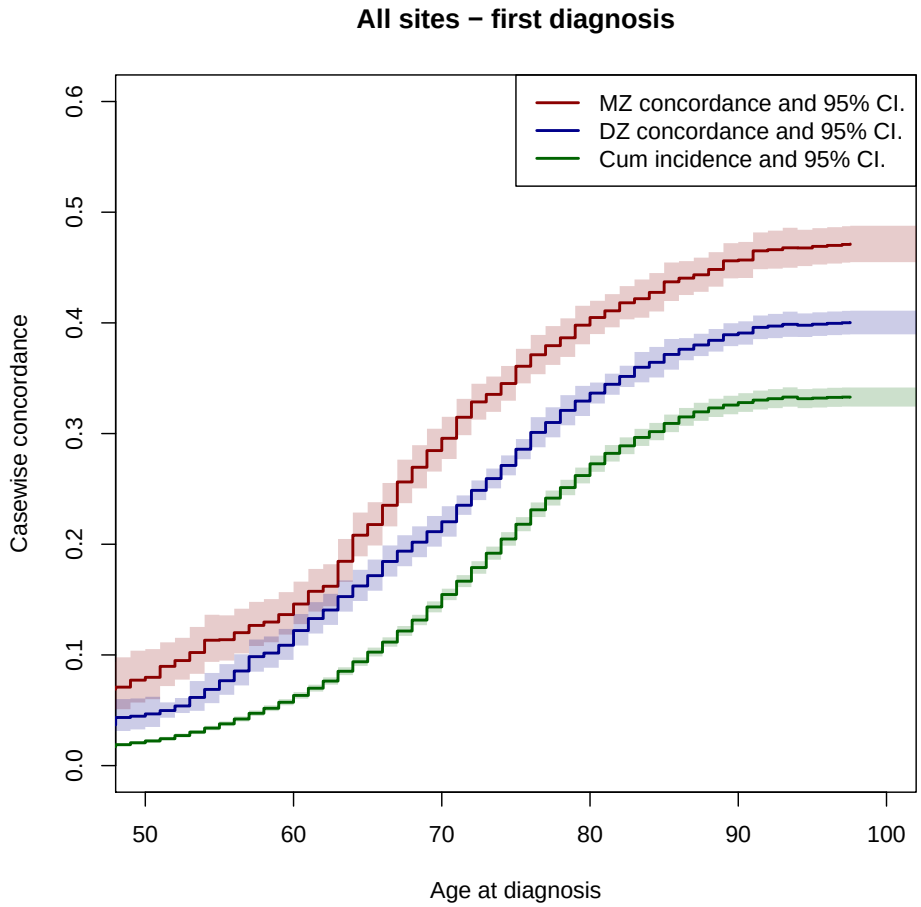
Analysis of Twin Data: Time to Event Models

J. Hjelmborg, T. Scheike and K. Holst

Epidemiology, Biostatistics and Biodemography, Institute of Public Health, University of Southern Denmark, Denmark

May 2018

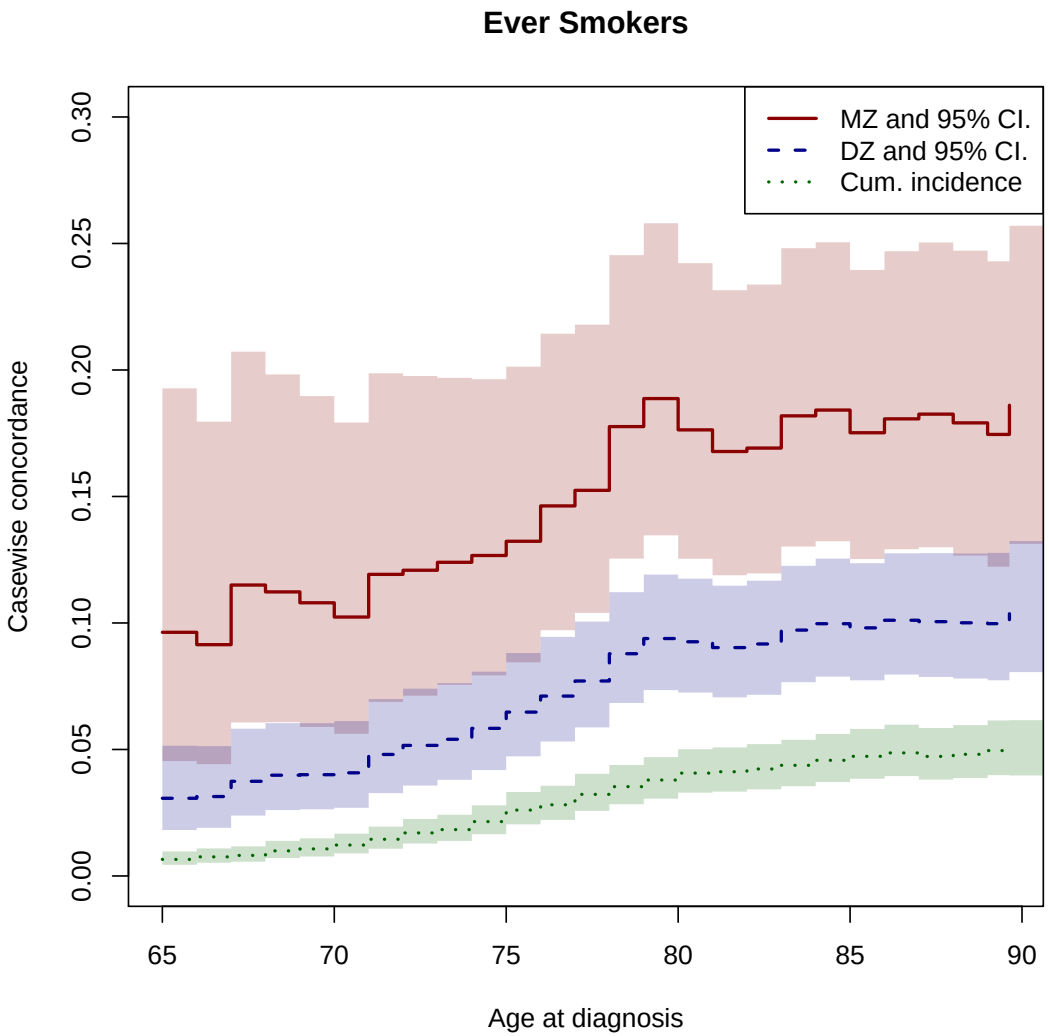
All Cancer



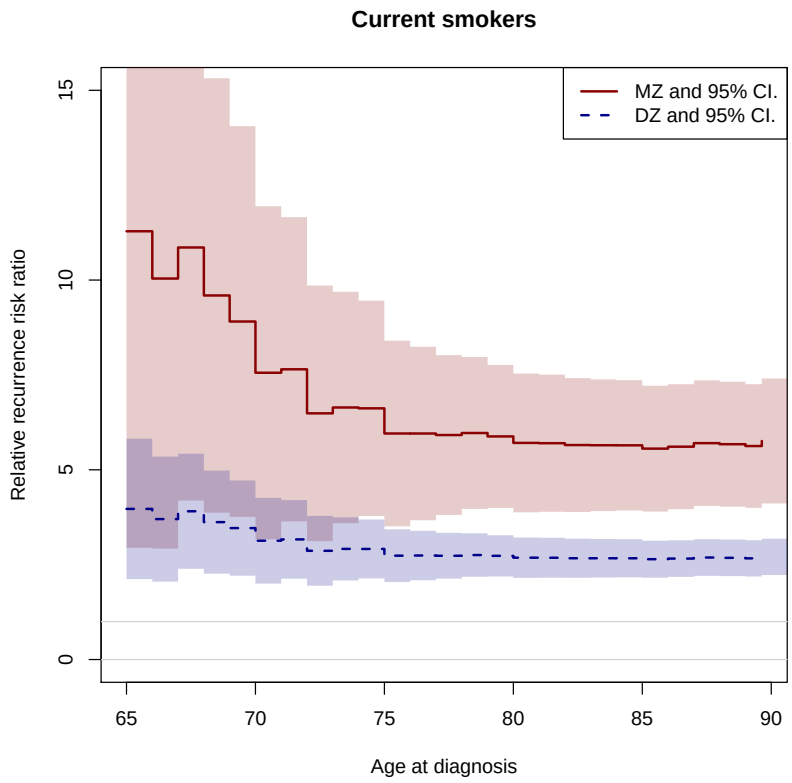
L. Mucci, J. Hjelmborg, T. Scheike, K. Holst, A. Skytthe, H. Adami, N. Holm, J. Harris, J. Kaprio et al. JAMA (2016)

