Prognosis walk-through

Gustaf Edgren, MD PhD
Associate professor of epidemiology
Karolinska Institutet

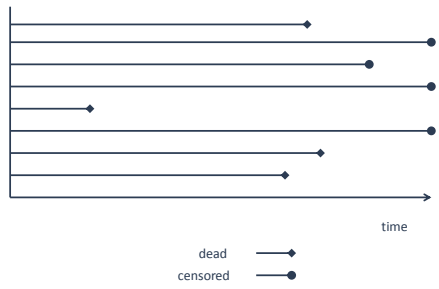
Overview

- (Brief) repetition of survival analysis
- What is a prognosis model?
- Prediction modeling vs. etiologic modeling
- How to build a prognosis model?
- Common misconceptions and errors
- Application

What is survival analysis?

- Methods to estimate the risk of death
- The time factor must be considered (somehow) since everyone dies in the end
- Typically this is done by considering the time until the occurrence of death
- Can be used for other outcomes than death

Survival data



Life table method

Year	N	# dead	Alive	Risk of death	Survival	Cumulative survival	Cumulative risk of death
1	100	8	92	8/100=**0,08**	92/100=**0,92**	**0,92**	**0,08**
2	92	7	85	7/92=**0,076**	85/92=**0,92**	0,92x0,92=**0,85**	1,0-0,85=**0,15**
3	85	5	80	5/85=**0,06**	80/85=**0,95**	0,92x0,92x0,95=**0,80**	1,0-0,80=**0,20**

Life table with censoring

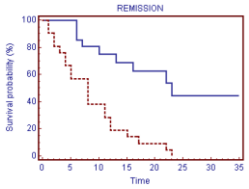
Year	N	Censored	# at risk	Dead	Alive	Risk of death	Survival	Cumulative survival	Cumulative mortality
1	100	5	97,5	8	89,5	8/97,5 =**0,082**	89,5/97,5 =**0,92**	**0,92**	1-0,92 =**0,08**
2	87	4	85	7	78	7/85 =**0,082**	78/85 =**0,92**	0,92x0,92 =**0,85**	1-0,85 =**0,15**
3	76	6	73	5	68	5/73 =**0,068**	68/73 =**0,93**	0,85x0,93 =**0,79**	1-0,79 =**0,21**

Kaplan-Meier method

Time to event	At risk	Survival	Period specific survival	Cumulative survival
7	20	17	17/20=**0,85**	**0,85**
8	17	16	16/17=**0,94**	0,85x0,94=**0,80**
10	16	15	15/16=**0,94**	0,85x0,94x0,94=**0,75**

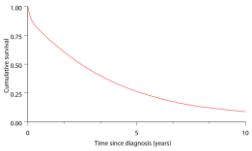
Kaplan-Meier method

- Blue line based on data from previous slide
- Survival changes at each event, but remains constant between these
- Censoring is normally not indicated in graph, but affects the size of the ”jump” at the next event



Modeling survival

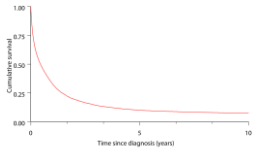
Myeloma



Median survival=2.44 years

10-year survival = 7.7%

- AML



Median survival=0.44 years

10-year survival = 7.3%

What is prognosis?

- Prognosis is a prediction of the future outcome based on the current state
- Typically deals with binary outcomes (where time is of the essence)
- Prognosis is typically dealt with using various multi-variate models

Why are we interested in prognosis?

- Patients want to know
- Doctors want to know
- Helps guide treatment allocation

How do we communicate prognosis?

