The Extended Liability-Threshold Model

Klaus K. Holst
<k.k.holst@biostat.ku.dk>
Thomas Scheike, Jacob Hjelmborg

2014-05-21

Heritability

Twin studies

Include both monozygotic (MZ) and dizygotic (DZ) twin pairs.
- DZ pairs on averages shares half of their genes
- MZ pairs are natural copies

Difference in similarity of DZ and MZ twins may indicate genetic influence!

Decomposition

What is contribution of genetic and environmental factors to the variation in the outcome?
The phenotype is the sum of genetic and environmental effects:

\[ Y = G + E \]
\[ \Sigma Y = \Sigma G + \Sigma E \]

Polygenic model for continuous trait

ACDE model

Decompose outcome into

\[ Y_i = A_i + D_i + C_i + E_i, \quad i = 1, 2 \]

- \( A \): Additive genetic effects of alleles
- \( D \): Dominant genetic effects of alleles
- \( C \): Shared environmental effects
- \( E \): Unique environmental effects

Assumptions
- No gene-environment interaction
- No gene-gene interaction (ACE)
- Same marginals of twin 1 and twin 2, and MZ and DZ. Equal environmental effects for MZ and DZ.

Model

\[ Y_i = A_i + C_i + D_i + E_i \]

\[ A_i \sim \mathcal{N}(0, \sigma_A^2), \quad C_i \sim \mathcal{N}(0, \sigma_C^2), \quad D_i \sim \mathcal{N}(0, \sigma_D^2), \quad E_i \sim \mathcal{N}(0, \sigma_E^2) \]

\[ \text{Cov}(Y_1, Y_2) = \begin{pmatrix} \sigma_A^2 & Z_A \sigma_A^2 \\ Z_A \sigma_A^2 & \sigma_A^2 \end{pmatrix} + \begin{pmatrix} \sigma_C^2 & \sigma_C^2 \\ \sigma_C^2 & \sigma_C^2 \end{pmatrix} + \begin{pmatrix} \sigma_D^2 & Z_D \sigma_D^2 \\ Z_D \sigma_D^2 & \sigma_D^2 \end{pmatrix} + \begin{pmatrix} \sigma_E^2 & 0 \\ 0 & \sigma_E^2 \end{pmatrix} \]

where

- \( Z_A = \begin{cases} 1, & \text{MZ} \\ 0.5, & \text{DZ} \end{cases} \)
- \( Z_D = \begin{cases} 1, & \text{MZ} \\ 0.25, & \text{DZ} \end{cases} \)
**Polygenic model**

- \( DZ \), 0.5/1
- \( MZ \), 0.25/1

\[ A_1, D_1, E_1, C, A_2, D_2, E_2 \]

\( Y_1 \), \( Y_2 \)

\( X \), \( Z \)

\( \lambda_A \), \( \lambda_D \), \( \lambda_E \), \( \lambda_C \)

\[ Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & Y_i^* \leq 0 \end{cases} \]

\[ Y_i^* = \beta^T X_i + A_i + C_i + E_i, \quad \text{Var}(E_i) = 1. \]

**Heritability**

(Broad-sense) Heritability

\[ h_Y^2 = \frac{\text{Var}(G)}{\text{Var}(Y)} = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2} \]

Shared environmental effect

\[ c_Y^2 = \frac{\sigma_C^2}{\sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2} \]

In the ACE model the heritability is given by

\[ h^2 = 2(\rho_{MZ} - \rho_{DZ}) \]

**Liability model/Threshold model for binary data**

For the dichotomous prostate cancer status outcome (cancer or death without cancer) we can use a Liability Model

Let \( Y_1 \) and \( Y_2 \) be cancer status of the two twins. Model based on Probit link:

\[ \mathbb{P}(Y_1 = 1, Y_2 = 1 \mid X) = \Phi(\beta^T X_1 + A_1 + C_1, \beta^T X_2 + A_2 + C_2) \]

where \( \Phi \) is bivariate standard normal CDF, e.g.

\[ Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & Y_i^* \leq 0 \end{cases} \]

\[ Y_i^* = \beta^T X_i + A_i + C_i + E_i, \quad \text{Var}(E_i) = 1. \]
Liability model/Threshold model for binary data

\[ P(Y_1 = 1, Y_2 = 1) = P(Y^*_1 > \tau_1, Y^*_2 > \tau_2) \]

Liability model in R

The bivariate probit model, biprobit, and Liability model, bptwin, is available in R via the mets package.

```r
> library(mets)
> data(twinstut)
> twinstut <- subset(twinstut, zyg != "os")
> biprobit(stutter ~ sex + strata(zyg), data=twinstut, id="tvparnr")
> bptwin(stutter ~ sex + strata(zyg), data=twinstut, id="tvparnr", zyg="zyg", DZ="dz")
```

Statistical inference via compare (LRT, Wald test), confint, AIC, ...