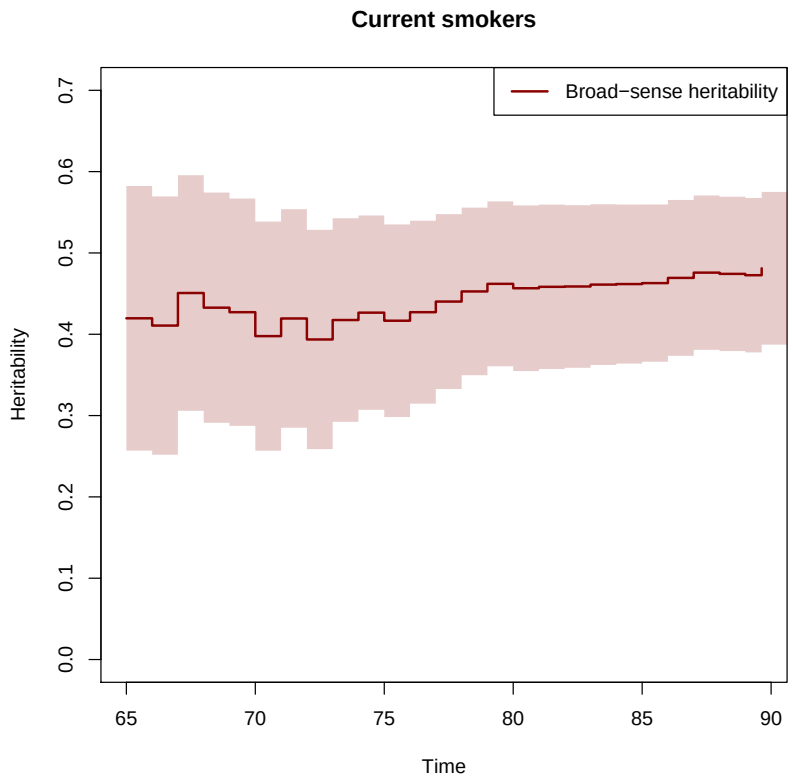
Cancer site	Cumulative risk[1] (%)	N Twin pairs concordant/ discordant		Familial risk[2] (95% CI) – MZ twins	Familial risk (95% CI) – DZ twi
		MZ	DZ		
Overall cancer	32.4%	1383/5887	1933/11461	45.9% (44.1%-47.7%)	37.1% (35.7-38.4
Head and neck[3]	0.8%	5/191	6/361	6.0% (2.4-14.4%)	5.1% (2.2-11.3%
Esophagus	0.4%	0/87	0/183	--	--
Stomach	1.6%	14/338	15/648	6.8% (3.9-11.4%)	4.4% (2.6-7.3%
Small intestine	0.1%	0/32	0/59	--	--
Colon	2.9%	30/577	31/1156	10.9% (7.4-15.8%)	7.9% (5.4-11.4%
Rectum and anus	1.9%	14/440	13/771	6.6% (3.7-11.4%)	5.8% (3.4-9.7%
Liver	0.5%	0/124	2/208	--	--
Gallbladder, extrahepatic bile duct	0.5%	1/110	1/187	0.5% (0-4.7%)	0.3% (0-1.0%)
Pancreas	1.1%	4/234	6/508	4.3% (1.5-11.6%)	3.7% (1.5-8.6%
Nose, sinuses	0.1%	0/21	0/36	--	--
Larynx	0.2%	2/53	1/113	8.4% (2.3-26.4%)	2.7% (1.1-6.1%
Lung, trachea and bronchus	3.2%	50/682	74/1366	17.5% (13.4-22.5%)	13.4% (10.8-16.6
Pleura	0.1%	1/22	0/38	--	--
Bone	0.1%	0/20	0/35	--	--
Melanoma of skin	1.2%	11/342	6/585	19.6% (11.5-31.3%)	6.1% (2.7-13.2%
Skin, non-melanoma	3.0%	16/395	10/618	14.5% (7.5-26.2%)	4.6% (2.4-8.6%
Connective and soft tissues	0.2%	0/57	0/110	--	--
Breast	9.4%	124/1175	141/2223	28.1% (23.9-32.8%)	19.9% (17.0-23.2
Cervix uteri	1.0%	1/210	3/324	--	--
Corpus uteri	2.2%	9/272	6/481	7.0% (3.4-14.0%)	3.6% (1.6-8.0%
Uterus, other	0.1%	0/24	0/36	--	--
Ovary	1.6%	6/234	4/427	8.7% (4.0-17.9%)	2.9% (1.1-7.4%
Other female genital organs	0.4%	0/47	1/84	--	--
Penis and other genital organs	0.1%	0/15	0/34	--	--
Prostate	10.5%	197/807	148/1719	38.0% (33.9-42.2%)	22.0% (18.8-25.7
Testis	0.5%	5/90	3/123	13.8% (5.7-29.6%)	6.0% (1.9-16.9%
Kidney	0.8%	5/196	2/374	6.7% (2.8-15.1%)	1.8% (0.4-6.8%
Bladder, other urinary organs	2.2%	18/471	13/870	9.9% (6.2-15.5%)	5.5% (3.1-9.7%
Eye	0.1%	2/30	0/64	--	--
Brain, central nervous system	0.9%	1/343	3/522	1.7% (0.5-6.2%)	1.8% (0.3-12.0%
Thyroid	0.2%	0/85	1/132	--	--
Hodgkin's disease	0.1%	0/57	0/69	--	--
Multiple myeloma	0.4%	0/114	0/174	--	--
Non-Hodgkin lymphoma	0.7%	1/254	3/466	--	--
Leukemia, acute	0.3%	0/77	0/139	--	--
Leukemia, other	0.6%	5/128	3/259	15.2% (6.1-33.2%)	4.1% (1.3-11.9%

Time-varying genetic influence? Lung Cancer



Time-varying genetic influence? Lung Cancer





J. Hjelmborg, J. Harris, J. Kaprio et al. Thorax (2017)

Effect?

Exposure→Outcome

- Outcome: Time to occuence of event. Event may not occur - can be censored at follow-up.
- What is the contribution of genetic and environmental factors to the variation in risk of outcome?

$$\begin{cases} Y = \text{Genes} + \text{Environment} \\ \Sigma_Y = \Sigma_{\text{Genes}} + \Sigma_{\text{Environment}} \end{cases}$$

- What kind of genetic and environmental influences to expect?
- How does this influence vary with time?

Time in Twin studies

- Suppose we're studying a dichotomous trait; Disease is present or not.
- Suppose data is complete in the sense that status of disease does not change anymore.
- Analysis: prevalence, concordance, correlation and biometric measures - Yes, We Can!
- Example: Stuttering in childhood (questionnaire answered by adults).
- -at least we do not hesitate to assume complete status.

Table: Genetic influence on Stuttering

	Liability threshold model			
	prevalence	concordance	tetrachorics	heritability (95% CI)
MZ females	.04	.47 (.38,.59)	.81 (.71,.87)	.78 (.68,.85)
DZ females	.04	.08 (.04,.16)	.17 (-.02,.35)	AE model
MZ males	.08	.54 (.46,.62)	.79 (.72,.85)	.75 (.66,.82)
DZ males	.08	.10 (.062,.16)	.07 (-.07,.23)	AE model

(Fibiger et al. 2008)

Prevalence and casewise Concordance

	pairs	concordant	discordant	**prevalence**	**concordance** (9
MZ females	1948	33	74	.04	.47 (.38,.59)
DZ females	2404	7	167	.04	.08 (.04,.16)

- casewise concordance rate; *Risk of being affected given that co-twin is affected*
- higher MZ than DZ concordance rate suggests genetic influence (but not how much).
- empirically based similarity-measure.
- easy to estimate (in next slide).
- easy to communicate(!)
- also applicable under casewise ascertainment (to be cont'd).

Prevalence and casewise Concordance

	n pairs	n_{11} concordant	n_{d} discordant	**prevalence**	**conc**
MZ females	1948	33	74	.04	.47 (
DZ females	2404	7	167	.04	.08 (

- Prevalence is proportion of affected individuals,

$$\hat{p} = \frac{2n_{11} + n_{\mathrm{d}}}{2n}$$

- casewise concordance rate,
  $P(\text{twin is affected} \mid \text{co-twin is affected})$,

$$\hat{p_c} = \frac{2n_{11}}{2n_{11} + n_{\mathrm{d}}}$$

Prevalence and casewise Concordance

	n pairs	n_{11} concordant	n_{d} discordant	**prevalence**	**conc**
MZ females	1948	33	74	.04	.47 (
DZ females	2404	7	167	.04	.08 (

- casewise concordance rate,
  $P(\text{twin is affected} \mid \text{co-twin is affected})$,

$$\hat{p}_c = \frac{2n_{11}}{2n_{11} + n_{\mathrm{d}}}$$

- Confidence intervals can be exact or approximate by asymptotic normality assumption using
  $\mathrm{Var}(\hat{p}_c) = \hat{p}_c{}^2(1-\hat{p}_c)^2(\frac{1}{n_{11}} + \frac{1}{n_{\mathrm{d}}})$.

casewise Concordance - Estimation

The dataset:

```
. list  v411 kon1 id1 v412 kon2 id2 zyg  zygbin  alder in 1/5
```

	v411	kon1	id1	v412	kon2	id2	zyg	zygbin	alder
1.	0	kvinde	30098851	0	kvinde	30098852	mz	1	39
2.	0	kvinde	30199351	0	kvinde	30199352	mz	1	27
3.	ja	mand	30186321	0	mand	30186322	mz	1	29
4.	0	kvinde	20044491	0	kvinde	20044492	mz	1	68
5.	0	kvinde	30072841	0	kvinde	30072842	mz	1	42

casewise Concordance - Estimation using Stata


```
. xi: glm ytwin i.zygbin*ycotwin if kon1==0, family(binomial) link(log) cluster( tvparnr)
                            (Std. Err. adjusted for 4352 clusters in tvparnr)
------------------------------------------------------------------------------
             |               Robust
       ytwin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Izygbin_1 |  -.6053746   .1396172    -4.34   0.000    -.8790193   -.3317298
     ycotwin |   .7622305   .3786476     2.01   0.044     .0200949    1.504366
_IzygXycot~1 |   2.412827   .4279456     5.64   0.000     1.574069    3.251585
       _cons |   -3.32167   .0773408   -42.95   0.000    -3.473255   -3.170085
------------------------------------------------------------------------------

. * casewise concordance mz
. lincom ycotwin + _Izygbin_1 + _IzygXycotw_1 + _cons
------------------------------------------------------------------------------
       ytwin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.7519877   .1106554    -6.80   0.000    -.9688683   -.5351071
------------------------------------------------------------------------------
. disp exp(r(estimate))
.47142857

. * 95% lower bound
. disp exp( r(estimate) - 1.96*r(se) )
.37951078
. * 95% upper bound
. disp exp( r(estimate) + 1.96*r(se) )
.58560892
.
. * casewise concordance dz
. lincom ycotwin+_cons
```

casewise Concordance - Estimation using R


```
bp1sex.u <- twinlm(stutter~+strata(sex)+age,data=stut,
                           id="tvparnr",
                           zyg="zyg",DZ="dz",OS="os", pairsonly = TRUE,
                           binary=TRUE,control=list(trace=0),
                           type="u")
score(bp1sex.u)
summary(bp1sex.u)
```

-will similarly estimate concordances. To be cont'd in Practicals.