”You have a 30% chance of surviving 5 years”

”Patients with your particular disease have a 30% chance of surviving 5 years”

” Patients with your particular disease, in your age group, of the same sex as you, with the same other diseases and with the same genetic mutations have a 30% chance of surviving 5 years”

…

Prognostic models

- Patients' prognosis typically depend on a range of known (and unknown) factors
- The combined effect of these factors can be measured and aggregated using multivariate statistical models
- The aim is to get a better understanding of what the prognosis is for a particular patient
- Problems with individualization of risk persists…

Prognosis vs. etiology

- Etiologic studies are typically also based on multivariate statistical models
- Importantly, for prediction/prognostication, one does not have to care (the slightest) about causality and confounding
- A prognosis model can be very accurate without inclusion of a single causal factor
- Also, absolute risks are generally preferred in prognostic studies

Prognostic models, cont.

- Depending on what you want to estimate, prognosis models are typically based on Cox regression
- The goal is to estimate what "weight" should be given to the different prognostic factors

Step-by-step development

1. Choice of possible (available) prognostic factors
2. Collection of data
3. Wait, wait, wait …
4. Design your model, choose your prognostic factors
5. Calibrate model
6. Validate model
7. Test in real life – does it matter in clinical practice?

Focus today

Choice of predictors

- Actually quite simple!
- Depends on:
  - What data do you have access to?
  - What do you think is important, clinically?
  - How many factors do you have power to estimate?
  - What is practical to use?
  - Previous experience…

Predictors, cont.

- Typically, one includes: demographic factors, clinical parameters, laboratory tests, previous treatment*, disease characteristics
  - Current treatment not suitable
- Objective and reproducible factors
- Keep in mind that the factors must be easy to acquire for future application

Predictors, cont.

- Depending on situations, and if there is sufficient power, it can be suitable to consider interactions
- But should be based on biological/clinical reasoning, not on statistical mass-testing

Study design

- Natural chocie is standard, prospective (retrospective) cohort design
- Theoretically possible to develop prognosis scores using a case-control design, but it is usually more complex

What outcome do I want to predict?

- Choice based on relevance and reproducibility
- The most robust is often all-cause mortality, but may be impractical and underpowered for most situations
- Proxy-/surrogate outcomes such as disease recurrence, disease progression or death*

Example dataset!

- Publicly available melanoma dataset on 1386 operated patients
- Available variables:
  - Time to death/recurrence, months (0-187; 0% missing)
  - Death/recurrence (0,1; 0% missing)
  - Clark's stage (1-4; 6,5% missing)
  - Sex (1,2; 0.4% missing)
  - Ulcerating tumor (0, 1; 25% missing)
  - Tumor thickness, mm (0-20 mm; 0.2% missing)
  - Localization (1-4; 0,1% missing)

Other variables

The FREQ Procedure

CLARK	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	90	6.49	90	6.49
1	7	0.51	97	7.00
2	260	18.76	357	25.76
3	423	30.52	780	56.28
4	606	43.72	1386	100.00

status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1062	76.62	1062	76.62
1	324	23.38	1386	100.00

sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	6	0.43	6	0.43
0	574	41.41	580	41.85
1	806	58.15	1386	100.00

ulcer	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	346	24.96	346	24.96
0	731	52.74	1077	77.71
1	309	22.29	1386	100.00

LOCCODE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	2	0.14	2	0.14
1	511	36.87	513	37.01
2	601	43.36	1114	80.38
3	240	17.32	1354	97.69
4	32	2.31	1386	100.00

So, how is this done in practice?

- Commonly the cohort is based on some sort of routine data collection (e.g. through surveys or from a clinical database)
- Follow-up can be done successively or at some time-point through record linkage with a cause of death register

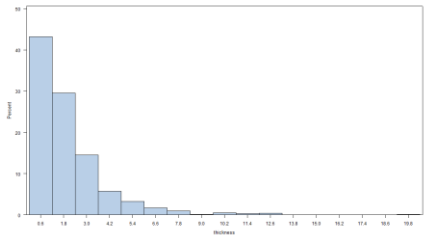
Practical details, cont.

- After the data collection, it is relevant to describe and analyze data roughly, one variable at a time:
  - Mean, median, distribution
  - Proportion missing?
  - Association with survival? Plot and consider!
  - Linear? Non-linear? Test different transformations! E.g. 1/x, log(x), x^2, etc.