Heritability of prostate cancer

Considerable interest in quantifying the genetic influence of prostate cancer


Reported case-wise concordance rates (MZ; DZ) of 0.20; 0.09, and a heritability of 0.42 (0.29; 0.50).
Prostate cancer - Danish cohorts

Table 1: Lifetime disease prevalence and lifetime concordance for MZ twins estimated using Liability threshold model ignoring censoring for different censorings patterns of MZ and DZ twins. Cross odds-ratio dependence parameter 3 and 2 for MZ and DZ, respectively.

<table>
<thead>
<tr>
<th>MZ &amp; DZ Status</th>
<th>Prostate cancer</th>
<th>No cancer, dead</th>
<th>No cancer, alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate cancer</td>
<td>26 &amp; 15</td>
<td>220</td>
<td>101</td>
</tr>
<tr>
<td>No cancer and dead</td>
<td>100</td>
<td>1224 &amp; 2314</td>
<td>1392</td>
</tr>
<tr>
<td>No cancer and alive</td>
<td>38</td>
<td>524</td>
<td>4024 &amp; 6884</td>
</tr>
</tbody>
</table>

Table: Number of pairs by status at time of follow-up with MZ pairs in lower left triangle (colored red) and DZ pairs in upper right triangle.

Massive amount of censoring which cannot be ignored!!!

Simulation, Scheike et al 2013, Lifetime Data An.

Table 2: Estimates of variance for genetic (A) and environmental components (C) of threshold liability model given different censorings patterns of MZ and DZ twins. Cross odds-ratio dependence parameter 3 and 1.5 for MZ and DZ, respectively.

<table>
<thead>
<tr>
<th>MZ &amp; DZ Status</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.374</td>
<td>0.000</td>
</tr>
<tr>
<td>21%</td>
<td>0.473</td>
<td>0.019</td>
</tr>
<tr>
<td>55%</td>
<td>0.627</td>
<td>0.198</td>
</tr>
<tr>
<td>82%</td>
<td>0.988</td>
<td>0.257</td>
</tr>
</tbody>
</table>

Modeling time to first prostate cancer

- Hazard scale model?
- Concordance scale?
- AFT/Tobit?

We observe the minimum of the failure time $T_i^*$ and the censoring time $C_i$

$$T_i = \min(T_i^*, C_i)$$

$$g(T_i^*) = \beta^T X_i + A_i + C_i + E_i$$
Multi-state model

Tobit framework complicated by competing risks

Dead

Alive

Prostate cancer

Estimating equations for the Liability model

Define indicator of an observation being an actual event time

\[ \Delta_{ij} = I \{ T^*_{ij} \leq C_i \}, \]

for twin \( j = 1, 2 \) of the \( i \)th twin-pair, and with same random censoring \( C_i \) (independent of \( T^*_{ij} \) given possible covariates) for the twin-pair.

Model for censoring mechanism

\[ G_c(t; Z_i) = \mathbb{P}(C_i > t \mid Z_i) \]

Define IPCW Estimating Equation based on complete pairs only

\[ U_w(\theta) = \sum_{i=1}^{n_c} \Delta_{i1} \Delta_{i2} \min \{ G_c(T_{i1}; Z_i), G_c(T_{i2}; Z_i) \} U_i(\theta) \]

OBS: positivity assumption!

Estimating equations for the Liability model

Nice properties of the Probit model gives us closed forms of the marginals of the twin-pair observation \( Y = (Y_1, Y_2)^T \) (up to evaluation of bivariate CDF).

\[
U(\theta) = U(\theta; Y) \\
= \frac{1}{\Phi_{\mu_\theta,\Sigma_\theta(0)}} \frac{1}{2} \left( \frac{\partial \text{vec } \Sigma_\theta}{\partial \theta^T} \right)^T \left[ -\text{vec}(\Sigma_\theta^{-1}) \Phi_{\mu_\theta,\Sigma_\theta(0)} + (\Sigma_\theta^{-1} \otimes \Sigma_\theta^{-1}) \text{vec}\{ V_{\mu_\theta,\Sigma_\theta(0)} \} \right] + \left( \frac{\partial \text{vec } \mu_\theta}{\partial \theta^T} \right)^T \Sigma_\theta^{-1} \text{vec}\{ M_{\mu_\theta,\Sigma_\theta(0)} \} \\
L = \text{diag}\{2Y_1 - 1, 2Y_2 - 1\} \\
\Sigma_\theta = \Sigma_\theta(Y) = L \Sigma_\theta^0 \Sigma_\theta^0, \quad \mu_\theta = \mu_\theta(Y) = L \mu_\theta^0
\]

Weights based on parametric survival model, stratified Kaplan-Meier, Cox Proportional Hazards Model, Aalen’s Additive Model, ...

