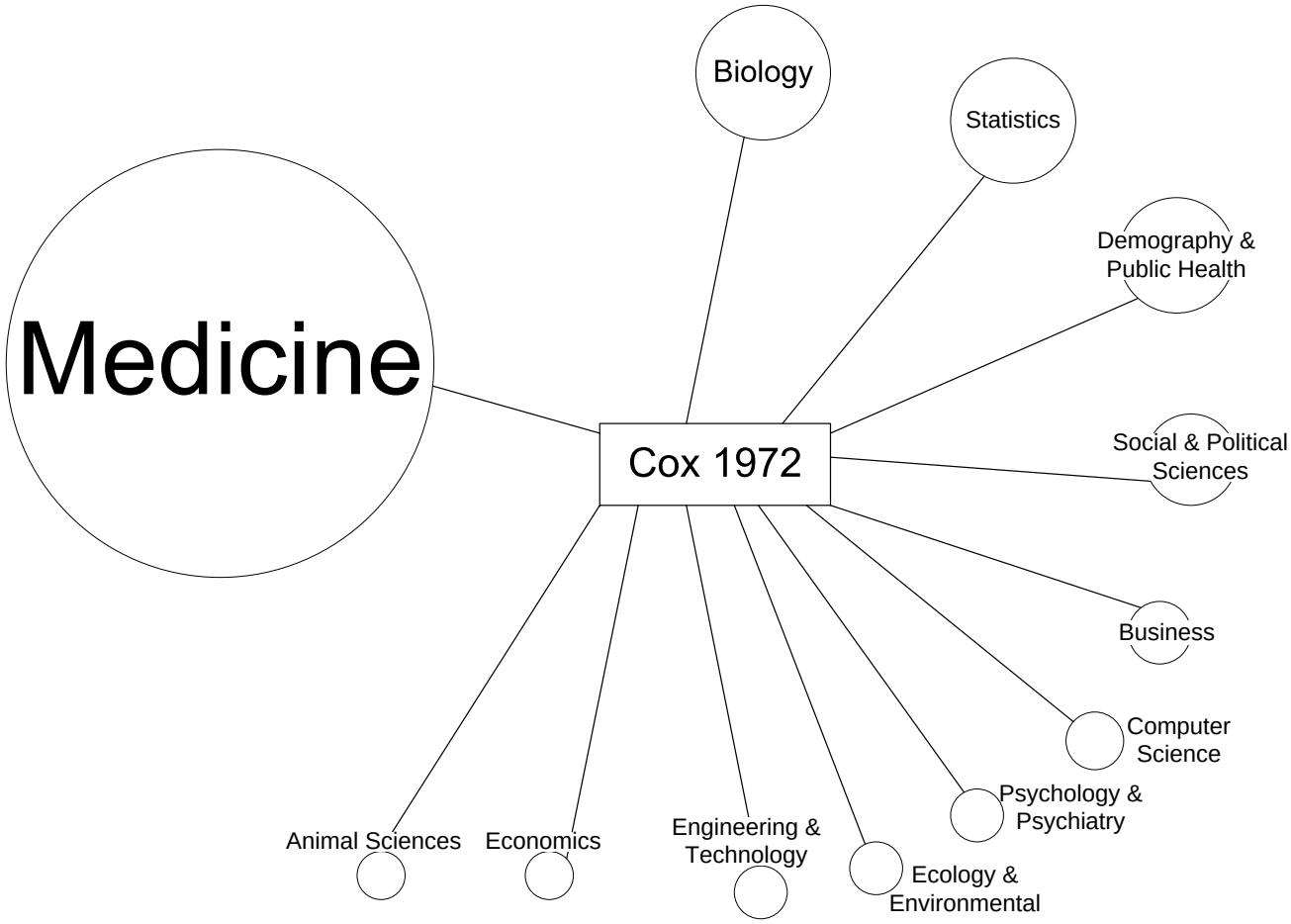
# Outline

# Applications

- The paper states that the proposed methodology will be for *"applications in industrial reliability studies and in medical statistics"*
- Was this a fair prediction?