MZ twins	Estimate	2.5%	97.5%
Concordance	0.03	0.02	0.04
casewise Concordance	0.49	0.39	0.59
Marginal	0.05	0.04	0.07

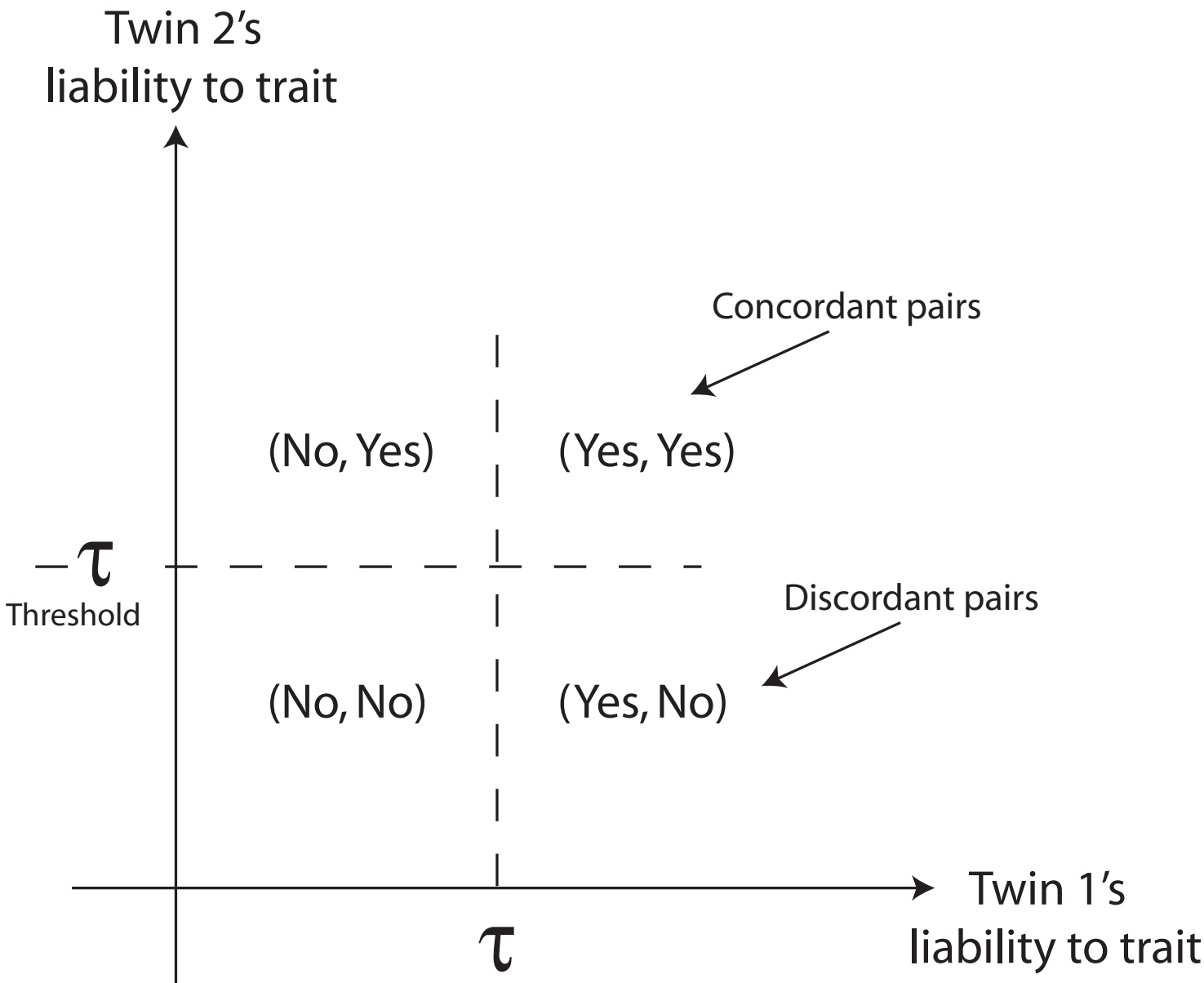
DZ twins	Estimate	2.5%	97.5%
Concordance	0.01	0.00	0.01
casewise Concordance	0.10	0.05	0.20
Marginal	0.05	0.04	0.07

Polychoric correlation - tetrachorics

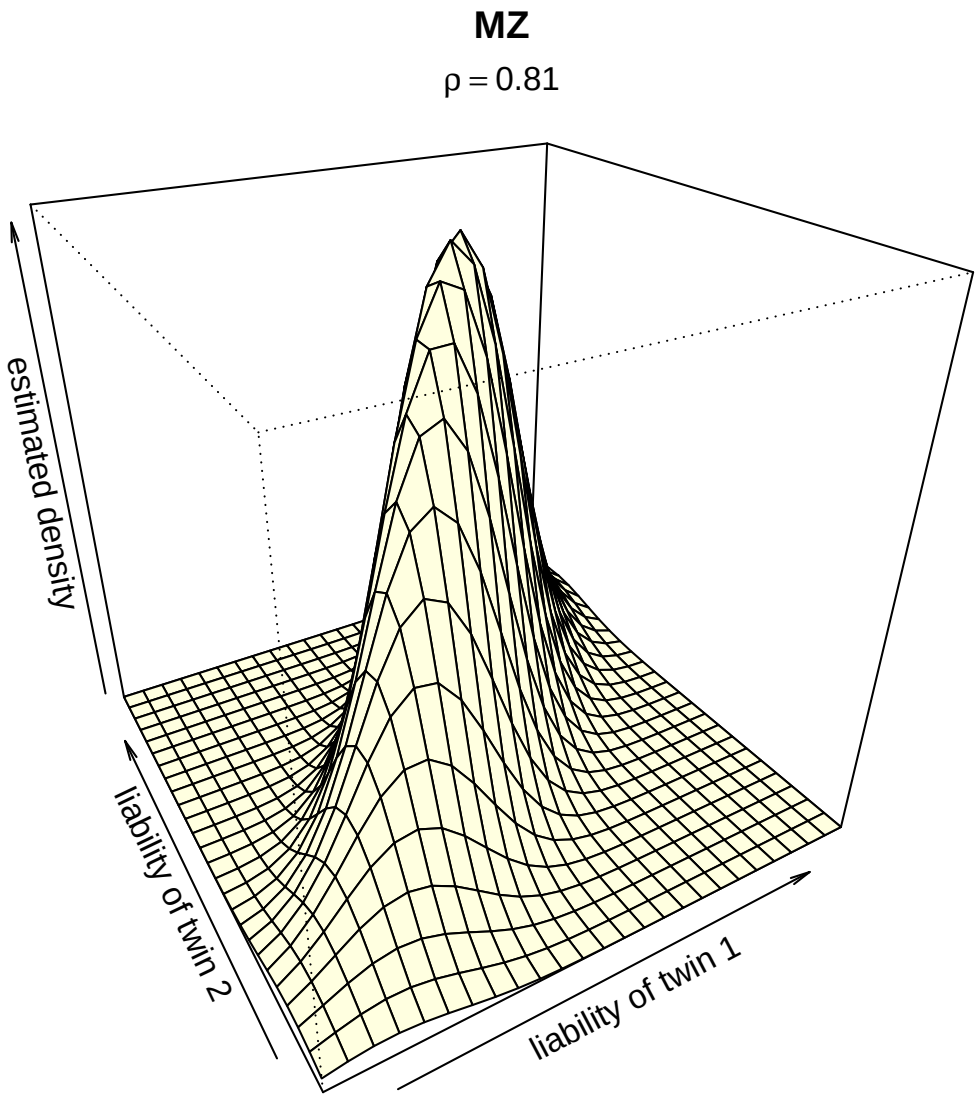
	Liability threshold model		
	prevalence	concordance	**tetrachorics**
MZ females	.04	.47 (.38,.59)	.81 (.71,.87)
DZ females	.04	.08 (.04,.16)	.17 (-.02,.35)

- measure of similarity of twin pairs defined via the *liability-threshold* model (will follow).
- does not depend on prevalence of trait.
- relates to the polygenic quantitative genetics model (ADCE model).

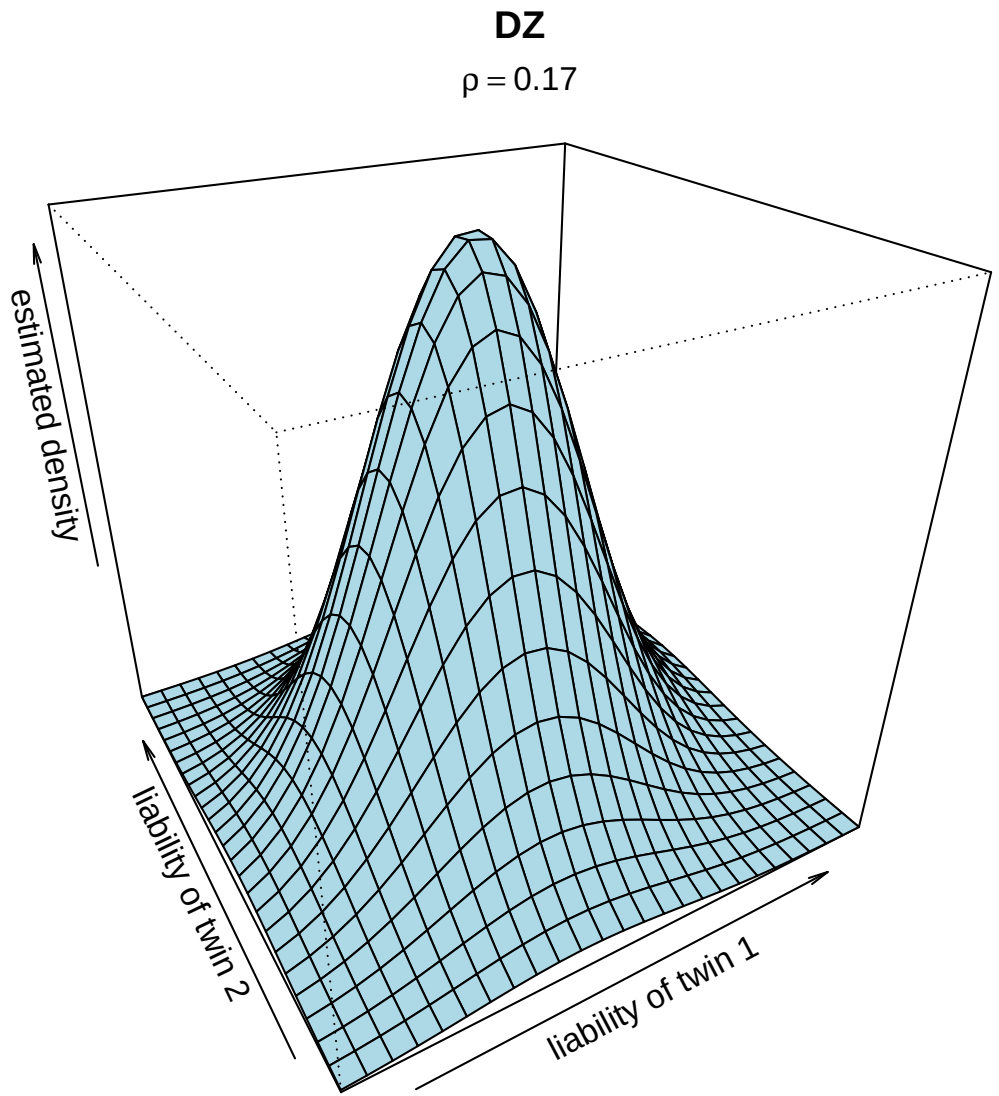
The liability treshold model for dichotomous twin data



The liability treshold model for dichotomous twin data



The liability treshold model for dichotomous twin data



Correlations - Estimation using R


```
bp1sex.u <- twinlm(stutter~+strata(sex)+age,data=stut,
                           id="tvparnr",
                           zyg="zyg",DZ="dz",OS="os", pairsonly = T
                           binary=TRUE,control=list(trace=0),
                           type="u")
```

-will similarly estimate correlations. To be cont'd in Practicals.