With correct model for the weights we obtain consistency

\[ \mathbb{E}(U_w(\theta)) = \mathbb{E}\{ \mathbb{E}(U_w(\theta) \mid Z, X) \} = \mathbb{E}(U_i(\theta)) = 0 \]
Estimating equations for the Liability model

Under mild regularity conditions the estimates, \( \hat{\theta} \), from solving \( U^w \) are consistent and asymptotically normal following from i.i.d. decomposition

\[
\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i + o_p(1)
\]

and asymptotic variance (sandwich type)

\[
\frac{1}{n} \sum_{i=1}^{n} \epsilon_i \otimes \epsilon_i
\]

Estimates of concordance, tetrachoric correlation, heritability, etc. can be obtained applying the Delta theorem (using variance stabilizing transformations \( \tanh \), \( \logit \), ...)

Implementation

```r
> library(mets)
> dw <- ipw(Surv(time,status==0)~strata(country),
          data=prostatedata)
> a <- bptwin(cancer ~ country, data=dw,
             id="tvparnr", zyg="zyg", DZ="DZ",
             weight="w", type="ace")
> summary(a)

Estimate Std.Err Z p-value
(Intercept) -2.372835 0.094033 -25.233941 0.0000
landFinland 0.605013 0.067857 8.915952 0.0000
landNorway 0.664485 0.070300 9.452122 0.0000
landSweden 0.557265 0.052939 10.526493 0.0000
log(var(A)) 0.260261 0.127288 2.044662 0.0409
log(var(C)) -26.669063 6.054746 -4.404655 0.0000

Total MZ/DZ Complete pairs MZ/DZ
8585/17115 3161/5894

... ...
```

Implementation, type="ace"

<table>
<thead>
<tr>
<th>Estimate 2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.56470 0.50270 0.62475</td>
</tr>
<tr>
<td>C</td>
<td>0.00000 0.00000 0.00000</td>
</tr>
<tr>
<td>E</td>
<td>0.43530 0.37525 0.49730</td>
</tr>
<tr>
<td>Correlation MZ</td>
<td>0.56470 0.50027 0.62291</td>
</tr>
<tr>
<td>Correlation DZ</td>
<td>0.28235 0.25141 0.31272</td>
</tr>
<tr>
<td>MZ: Estimate 2.5%</td>
<td>97.5%</td>
</tr>
<tr>
<td>Concordance</td>
<td>0.01774 0.01446 0.02174</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.30209 0.26012 0.34766</td>
</tr>
<tr>
<td>Marginal</td>
<td>0.05873 0.05272 0.06538</td>
</tr>
<tr>
<td>DZ: Estimate 2.5%</td>
<td>97.5%</td>
</tr>
<tr>
<td>Concordance</td>
<td>0.00878 0.00722 0.01068</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.14954 0.13389 0.16666</td>
</tr>
<tr>
<td>Marginal</td>
<td>0.05873 0.05272 0.06538</td>
</tr>
<tr>
<td>Heritability</td>
<td>0.56470 0.50270 0.62475</td>
</tr>
</tbody>
</table>

Comparison with competing risks model (ref:SE)

Cumulative Incidence Function (MZ)

![Cumulative Incidence Function (MZ)](image)

Time

Probability

- IPW
- Naive
- bicomprisk

<table>
<thead>
<tr>
<th>Time</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPW</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
<td>0.20</td>
</tr>
<tr>
<td>Naive</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
<td>0.20</td>
</tr>
<tr>
<td>bicomprisk</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
<td>0.20</td>
</tr>
</tbody>
</table>
Comparison with bivariate competing risks model

**Implementation. type="cor"**

<table>
<thead>
<tr>
<th></th>
<th>Estimate 2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation MZ</td>
<td>0.56718</td>
<td>0.49546 0.63121</td>
</tr>
<tr>
<td>Correlation DZ</td>
<td>0.27713</td>
<td>0.20442 0.34680</td>
</tr>
</tbody>
</table>