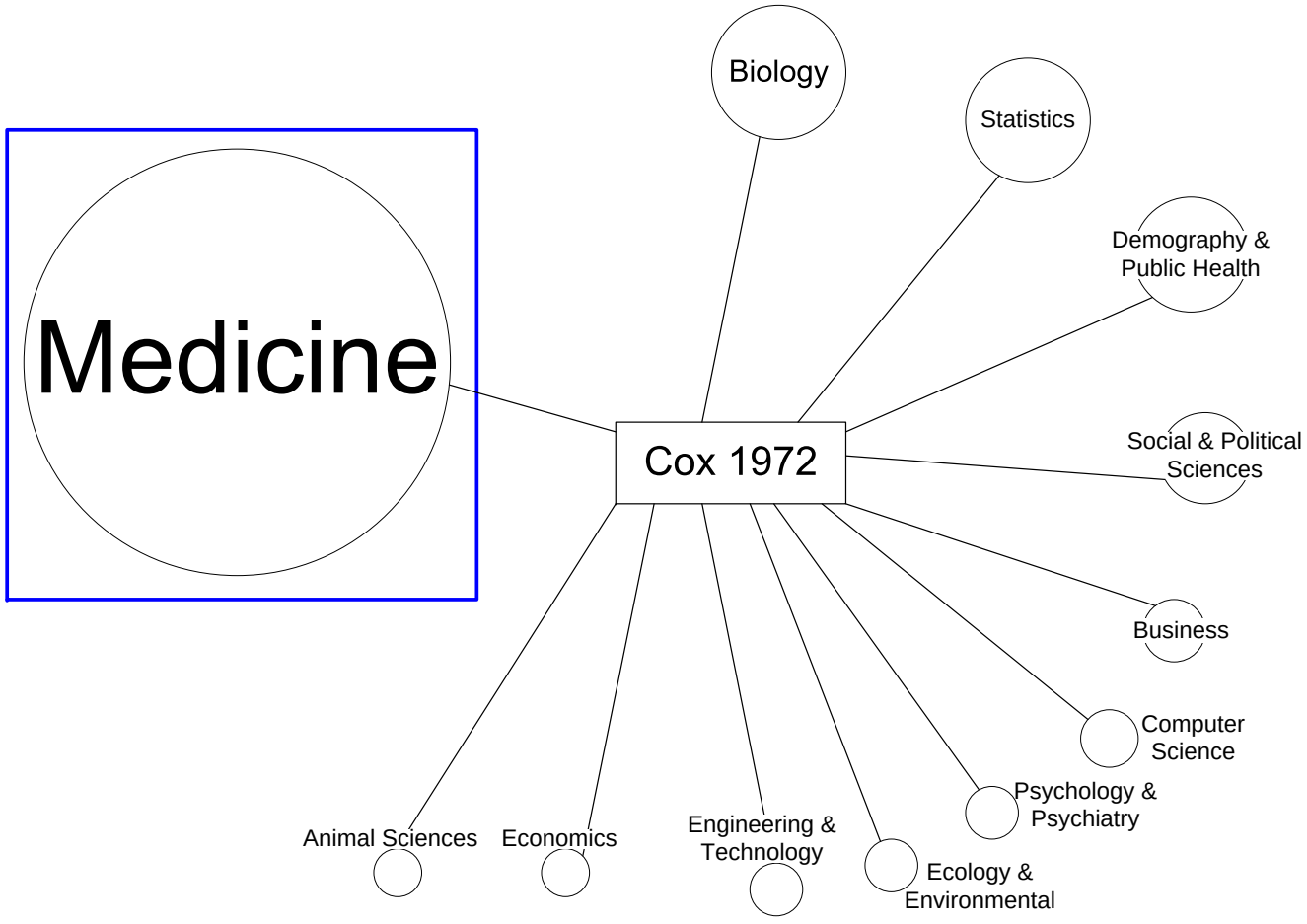
# Citations from Web of Science

## Area proportional to number of citations



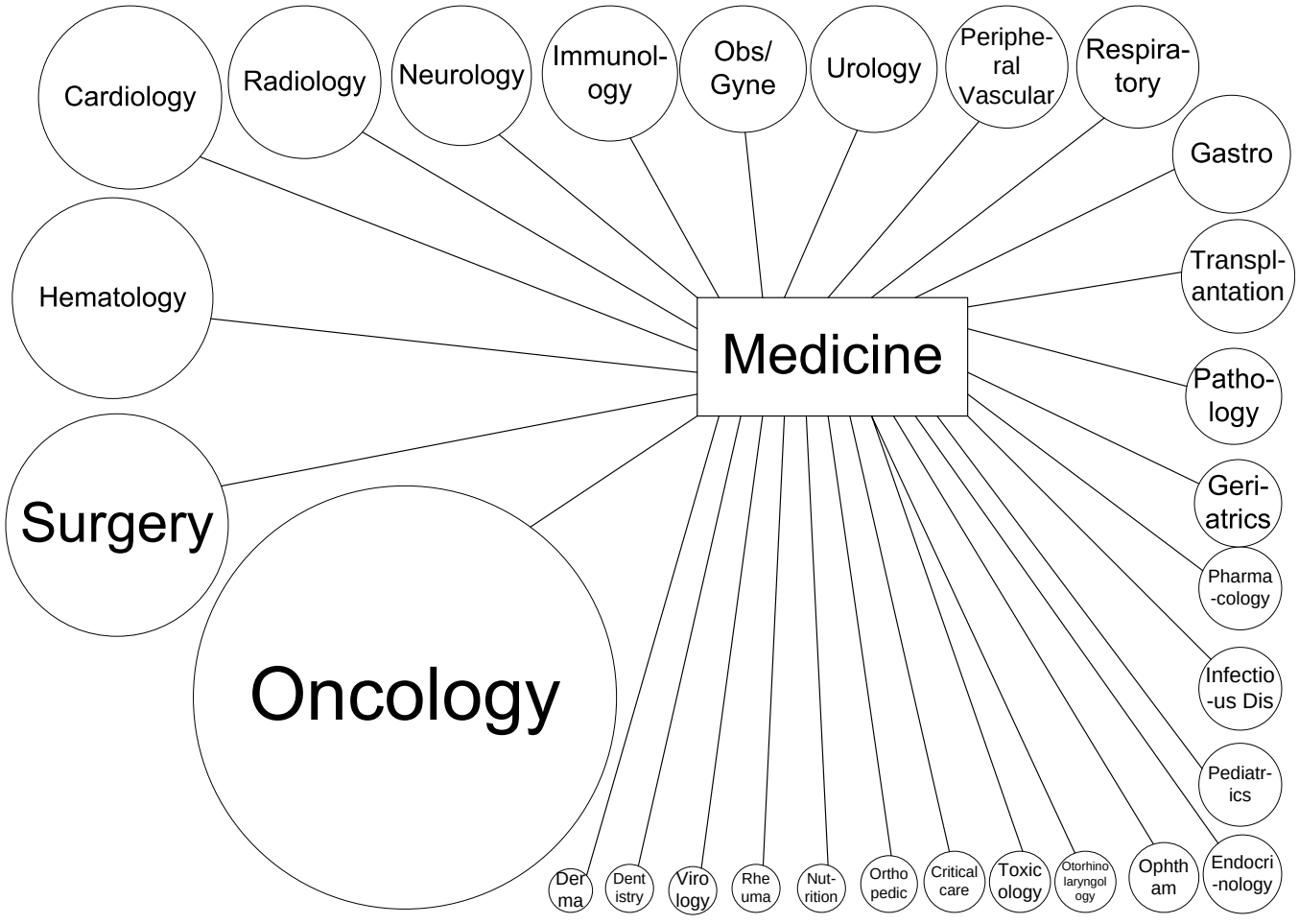
# Citations from Web of Science

## Area proportional to number of citations



# Citations from Web of Science

## Area proportional to number of citations



# Outline

# Final thoughts

Votes of thanks aired on 8 March 1972:

> *As usual [David Cox's] statistical ideas are of both theoretical interest and great practical importance.*
>
> *(F. Downton)*

# Final thoughts

Votes of thanks aired on 8 March 1972:

*As usual [David Cox's] statistical ideas are of both theoretical interest and great practical importance.*

*(F. Downton)*

*. . . he has opened up new territories to common sense.*

*(R. Peto)*

# Sources


Andersen PK (1991)

Survival analysis 1982–1991: the second decade of the proportional hazards regression model.

*Statistics in Medicine*, 10:1921–1941.

Cox DR, Oakes D (1984)

*Analysis of Survival Data.*

Chapman and Hall: London.

Kalbfleisch JD, Prentice RL (2000)

*The Statistical Analysis of Failure Time Data.*

John Wiley: Canada.

Oakes D (1991)

