ROeS Dornbirn 2013

Max-test to evaluate genetic association studies for continuous and time-to-event traits

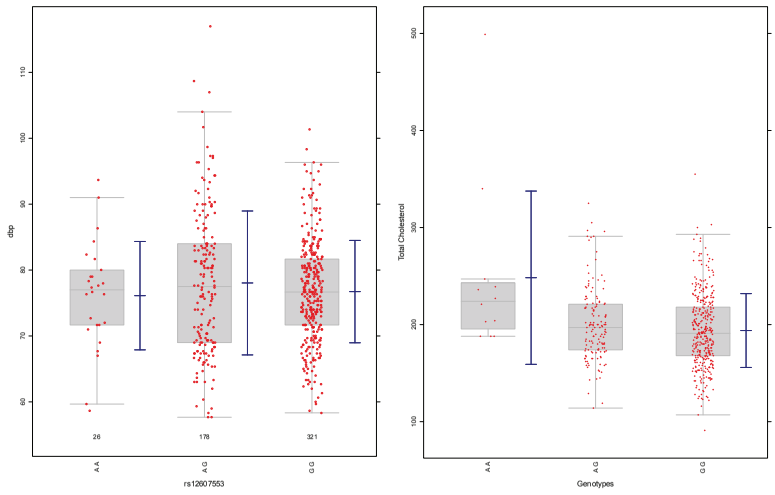
Ludwig A. Hothorn
Leibniz University Hannover
hothorn@biostat.uni-hannover.de
Joint work with E. Herberich (Ingelheim) and F. Konietschke (Goe)

August 29, 2013

A motivating example I

- The majority of genetic association studies used population-based recruitment and a case-control design, i.e. diseased vs. healthy subjects
- Today, I will focus on cohort design with continuous endpoints, i.e. phenotypes as dbp, ...(quantitative traits)
- Particular challenging traits are scores in psychiatric studies and time-to-event outcomes
- Motivating example: re-analyzing the Bogalusa Heart Study [SCK10]. From the longitudinal study, 12 selected clinical endpoints (at study end) were used from N=525 individuals together with 545,821 SNPs
- Two examples are discussed more in detail in the following for the phenotype *diastolic blood pressure* (SNP rs12607553) and *total cholesterol* on *SNP rs7738656*

A motivating example II



The problem I

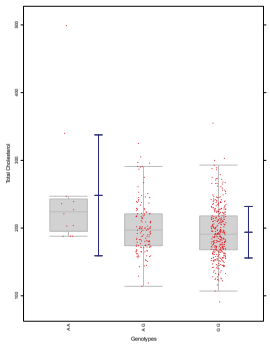
- Simplified: per-SNP consideration (ignoring many SNPs, their correlation or interaction)
- Simplified: one selected phenotype *total cholesterol* and one selected SNP *rs7738656* in the gene C6orf170/GJA1
- Simplified design: one-way layout with the 3 qualitative levels, i.e. **genotypes AA, AG, GG:**
  Homozygote risk allele,
  heterozygote allele,
  homozygote non-risk allele
  Here, AA is high risk genotype. No covariates (population stratification, subject characteristics)
- Seems to be rather simple ...

The problem II

- Common analysis in genetic papers (PLINK-style):
  1. Common linear regression using $0, 0.5, 1$ allele scores assuming an additive mode. How realistic, how robust?
  2. ANOVA-F-test for any heterogeneity
  3. Maximum test on three mode-of-inheritance-specific tests: additive, recessive, dominant.
     First introduce by [FZLG02] for 2-by-3 table data as min-p test.
  4. Notice, inherently an one-sided test problem, but performed two-sided, because the risk allele is a-priori unknown in each of the 545,821 SNPs
- But, serious assumptions:
  - i the validity of the additive mode of inheritance ... questionable
  - ii normally distributed errors ... rare
  - iii homogeneous variances ... not per definition
  - iv AND, the robustness of standard test is limited by the rather unbalanced designs: high risk allele is sometimes extreme rare

The problem III

- Query: can be assume i)-iv) in our example?