	Estimate	2.5%	97.5%
Correlation MZ	0.79	0.69	0.86
Correlation DZ	0.19	-0.001	0.36

The liability treshold model for dichotomous twin data

In summary,

- we assume bivariate standard normality of liabilities
  $(Z_1, Z_2) \sim \mathrm{MvN}\{(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})\}$
- Tetrachoric correlation of categorical variables is by definition the usual correlation in liabilities to outcomes, ρ.
- Thresholds and tetrachorics are estimated from the liability-threshold model .

	prevalence	concordance	**tetrachorics**
MZ females	.04	.47 (.38,.59)	.81 (.71,.87)
DZ females	.04	.08 (.04,.16)	.17 (-.02,.35)

Biometric analyses - polygenic model

	Liability threshold model		
	concordance	tetrachorics	**heritability** (95% CI)
MZ females	.47 (.38,.59)	.81 (.71,.87)	.78 (.68,.85)
DZ females	.08 (.04,.16)	.17 (-.02,.35)	AE model

- Decomposing the liability: $Z_i = A_i + D_i + C_i + E_i$
- Gives usual variance components in polygenic model.
- What is the contribution of genetic and environmental factors to the variation:

$$H_Z^2 = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2}$$

Biometric analyses - polygenic model

- Variance components in polygenic ADCE model satisfies

$$I = \begin{pmatrix} \sigma_A^2 & z\sigma_A^2 \\ z\sigma_A^2 & \sigma_A^2 \end{pmatrix} + \begin{pmatrix} \sigma_D^2 & u\sigma_D^2 \\ u\sigma_D^2 & \sigma_D^2 \end{pmatrix} + \begin{pmatrix} \sigma_C^2 & \sigma_C^2 \\ \sigma_C^2 & \sigma_C^2 \end{pmatrix} + \begin{pmatrix} \sigma_E^2 & 0 \\ 0 & \sigma_E^2 \end{pmatrix}$$

where $z = u = 1$ for MZ pairs, $z = \frac{1}{2}$ and $u = \frac{1}{4}$ for DZ pairs.

Estimation using R

The estimation of the polygenic models, ACE, ADE and AE can be done with:

```
# ACE
bp1.ace <- twinlm(stutter~+sex+age,data=stut,
                          id="tvparnr",
                          zyg="zyg",DZ="dz",OS="os", pairsonly = TRUE,
                          binary=TRUE,control=list(trace=0),
                          type="ace")
score(bp1.ace)
summary(bp1.ace)
AIC(bp1.u,bp1.ace)

# ADE
bp1.ade <- twinlm(stutter~+sex+age,data=stut,
                          id="tvparnr",
                          zyg="zyg",DZ="dz",OS="os", pairsonly = TRUE,
                          binary=TRUE,control=list(trace=0),
                          type="ade")
score(bp1.ade)
summary(bp1.ade)
AIC(bp1.u,bp1.ade)
AIC(bp1.ade,bp1.ace)

# AE
bp1.ae <- twinlm(stutter~+sex+age,data=stut,
                          id="tvparnr",
                          zyg="zyg",DZ="dz",OS="os", pairsonly = TRUE,
                          binary=TRUE,control=list(trace=0),
                          type="ae")
score(bp1.ae)
summary(bp1.ae)
```

Estimation using OpenMx in R (extract)


```
# Matrices for expected Means & Thresholds (on liabilities)
meanG    <-mxMatrix( type="Zero", nrow=1, ncol=ntv, name="expMean" )
threT    <-mxMatrix( type="Full", nrow=1, ncol=ntv, free=TRUE, values=thVals, label="thre", name

# Algebra to compute Total Variance
covP     <-mxAlgebra( expression=A+C+E, name="V" )

# Algebra for expected Variance/Covariance Matrices in MZ & DZ twins
covMZ    <-mxAlgebra( expression= rbind( cbind(A+C+E , A+C),
                                         cbind(A+C   , A+C+E)), name="expCovMZ" )
covDZ    <-mxAlgebra( expression= rbind( cbind(A+C+E     , 0.5%x%A+C),
                                         cbind(0.5%x%A+C , A+C+E)), name="expCovDZ" )

# Constraint on variance of the liability of Binary variables (assumed to have a SND)
matUnv  <-mxMatrix( type="Unit", nrow=nv, ncol=1, name="Unv1" )
var1 <-mxConstraint( expression=diag2vec(V)==Unv1, name="Var1" )

# Data objects for Multiple Groups
dataMZ   <-mxData( observed=mzData, type="raw" )
dataDZ   <-mxData( observed=dzData, type="raw" )

# Objective objects for Multiple Groups
objMZ <-mxFIMLObjective( covariance="expCovMZ", means="expMean", dimnames=selVars, thresholds="e
objDZ <-mxFIMLObjective( covariance="expCovDZ", means="expMean", dimnames=selVars, thresholds="e

#Combine groups (...)
```

Polygenic model - model selection

Liability threshold models							
Females	-2LL	df	ΔX^2	Δdf	p	AIC	note
Saturated	2602.092	8698	0.920	3	0.821	-5.080	
ACE	2609.629	8701					
AE (*)	2609.629	8702	0.002	1	$0.968^{\dagger}$	-1.998	without D
ADE	2609.627	8701					

- Most likely in terms of $-2\log(\text{likelihood})$ with fewest parameters, ie., most parsimonious model, is chosen.
- Saturated model: same treshold for mz and dz (twin 1 and twin 2).
- -is compared to full model, p-value is 0.821, and gives tetrachorics.
- The additive genetic effect is significant in all models
- †this p-value is too conservative and can be halved (Dominicus et al. 2006).
- The AE model is chosen by comparison with ADE.

Biometric analyses - polygenic model

- Model selection - notes:
- Most likely in terms of $-2\log(\text{likelihood})$ with fewest parameters, ie., most parsimonious model, is chosen.
- Testing for a vanishing variance component: Using the χ^2_1 distribution as approximation to likelihood-ratio distribution gives conservative p-values.
- Eg., when dropping 'C' in 'ACE' model the p-value should be halved (Dominicus et al. 2006).

Table: Genetic influence on Stuttering

	Liability threshold model			
	prevalence	concordance	tetrachorics	heritabi
MZ females	.04	.47 (.38,.59)	.81 (.71,.87)	.78 (.68
DZ females	.04	.08 (.04,.16)	.17 (-.02,.35)	AE mod
MZ males	.08	.54 (.46,.62)	.79 (.72,.85)	.75 (.66
DZ males	.08	.10 (.062,.16)	.07 (-.07,.23)	AE mod

(Fibiger et al. 2008)

- These results may be recovered in R using scripts 'stut.R' or 'stutOpenMx.R'.
- -we are done.

Some Finnish males

- Finding the mutations or set of genes
- Classic twin methodology may indicate where to look!

Given ordinal categorical data

- Casewise concordance rates may indicate genetic effects, but not the magnitude and type.
- The liability-threshold model allows for adapting classical measures.
- Tetrachorics are polychoric correlation of dichotomous trait.
- -which is the usual within-pair correlation in liability of trait.
- Model selection and estimation is analogues to continuous case.