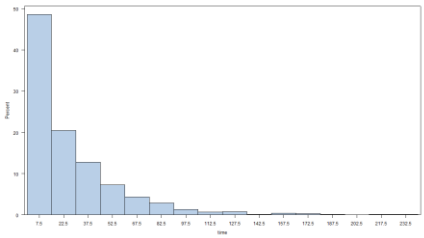
Practical details, cont.

- Some simple rule of thumb:
  - Consider the type of variable (ordinal, nominal, etc.)
  - Avoid categorization of continuous variables (unless necessary)
  - Consider truncating variables that are extremely distributed
  - Avoid including several strongly correlated variables in the same model

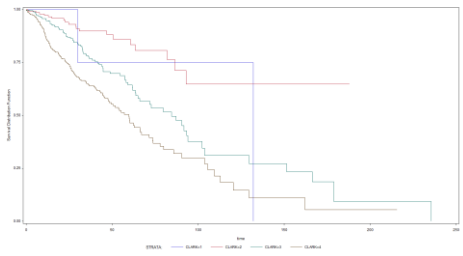
Thickness



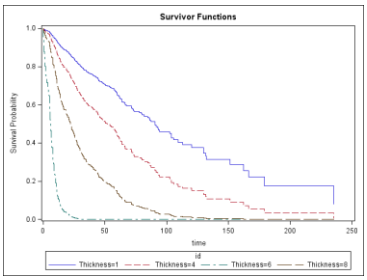
Follow-up time



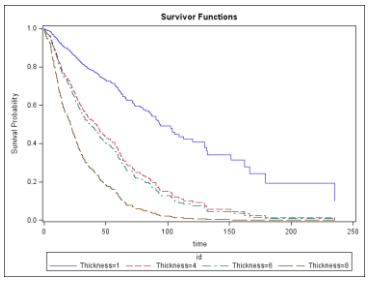
Clark (categorical)



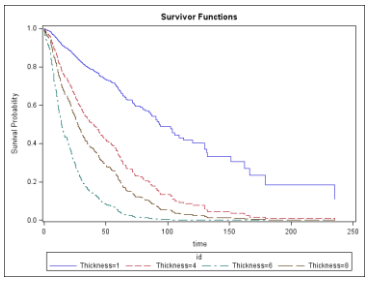
Thickness (linear function)



Thickness (linear + quadratic)



Thickness (natural spline)



Missing data

- Dependeing on how data is collected, some variables will always have missing values
- Why are they missing? Random? Related to outcome or prognosis? Important!
- How much is missing?
- Be cautious before excluding subjects solely based on missing data!

Missing data, forts.

- Because missing data may render an observation useless (depending on method) , researchers often use some method for missing data imputation
- Imputation, in simple terms, means that one makes a qualified guess what the value should be, based on one or many other variables
- Better than throwing out the variable, but leads to loss of power and decreased predictive ability

Estimation

- Once you've picked your predictors and how they should be analyzed, the estimation is more straightforward (…), i SAS:

```
proc phreg data=melanoma;
class clark loccode;
model Time*status(0)=
  thickness t2
  clark
  sex
  ulcer
  loccode
  ;
run;
```

Result of the first estimation

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	116.6422	10	<.0001
Score	132.8765	10	<.0001
Wald	114.4569	10	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
thickness	1	11.8628	0.0006
t2	1	0.4175	0.5182
CLARK	3	16.1900	0.0010
sex	1	3.9614	0.0466
ulcer	1	2.9613	0.0853
LOCCODE	3	2.0902	0.5539