Life table analysis. In: *Statistical Theory and Modelling*, Hinkley DV, Reid N, Snell EJ (eds).

Chapman and Hall: London.

# References (1)


Aalen OO (1975)

Statistical inference for a family of counting processes.

Ph.D. thesis, University of California, Berkeley.

Andersen PK, Gill RD (1982)

Cox's regression model for counting processes: a large sample study.

*Annals of Statistics*, 10:1100–1120.

Bickel et al (1998)

*Efficient and Adaptive Estimation for Semiparametric Models*.

Springer: New York.

Breslow N, Crowley J (1974)

A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship.

*Annals of Statistics*, 2:437–453.

# References (2)


Cox DR (1972)

Regression models and life-tables.

*JRSS B*, 34:187–220.

Cox DR (1975)

Partial likelihood.

*Biometrika*, 62:269–276.

Kaplan EL, Meier P (1958)

Nonparametric estimation from incomplete observations

*JASA*, 53:457–481.

Peto R, Peto J (1972)

Asymptotically efficient rank invariant test procedures.

*JRSS B*, 135:185–207.

# References (3)

Tsiatis AA (1981)

A large sample study of Cox's regression model.

*Annals of Statistics*, 9:93–108.

Tsiatis AA (2006)

*Semiparametric Theory and Missing Data*.

Springer: New York.