- Answer: no

The problem IV

- A general approach: **multiple contrast test** in **GLM** providing **simultaneous confidence intervals** using **R packages**

  - The null hypothesis $H_0 : \mu_{GG} = \mu_{GA} = \mu_{AA}$ can be tested against three types of H_1:

    i Any heterogeneity $H_0^{het} : \mu_i \neq \mu_j i, j \in \{AA, AG, GG\}$
    ii Just an additive mode of inheritance $H_1^{add} : \mu_{GG} < \mu_{GA} < \mu_{AA}$
    iii The most likely out of the three main mode of inheritances

    $$H_1^{add} : \mu_{GG} < \mu_{GA} < \mu_{AA}$$

    $$H_1^{dom} : \mu_{GG} < \mu_{GA} = \mu_{AA}$$

    $$H_1^{rec} : \mu_{GG} = \mu_{GA} < \mu_{AA}$$

    I.e. order restricted alternatives

  - A **contrast** is a suitable linear combination of means:

    $$\sum_{i=0}^{k} c_i \bar{x}_i$$

The problem V

- A **contrast test** is standardized:

  $$t_{Contrast} = \sum_{i=0}^{k} c_i \bar{x}_i / S \sqrt{\sum_{i}^{k} c_i^2 / n_i}$$

  where $\sum_{i=0}^{k} c_i = 0$ guaranteed a $t_{df,1-\alpha}$ distributed level-α-test and to achieve compatible sCIs. To guarantee comparable simultaneous confidence intervals is needed: $\sum sign^+(c_j) = 1, \sum sign^-(c_j) = 1$

- A **multiple contrast test is defined as maximum test**:

  $$t_{MCT} = max(t_1, ..., t_q)$$

  which follows jointly $(t_1, \ldots, t_q)'$ a q-variate t- distribution with degree of freedom df and the correlation matrix R.
  Notice, to use Bonferroni (i.e. $t_{df,1-\alpha/3}$) is not a powerful approach, since the 3 contrasts are highly correlated, at least each 2

The problem VI

- Now, just the choice of a particular contrast matrix defines the MCT (some in the literature denoted as MCP), e.g.
- Dunnett one-sided [Dun55]

c_i	C	T_1	T_2
c_a	-1	0	1
c_b	-1	-1	0

- Association max test for an unbalanced design

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}'_{dom} \\ \mathbf{c}'_{add} \\ \mathbf{c}'_{rec} \end{pmatrix} = \begin{pmatrix} -1 & \frac{n_2}{n_2+n_3} & \frac{n_3}{n_2+n_3} \\ -1 & 0 & 1 \\ -\frac{n_1}{n_1+n_2} & -\frac{n_2}{n_1+n_2} & 1 \end{pmatrix}$$

- **One-sided (lower) simultaneous confidence limits**:
  $[\sum_{i=0}^{k} c_i \bar{x}_i - St_{q,df,R,2-sided,1-\alpha} \sqrt{\sum_i^k c_i^2/n_i}]$

The problem VII

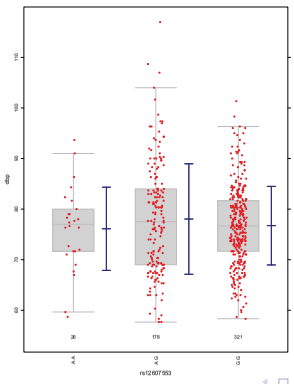
- Modification for unbalanced, heteroscedastic data [Has08]

$$S^{2*} = \frac{\omega_i^2}{n_0} S_0^2 + \sum_{h=q+1-i}^{q} \frac{n_h}{\widetilde{n_i^2}} S_h^2.$$

  Approximate multi-t-distributed with Satterthwaite-type ν, whereas R depend on the unknown variances σ_i^2
  Plug-in modification: *sci.ratioVH* function in the R package *mratios*

Evaluation of the spb-example I

- The association between the phenotype *systolic blood pressure* and the SNP *rs726914* is characterized by
  1. a symmetric distribution
  2. an unbalanced design, even not too unbalanced: 26/178/321
  3. heterogeneous variances (p-value of the Levene test 0.01)



Evaluation of the spb-example II

- Multiple contrast tests modified for heterogeneous variances (and adjusted against covariates sex, weight and age):

Mode of inherit.	Mean differ. / *mm Hg*	Sim. confid. interval / *mm Hg*	adj. p-value
Recessive	2.4	[-0.7; 5.4]	0.15
Additive	4.4	[0.8; 7.5]	0.012
Dominant	3.6	[1.3; 5.8]	0.00092

- Most likely: dominant mode, since the lower CL of 1.3 mm Hg is most distant to zero of H_0 (or reveals the smallest adjusted p-value of 0.00092).
- Question: how clinically relevant is an increase of at least 1.3 mm Hg systolic blood pressure caused by SNP *rs726914*?