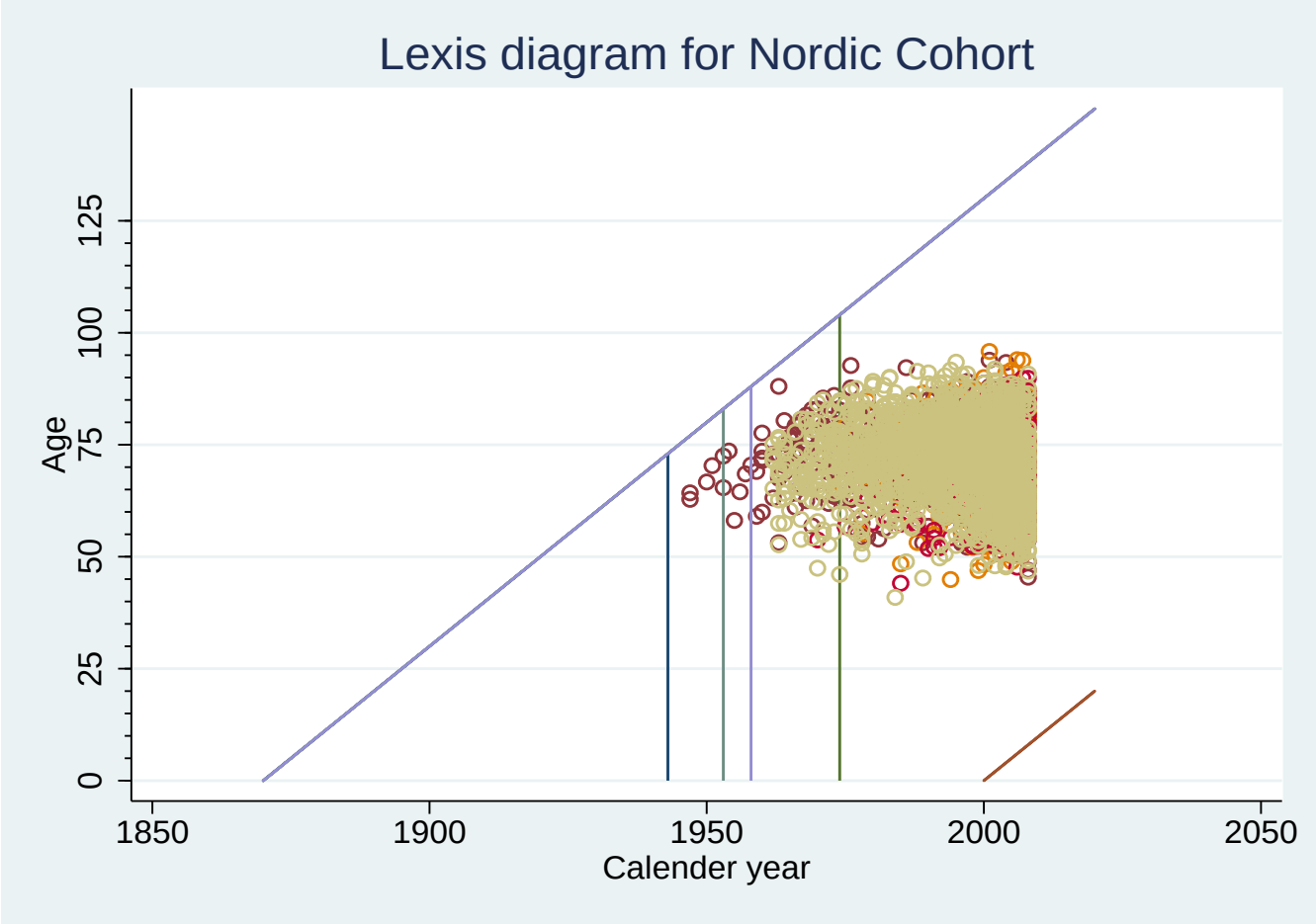
Special topics

- Covariates (modelled via the threshold)
- Probandwise ascertainment: No negatively concordant pairs
- More traits - Pleiotropy? (Multivariate categorical twin data)
- Sex limitation model to include opposite sexed DZ's.
- Other measures of similarity
- Analysis cookbook

Time in twin studies

- Can you think of a study, ie. trait and design, that is not governed by this?
- Data often contain registration of time of events!

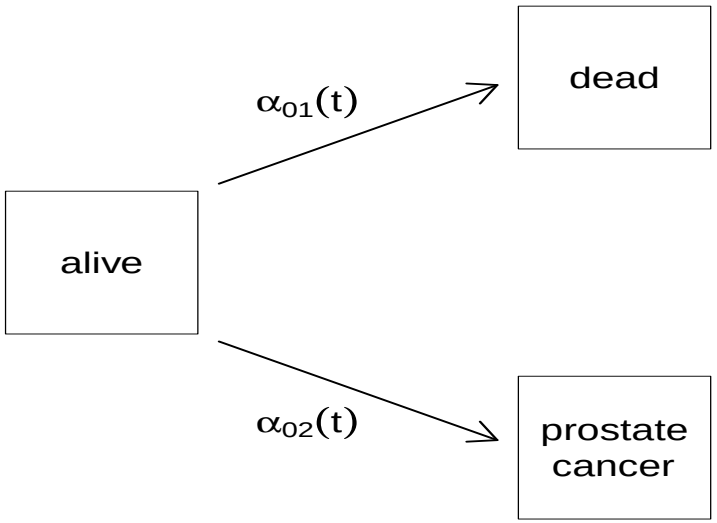
Lexis diagram - Nordic data on prostate cancer



- More than 70% are alive without cancer at follow-up.
- -also, delayed entry due to initiation of cancer registration.

Time in Twin studies

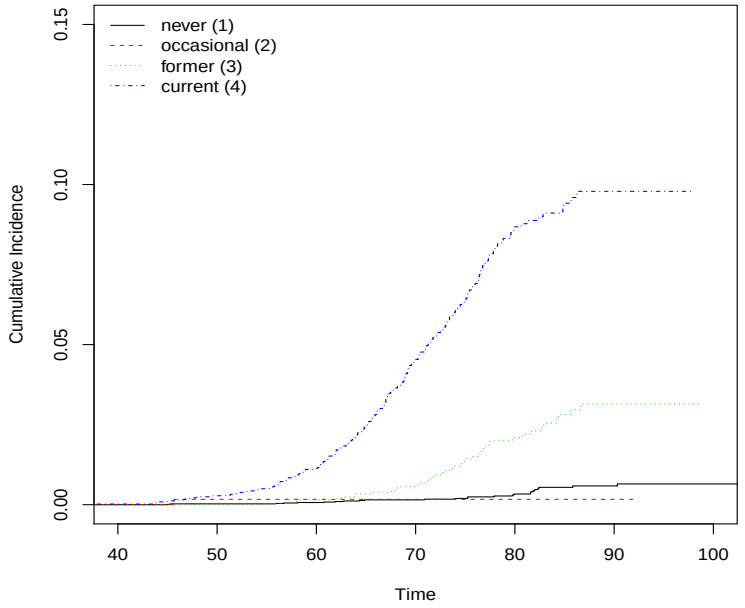
- We borrow methods from *survival analysis.*
- The Zoo: *events, censorings, competing risks,...*
- -a classic dichotomous trait is now an event.
- There may be multiple outcomes at each time point:

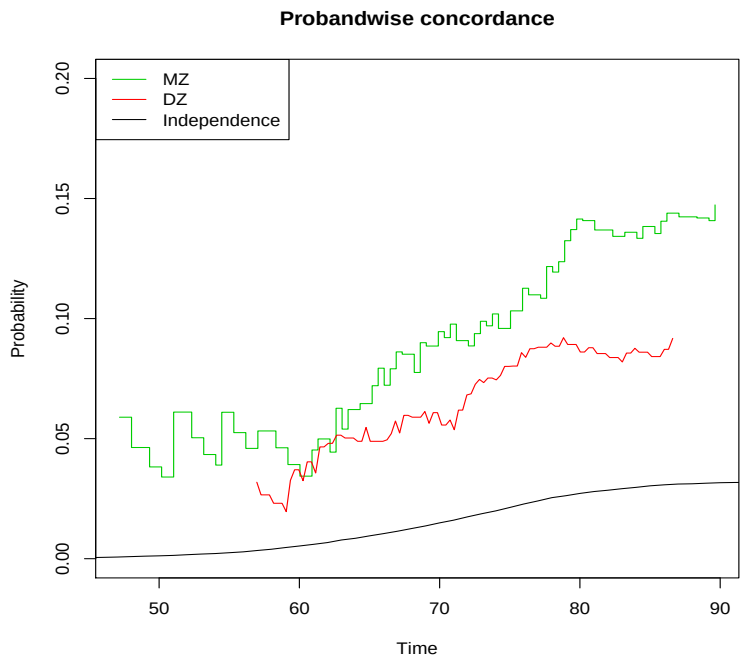


Time in Twin studies

Goals

- The cumulative incidence: *Risk of event before time t*
- The casewise concordance: *Risk of event in twin before time t given event in co-twin before time t*





Summary of sources of bias			
	prevalence	concordance	casewise
All complete data (1)	biased (low or high)	biased	biased
All data (2)	too low	too low	biased
-and modelling censorings (3)	ok	ok	ok

1. In case (1) all complete data at follow-up is used, that is, censored data is excluded.
2. In case (2) all observed data is used including censored observations at follow up, that is, censored observations are ignored.
3. In case (3) censorings and competing events (eg. death before cancer) are modelled.

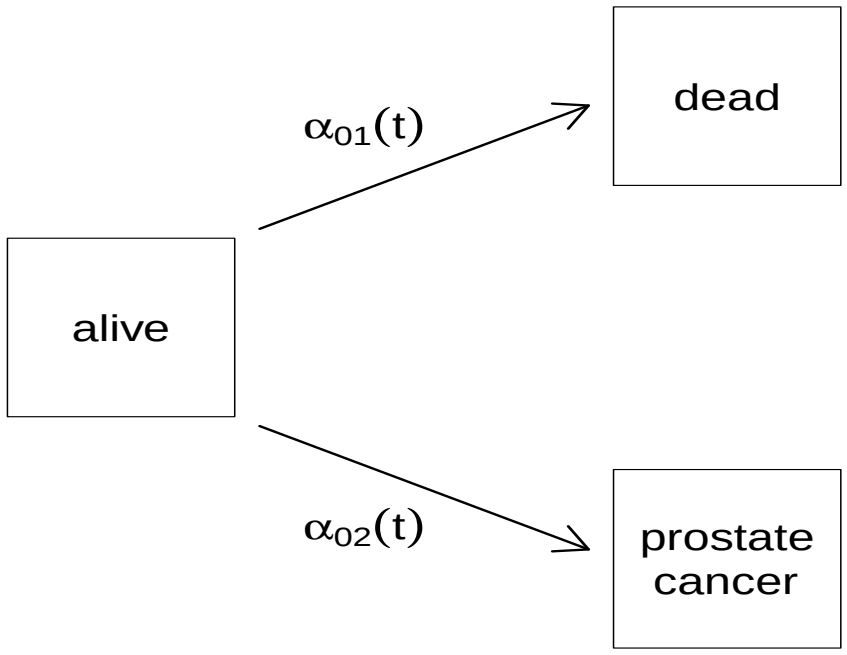
Sources of bias - breast cancer

Summary of sources of bias			
	prevalence	concordance	casewise
All complete data (1)	biased (low or high)	biased	biased
All data (2)	too low	too low	biased
-and modelling censorings (3)	ok	ok	ok