**MZ:**

<table>
<thead>
<tr>
<th></th>
<th>Estimate 2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance</td>
<td>0.01784</td>
<td>0.01441 0.02207</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.30378</td>
<td>0.25774 0.35413</td>
</tr>
<tr>
<td>Marginal</td>
<td>0.05872</td>
<td>0.05271 0.06537</td>
</tr>
</tbody>
</table>

**DZ:**

<table>
<thead>
<tr>
<th></th>
<th>Estimate 2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance</td>
<td>0.00865</td>
<td>0.00661 0.01132</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.14736</td>
<td>0.11868 0.18154</td>
</tr>
<tr>
<td>Marginal</td>
<td>0.05872</td>
<td>0.05271 0.06537</td>
</tr>
</tbody>
</table>

Heritability 0.58009 0.38028 0.75669

**Implementation. type="flex"**

<table>
<thead>
<tr>
<th></th>
<th>Estimate 2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation MZ</td>
<td>0.57110</td>
<td>0.49992 0.63462</td>
</tr>
<tr>
<td>Correlation DZ</td>
<td>0.27485</td>
<td>0.20253 0.34419</td>
</tr>
</tbody>
</table>

**MZ:**

<table>
<thead>
<tr>
<th></th>
<th>Estimate 2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance</td>
<td>0.01854</td>
<td>0.01369 0.02507</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.30886</td>
<td>0.25980 0.36266</td>
</tr>
<tr>
<td>Marginal</td>
<td>0.06003</td>
<td>0.04953 0.07258</td>
</tr>
</tbody>
</table>

**DZ:**

<table>
<thead>
<tr>
<th></th>
<th>Estimate 2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance</td>
<td>0.00845</td>
<td>0.00629 0.01136</td>
</tr>
<tr>
<td>Conditional</td>
<td>0.14546</td>
<td>0.11645 0.18022</td>
</tr>
<tr>
<td>Marginal</td>
<td>0.05811</td>
<td>0.05099 0.06615</td>
</tr>
</tbody>
</table>

Heritability 0.59250 0.39270 0.76577

**R Implementation**

```r
> a <- bptwin(cancer ~ strata(country), data=dw,  
  id="tvparnr", zyg="zyg", DZ="DZ",  
  weight="w", type="cor")
```

**Stratified analysis**

```r
> mean(score(a)^2) ## Close to zero?  
> a$opt ## Messages from optimization routine  
> a <- bptwin(..., control=list(trace=1,iter.max=100,  
  start=mystart,grtol=1e-10,...))
```
Hypothesis testing

Estimator is not MLE and usual likelihood ratio test are no longer an option.
Instead general linear hypotheses of the form

\[ H_0: \theta = \hat{\theta}_0 \]  

(1)

can be tested using a Wald test

\[
(B\hat{\theta} - \hat{\theta}_0)^T(B\Sigma_\theta B^T)^{-1}(B\hat{\theta} - \hat{\theta}_0) \sim \chi^2_{\text{rank}(B)}
\]

(2)

In R we can specify the contrast matrix \( B \) (see matrix, cbind, diag) and use the compare function

\[
> \text{compare}(a, \text{contrast}=B, \text{null}=b0)
\]

\[
> \text{compare}(a, \text{par}) \quad \# \text{ Index of parameters to test } b[i]=0
\]

Cumulative Heritability

Introducing time again...

\[
\Pr(Y_1 = 1, Y_2 = 1, T_1 \leq \tau, T_2 \leq \tau) = \Phi(A_1^T + C_1^T, A_2^T + C_2^T)
\]

\[
> \text{twinlm.time(cancer~1, zyg="zyg", DZ="DZ", id="id", type="ace", data=prt, cens.formula=Surv(time,status==0)~zyg, breaks=c(70,100,by=4))}
\]

Prostate Cancer Occurrence, Danish Twin Cohorte

Age

Probability

\( F_1(t) \)

\( C_{MZ}(t) \)

\( C_{DZ}(t) \)

\( F_2\)

Relative recurrence risk

\( MZ \)

\( DZ \)
Prostate Cancer Occurrence, Danish Twin Cohorte

Heritability under the presence of censoring

- Censoring in cancer registers cannot be ignored
- The IPCW Liability model corresponds to the bivariate competing risks model evaluated at infinity
- Extends the classical polygenic model for dichotomous endpoints (Liability model) to model with missing data due to right censoring
- Estimation straightforward in R
- Time effects can be pragmatically modeled using the cumulative heritability
- Limitations: Interpretation on the liability scale; lack of identification of full model; delayed entry
- Estimating equations could be based on the non-parametric same cens. concordance estimator