Analysis of Maximum Likelihood Estimates

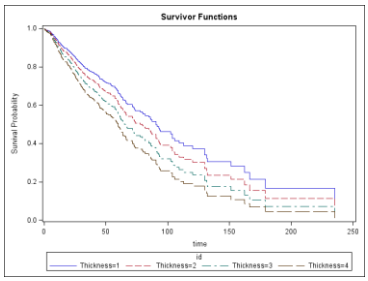
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
thickness	1	0.18312	0.05317	11.8632	0.0006	1.201	
t2	1	-0.00406	0.00628	0.4180	0.5179	0.996	
CLARK 1	1	0.81983	0.74950	1.1965	0.2740	2.270	CLARK 1
CLARK 3	1	0.70224	0.25254	7.7323	0.0054	2.018	CLARK 3
CLARK 4	1	1.01026	0.26101	14.9815	0.0001	2.746	CLARK 4
sex	1	0.24226	0.12172	3.9614	0.0466	1.274	
ulcer	1	0.24241	0.14087	2.9612	0.0853	1.274	
LOCCODE 1	1	-0.39649	0.30658	1.6725	0.1959	0.673	LOCCODE 1
LOCCODE 2	1	-0.43155	0.30315	2.0265	0.1546	0.650	LOCCODE 2
LOCCODE 3	1	-0.43963	0.32563	1.8228	0.1770	0.644	LOCCODE 3

Result of first estimation, cont.

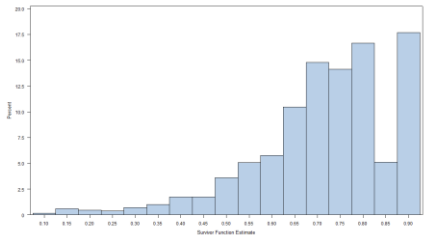
Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	3906.469	3789.827
AIC	3906.469	3809.827
SBC	3906.469	3847.635

Prediction, based on final model



Predicted 3-year survival

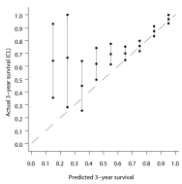


Estimation, cont.

- Techniques to reduce the number of variables are more tricky, can be forward or backwards selection
  - However, as they are based on multiple significance tests, they "consume" statistical power

Calibration (advanced)…

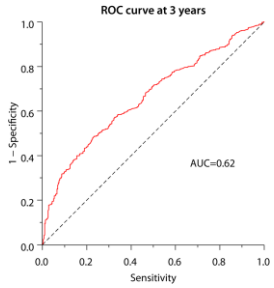
- In principle, calibration is done bycomparinge the model prediction with "the truth"
- The parameters are then corrected to give a more accurate prediction



Model discrimination

- Model discrimination relates to how well the model is at discriminating between those who will die and those who will not
- Traditionally, this is measured and presented as ROC-curves with AUC / c-statistic
- For survival data, where the outcome happens over time, standard methods for assessment of discriminatory ability must be estimated at a fixed time point

ROC-curve at 3-year follow-up



Validation

- The final step is validation
- This can be done both internally or externally (the latter is preferable)

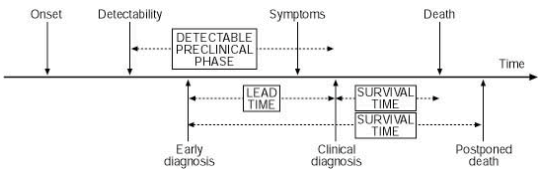
Common errors

- Power
- Systematic errors
  - External validity (due to systematic differences with other populations, e.g. due to lead time)
  - Changing treatment
- Model assumptions

Power

- Can be difficult to assess a priori
- A rule of thumb is that you need 10 patients with the outcome for each included variable – more data, however, is always better!
- More power is needed if one plans to employ variable reduction or test for interactions

Consideration of lead time



Application

- Generally difficult to apply a prognosis model without access to a computer, but examples can be found online:
  http://www3.mdanderson.org/app/medcalc/index.cfm?pagename=coloncancer

Recommended reading

- Steyerberg, Clinical prediction models
- Harrell, FE *et al*. Multivriable prognostic models: …, Stats in medicine, 1996
- Royston P, *et al.* Diagnosis and prognostic research…, BMJ 2009