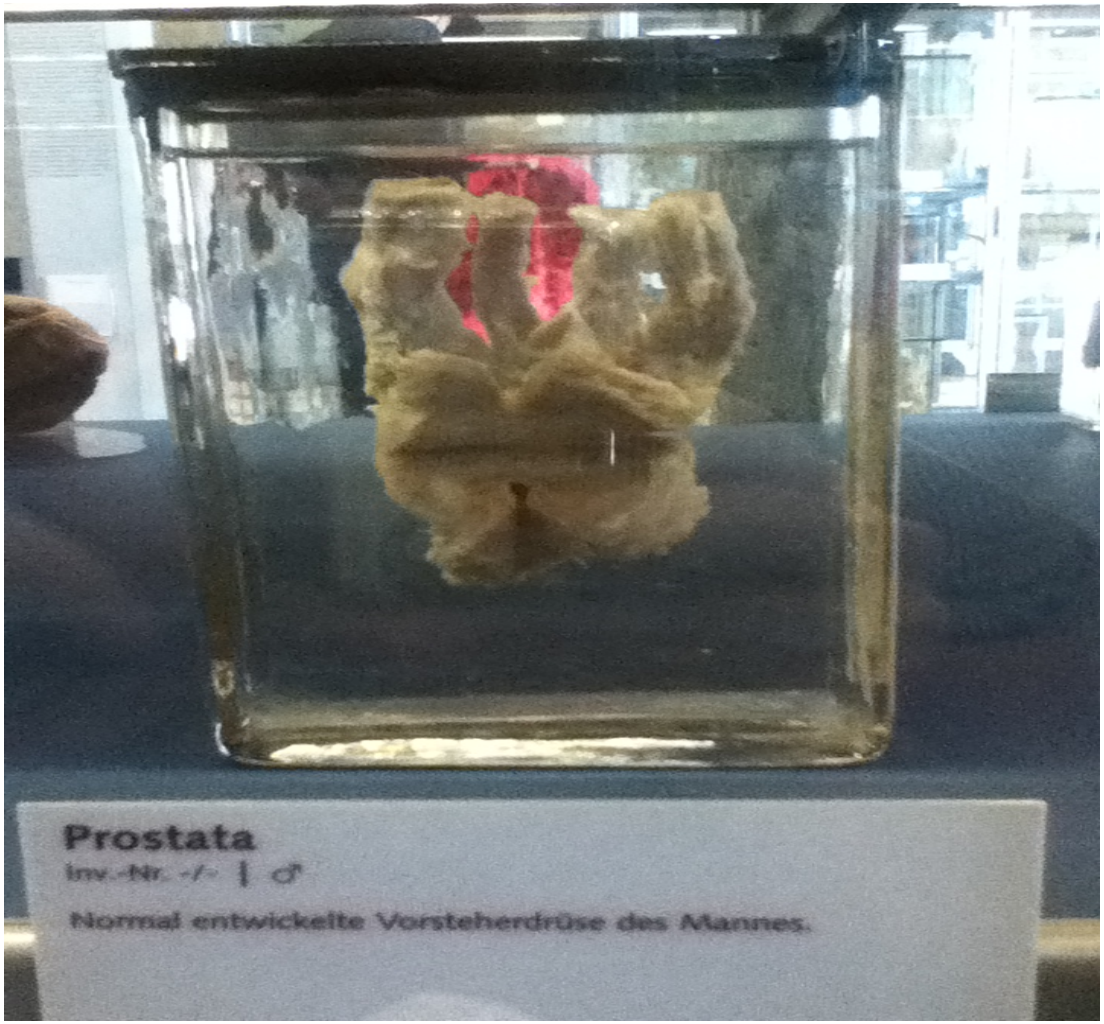
Breast cancer risk and sources of bias				
	Prevalence		casewise concordance	
	MZ twins	DZ twins	MZ twins	DZ twins
Complete data (1)	0.090 (0.005)	0.080 (0.004)	0.33 (0.04)	0.21 (0.03)
All data (2)	0.032 (0.002)	0.035 (0.001)	0.21 (0.03)	0.13 (0.02)
-and modelling censorings (3)	0.11 (0.004)	0.11 (0.004)	0.25 (0.04)	0.16 (0.03)

Methods - Competing risks

- 'the individual can experience more than one type of event?.
- 'when time to event is not independent of censoring-mechanism?.
- 'when other events precludes or interacts with event of interest?.



Example - Prostate cancer in twins



R Kioski - Package 'mets'

Prostate cancer in twins - Description


```
#Date: 2012-11-24

#Author: Klaus K. Holst, Thomas Scheike and
#        Jacob Hjelmborg
#Modified 2015-05-24

library(etm)

## Loading required package:  survival

library(prodlim)
library(mets)

## Loading required package:  timereg
## Loading required package:  lava
## lava version 1.6.1
## mets version 1.2.3.1
```

Prostate cancer in twins - Description


```
data(prt)  # simulated prostate cancer data
head(prt)

##    country     time status zyg id cancer
## 31 Denmark 96.98833      1  DZ  1      0
## 32 Denmark 80.88885      1  DZ  1      0
## 39 Denmark 68.04498      1  DZ  3      0
## 40 Denmark 61.45903      1  DZ  3      0
## 51 Denmark 78.78068      1  DZ  5      0
## 52 Denmark 90.36252      1  DZ  5      0
```

Prostate cancer in twins - Description


```
kable(with(prt, table(status,country)))
```

	Denmark	Finland	Norway	Sweden
0	7300	2533	3102	8348
1	2223	1209	876	2689
2	148	184	129	481

Prostate cancer in twins - Description


```
kable(with(prt, table(cancer,zyg)))
```

	DZ	MZ
0	17408	10872
1	583	359

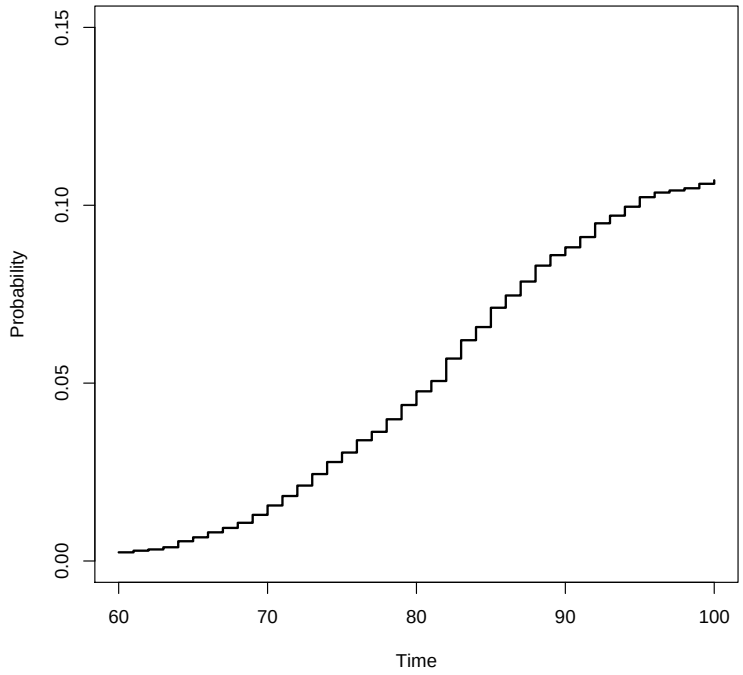
```
out <- lm(cancer~-1+zyg,prt)  # lifetime risk (!).
kable(summary(out)$coef, digits=2)
```

	Estimate	Std. Error	t value	Pr(>|t|)
zygDZ	0.03	0	24.61	0
zygMZ	0.03	0	19.18	0

Prostate cancer in twins - cumulative incidence


```
times <- seq(60,100,by=1)  # set time-range - elderly males
cifmod <- comp.risk(Event(time,status)~+1+cluster(id),data=prt,cause=2,n.sim=0,
                    times=times,conservative=1,max.clust=NULL,model="fg")
pcif <- predict(cifmod,X=1,resample.iid=0,uniform=0,se=0)
```

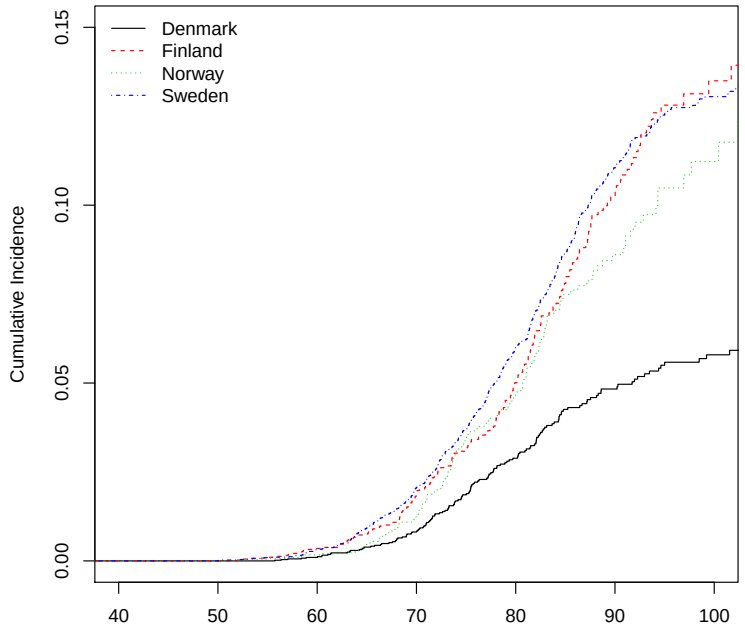
```
plot(pcif,multiple=1,se=0,uniform=0,ylim=c(0,0.15))
```



Prostate cancer in twins - cumulative incidence


```
# By non-parametric Aalen-Johansen estimator.
addprtmenetmd<-etmCIF(Surv(time,status!=0)~+factor(country),
                      data=prt,etype=status,failcode=2)
```

```
plot(addprtmenetmd,ylim=c(0,0.15),col=1:4,xlim=c(40,100),
     curvlab = c("Denmark", "Finland", "Norway","Sweden"))
```



Prostate cancer in twins - concordance


```
### ignoring country
### marginal cumulative incidence of prostate cancer##'
outm <- prodlim(Hist(time,status)~+1,data=prt)

times <- 60:100
cifmz <- predict(outm,cause=2,time=times,newdata=data.frame(zyg="MZ")) ## cause is 2 (second ca
cifdz <- predict(outm,cause=2,time=times,newdata=data.frame(zyg="DZ"))

### concordance for MZ and DZ twins
cc <- bicomprisk(Event(time,status)~strata(zyg)+id(id),data=prt,cause=c(2,2),prodlim=TRUE)


## Strata 'DZ'
## Strata 'MZ'

cdz <- cc$model$"DZ"
cmz <- cc$model$"MZ"

cdz <- casewise(cdz,outm,cause.marg=2)
cmz <- casewise(cmz,outm,cause.marg=2)
```

Prostate cancer in twins - concordance


```
plot(cmz,ci=NULL,ylim=c(0,0.6),xlim=c(60,100),legend=TRUE,col=c(3
par(new=TRUE)
plot(cdz,ci=NULL,ylim=c(0,0.6),xlim=c(60,100),legend=TRUE)
```

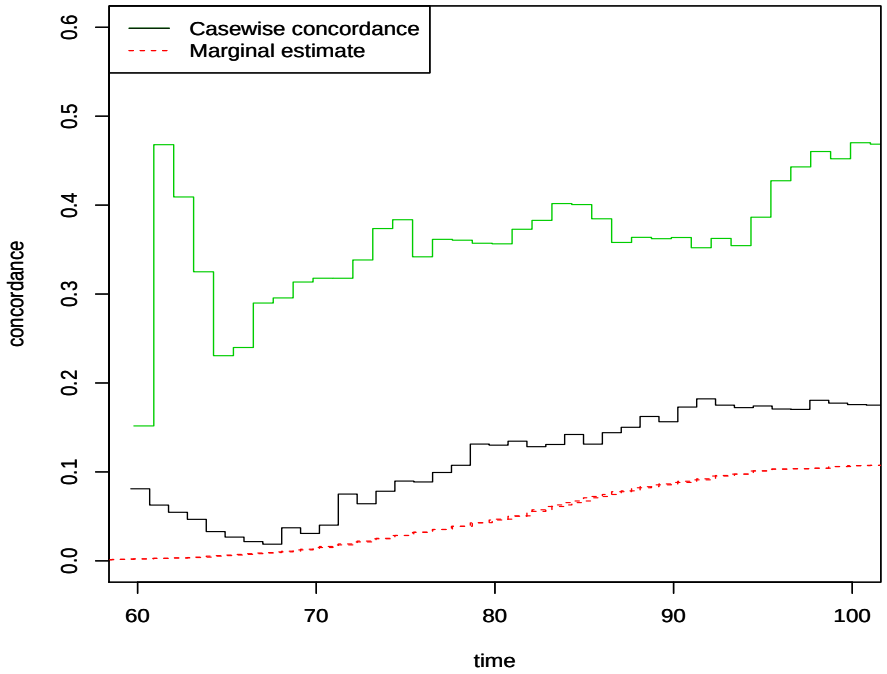


Figure: Casewise concordance

Prostate cancer in twins - Concordance

More information from

```
summary(cmz)
summary(cdz)
```

Further, Relative recurrence risk, multiple locus index and other measures can be obtained.

Time to event - biometric modeling?

The New England Journal of Medicine

VOLUME 343 JULY 13, 2000 NUMBER 2

ENVIRONMENTAL AND HERITABLE FACTORS IN THE CAUSATION OF CANCER

Analyses of Cohorts of Twins from Sweden, Denmark, and Finland

PAUL LICHTENSTEIN, PH.D., NIELS V. HOLM, M.D., PH.D., PIA K. VERKASALO, M.D., PH.D., ANASTASIA ILIADOU, M.SC., JAAKKO KAPRIO, M.D., PH.D., MARKKU KOSKENVUO, M.D., PH.D., EERO PUKKALA, PH.D., AXEL SKYTTHE, M.SC., AND KARI HEMMINKI, M.D., PH.D.

- NEJM 2000 landmark paper report heritabilities for all cancer sites.
- Prostate cancer: case-wise concordance rates (MZ; DZ) of 0.20; 0.09, and a heritability of 0.42 (0.29; 0.50).
- Biometric model: Liability threshold (ignoring censored data, ∼70%).
- Let's take censoring into account - Aim for NorTwinCan Study.

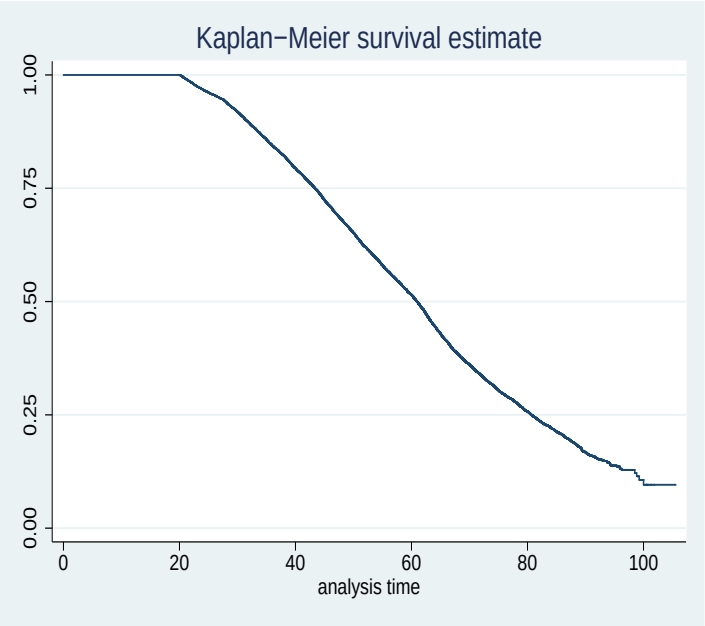
Time to event - biometric modeling - a first attempt

Genetic influence on risk scale, how about heritability?

- Liability-threshold polygenic ADCE model.:

$$\text{probit}(P(\text{twin j gets cancer}|X_j, Z)) = X_j^T\beta + Z, \quad j = 1, 2$$

- Extension: Weights from inverse probability of censoring:



Liability threshold model with IPW

- Liability model with Inverse Probability Weighting and adjusting for covariates
- Probabilities of being censored - we weight complete observations with these. In analogy with missing data analysis assuming missing at random (MAR). Probability weights based on Aalen's additive model

Liability threshold: Eq. marginals for twins


```
bp.flex <- twinlm.time(cancer~country,zyg="zyg",
    DZ="DZ",id="id",
    cumulative = TRUE, binary=TRUE,
    type="flex",data=prt,
    cens.formula=Surv(time,status==0)~1+zyg+country,
    breaks=Inf,
       control=list(refit=TRUE))
round(summary(bp.flex)$coef,2)

##                              Estimate Std.Err 2.5% 97.5%
## Tetrachoric correlation MZ       0.70    0.05 0.58  0.78
## Tetrachoric correlation DZ       0.27    0.06 0.14  0.39
```

Liability threshold: Eq. marginals for twins MZ and DZ


```
bp.u <- twinlm.time(cancer~country,zyg="zyg",
    DZ="DZ",id="id",
    cumulative = TRUE, binary=TRUE,
    type="u",data=prt,
    cens.formula=Surv(time,status==0)~1+zyg+country,
    breaks=Inf,
       control=list(refit=TRUE))

round(summary(bp.u)$coef,2)

##                              Estimate Std.Err 2.5% 97.5%
## Tetrachoric correlation MZ       0.69    0.05 0.58  0.78
## Tetrachoric correlation DZ       0.28    0.07 0.14  0.40
```

Liability threshold: Eq. marginals for twins MZ and DZ

We can compare above models directly since nested:

```
compare(bp.u,bp.flex)

##
##  - Likelihood ratio test -
##
## data:
## chisq = 19.836, df = 4, p-value = 0.000538
## sample estimates:
## log likelihood (model 1) log likelihood (model 2)
##                -8323.003                -8313.085
```

Liability threshold: ACE with IPW


```
bp.ace <- twinlm.time(cancer~country,zyg="zyg",
    DZ="DZ",id="id",
    cumulative = TRUE, binary=TRUE,
    type="ace",data=prt,
    cens.formula=Surv(time,status==0)~1+zyg+country,
    breaks=Inf,
              control=list(refit=TRUE))

score(bp.ace)

## [1]  1.089056e-04  3.706803e-05  2.993500e-05 -2.206749e-06  8

round(summary(bp.ace)$coef,2)

##                     Estimate Std.Err 2.5% 97.5%
## A                       0.67    0.05 0.58  0.77
## C                       0.00    0.00 0.00  0.00
## E                       0.33    0.05 0.23  0.42
## MZ Tetrachoric Cor      0.67    0.05 0.56  0.76
```

Liability threshold: ADE with IPW


```
bp.ade <- twinlm.time(cancer~country,zyg="zyg",
    DZ="DZ",id="id",
    cumulative = TRUE, binary=TRUE,
    type="ade",data=prt,
    cens.formula=Surv(time,status==0)~1+zyg+country,
    breaks=Inf,
       control=list(refit=TRUE))

round(summary(bp.ade)$coef,2)

##                     Estimate Std.Err  2.5% 97.5%
## A                       0.42    0.27 -0.11  0.95
## D                       0.27    0.28 -0.29  0.83
## E                       0.31    0.05  0.21  0.41
## MZ Tetrachoric Cor      0.69    0.05  0.58  0.78
## DZ Tetrachoric Cor      0.28    0.07  0.14  0.40
```

Liability threshold: ACE versus ADE

We can compare above models via the Akaike Information Index:

```
AIC(bp.ace, bp.ade)
```

```
##         df      AIC
## bp.ace   6 16662.82
## bp.ade   6 16658.01
```

Liability threshold: Stratified analysis


```
bp.ace.strata <- twinlm.time(cancer~strata(country),zyg="zyg",
    DZ="DZ",id="id",
    cumulative = TRUE, binary=TRUE,
    type="ace",data=prt,
    cens.formula=Surv(time,status==0)~1+zyg+country,
    breaks=Inf,
       control=list(refit=TRUE))

## Strata 'Denmark'
## Strata 'Finland'
## Strata 'Norway'
## Strata 'Sweden'

summary(bp.ace.strata)

## ------------------------------
## Strata 'Denmark'
## ------------------------------
## Strata 'Finland'
## ------------------------------
```

Liability threshold: Cumulative heritability


```
bp.ace.cum <- twinlm.time(cancer~country,zyg="zyg",
    DZ="DZ",id="id",
    cumulative = TRUE, binary=TRUE,
    type="ace",data=prt,
    cens.formula=Surv(time,status==0)~1+zyg+country,
    breaks=seq(60,90, by=2),
       control=list(refit=TRUE))
names(bp.ace.cum)
bp.ace.cum$summary
summary(bp.ace.cum)
```

Prostate cancer in twins - casewise concordance


```
plot(bp.ace.cum,which=c(8,13),ylim=c(0,0.5),legendpos="topright",
     col=c("darkred","darkblue"),lty=c(1,2),
     legend=c("MZ and 95% CI.","DZ and 95% CI."),
     ylab="Casewise concordance") #,    main="Nordic twin cohorts
```

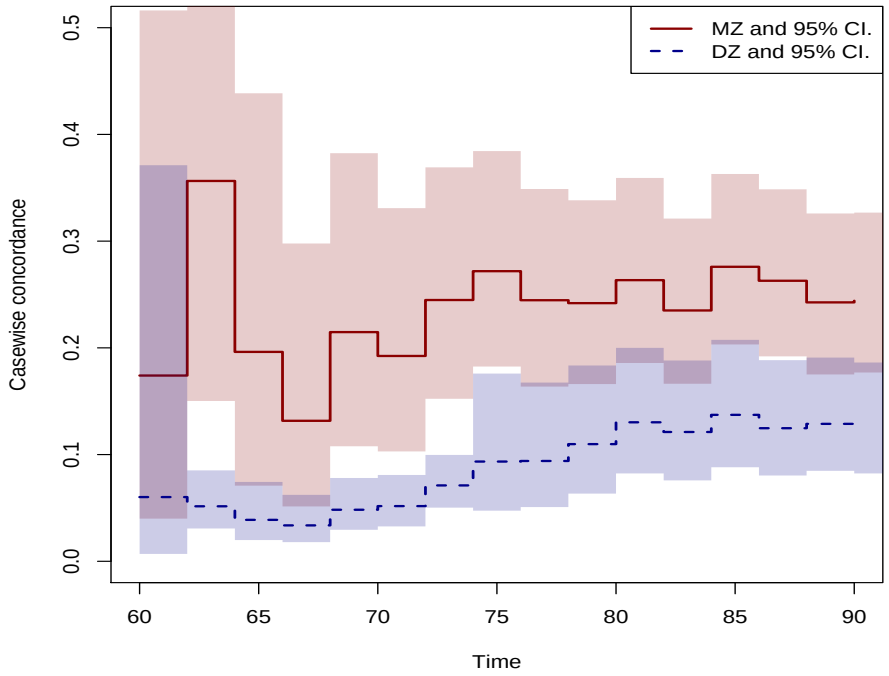


Figure: Casewise concordance

Prostate cancer in twins - heritability


```
plot(bp.ace.cum,which=c(1),ylim=c(0,1),legendpos="bottomright",
     col=c("darkred"),lty=c(1),
     legend=c("H2 and 95% CI."),
     ylab="Heritability") #,    main="Nordic twin cohorts")
```

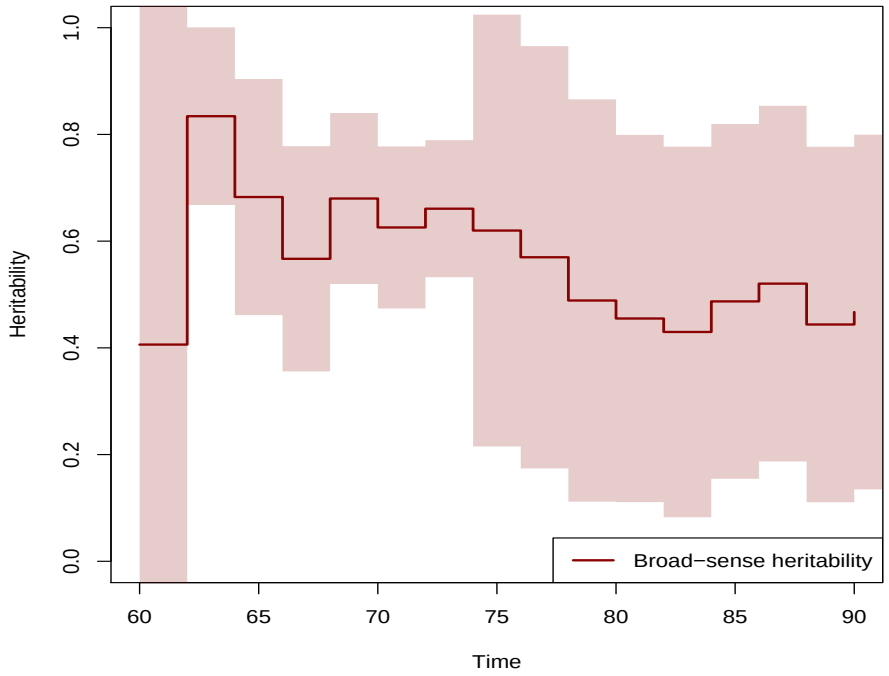


Figure: Cumulative heritability

Exercise

- Create above plots of cumulative casewise concordance and heritability from the liability threshold ADE model with IPW for censoring.
- What does the above stratified analysis add?

Exercise

- What would happen if time to event was ignored?
- This can be investigated by repeating the analysis without IPW.
- See the following slides for implementation.

Liability threshold: Saturated model - ignoring time


```
bp0 <- biprobit(cancer~country + cluster(id)+strata(zyg),
                data=prt)


## Strata 'DZ'
## Strata 'MZ'

summary(bp0)

## ------------------------------
## Strata 'DZ'
## ------------------------------
## Strata 'MZ'
```

Liability threshold: Eq. marginals - ignoring time


```
bp1 <- bptwin(cancer~country,zyg="zyg",DZ="DZ",id="id", binary=TRUE, type="u",data=prt)
summary(bp1)$probMZ
```

```
##                          Estimate          2.5%        97.5%
## Concordance           0.004467324  0.003292577  0.006058658
## Casewise Concordance  0.293233128  0.234208848  0.360136998
## Marginal              0.015234718  0.012860807  0.018038809
## Rel.Recur.Risk       19.247690103 14.645257727 23.850122480
## log(OR)               3.625153734  3.234286445  4.016021024
```

```
summary(bp1)$probDZ
```

```
##                          Estimate          2.5%       97.5%
## Concordance           0.001440254 0.0009503536 0.002182143
## Casewise Concordance  0.094537629 0.0667986510 0.132164199
## Marginal              0.015234718 0.0128608072 0.018038809
## Rel.Recur.Risk        6.205407284 4.0723812102 8.338433358
## log(OR)               1.994581307 1.5853517141 2.403810901
```

```
summary(bp1)$coef
```

```
##                              Estimate    Std.Err      2.5%     97.5%
## Tetrachoric correlation MZ 0.6988528 0.03375873 0.6265551 0.7592258
## Tetrachoric correlation DZ 0.3706259 0.04339034 0.2826528 0.4524161
```

```
compare(bp0,bp1) # LRT
```

Liability threshold: ACE model - ignoring time


```
bp2 <- bptwin(cancer~country,zyg="zyg",DZ="DZ",id="id", binary=TRUE,type="ace",data=prt)
summary(bp2)$probMZ
```

```
##                          Estimate         2.5%        97.5%
## Concordance           0.004467383  0.003292626  0.006058727
## Casewise Concordance  0.293234795  0.234210425  0.360138679
## Marginal              0.015234832  0.012860918  0.018038923
## Rel.Recur.Risk       19.247655569 14.645254717 23.850056421
## log(OR)               3.625156485  3.234289704  4.016023266
```

```
summary(bp2)$probDZ
```

```
##                          Estimate         2.5%       97.5%
## Concordance           0.00144021 0.0009503172 0.002182094
## Casewise Concordance 0.09453405 0.0667954867 0.132160421
## Marginal              0.01523483 0.0128609183 0.018038923
## Rel.Recur.Risk        6.20512574 4.0721466620 8.338104827
## log(OR)               1.99452774 1.5852906674 2.403764821
```

```
summary(bp2)$coef
```

```
##                         Estimate    Std.Err       2.5%     97.5%
## A                     0.65647764 0.10956971  0.4417250 0.8712303
## C                     0.04237639 0.09289080 -0.1396862 0.2244390
## E                     0.30114597 0.03375863  0.2349803 0.3673117
## MZ Tetrachoric Cor 0.69885403 0.03375863  0.6265565 0.7592268
## DZ Tetrachoric Cor 0.37061521 0.04339116  0.2826405 0.4524070
```

Analysis cookbook

- Empirical descriptives: Counts, prevalences and cumulative incidences.
- Factors for cumulative incidence? Intermediate or confounder? Include covariates or stratification may be necessary. Univariate analysis using twins: logistic regression correcting for within pair dependence by robust variance estimation. Alternative more elaborate approach: Covariates may be tested influential on thresholds in liability threshold model below. Survival model to take censoring into account. Competing risks?
- Similarity measures: Concordance rate and polychoric correlation over time.
- Biometric modelling: Liability threshold model with inverse probability weightning (if censoring).
- same threshold for twin 1 and twin 2.
- same threshold for MZ and DZ twins.
- Polychoric correlation estimation.
- Polygenic best fitting model: Most likely model with fewest parameters (parsimony).
- Conclusion: Familial risks by time, heritabilities, pleiotropy,...

The Heritability of Prostate Cancer in the Nordic Twin Study of Cancer Hjelmborg, Scheike, Kaprio, Mucci et al.
`http://cebp.aacrjournals.org/content/23/11/2303`
Cancer Epidemiology, Biomarkers & Prevention (2014)

References

- *Estimating heritability for cause specific mortality based on twin studies*; Scheike, Holst and Hjelmborg; *LIDA* (2013).
- *Estimating twin concordance for bivariate competing risks twin data*; Scheike, Holst and Hjelmborg; *Stat Med* (2014)
- *Measuring early or late dependence for bivariate lifetimes of twins* Scheike, Holst and Hjelmborg; *LIDA* (2014).
- *Revisiting the Concordance for Twin Pairs*; Hjelmborg, Scheike, Holst and Möller; *Hum Genet Twin Research* (2015), in preparation
- *The liability threshold model for censored twin data* Holst, Scheike and Hjelmborg, *Computational Statistics & Data Analysis* (2015)