A non-parametric approach I

- Focus on appropriate effect sizes, e.g.
  - i OR for a case-control study
  - ii $\mu_i - \mu_j$ and μ_i/μ_j for normal distributed traits
  - iii relative effect size
    $p_j = \frac{1}{3}\sum_{i=1}^{3}[P(X_{i1} < X_{j1}) + 0.5P(X_{i1} = X_{j1})]\,,\ j \in \{AA, AG, GG\}$
    for any distributed traits [KLH12]
  - iv Hazard rates for (censored) time-to-event data (see next chapter)
- Test statistic for relative effect size:

$$T_\ell = \sqrt{N}\frac{\mathbf{c}_\ell'(\widehat{\mathbf{p}} - \mathbf{p})}{\sqrt{\widehat{v}_\ell}}.$$

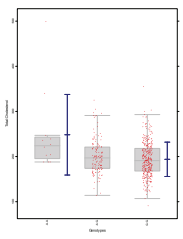
- Compatible simultaneous confidence intervals for the three genetic effects p_{dom}, p_{rec} and p_{add}:

$$\widehat{p}_\ell \pm z(1-\alpha, \widehat{\mathbf{R}}) \cdot SE(\widehat{p}_\ell),$$

A non-parametric approach II

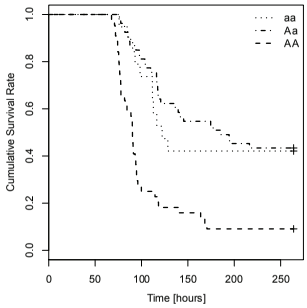
- Range preserving confidence intervals can be easily constructed by using the delta method [Kon09]
  - i a non-parametric approach
  - ii a Behrens-Fisher modification
  - iii a max-test sensitive against more then additive model
- Analysis of the total cholesterol example

Model	Est	95%-sCI	adj. p-Value
Dom	-0.24	[-0.38; -0.08]	0.0058
Add	-0.25	[-0.39; -0.09]	0.0043
Rec	-0.05	[-0.12; 0.02]	0.13



Time-to-event data I

- Example: survival of 116 female mice with the three genotypes *aa*, *Aa* and *AA* at the marker *DM13D147* in chromosome 13 after an infection with *Listeria monocytogenes* [Bro03]



- Effect size: hazard rate. Using Cox-model
- Single linear combination $\hat{L} = \sum_{i \in \{aa,Aa,AA\}} c_i \hat{\beta}_i$

Time-to-event data II

- A lower simultaneous $(1 - \alpha)$ Wald confidence limit for the hazard ratio $\exp(L)$:

$$\sum_{i \in \{aa,Aa,AA\}} c_{mi} \hat{\beta}_i - z_{3,\mathbf{R},1-\alpha} \sqrt{\sum_{i \in \{aa,Aa,AA\}} \sum_{j \in \{aa,Aa,AA\}} c_{mi} c_{mj} \hat{v}(\hat{\boldsymbol{\beta}})_{ij}},$$

- where $z_{3,\mathbf{R},1-\alpha}$ is the upper equicoordinate $(1 - \alpha)$ quantile of the multivariate normal distribution with expectation $\mathbf{0}$ and correlation matrix $\mathbf{R}$ [HH13]
- Analysis:

Inheritance-specific contrast	Hazard ratio	Lower confidence limit
C_{dom}	1.56	0.82
C_{add}	3.16	1.60
C_{rec}	3.50	2.23

- Recessive mode likely

R libraries - LUH and friends I

- multcomp
- mvtnorm
- mratios
- MCPAN
- SimComp
- nparcomp

Summary I

- Max-test using GLM or for relative effect size can be used-asymptotically
- R packages exist
- Can be used for specific analysis, not for genome-wise screening
- Use the clinical interpretation of sCI, instead of reporting tiny p-values
- Extensions available, e.g.
  1. mode-specific genotype-by-environmental interactions
  2. ratio-to-common risk test
  3. max-4 test, considering over-dominance

References I

[Bro03] Broman, KW: Mapping Quantitative Trait Loci in the Case of a Spike in the Phenotype Distribution. In: *Genetics* 163 (2003), MAR, Nr. 3, S. 1169–1175. – ISSN 0016–6731

[Dun55] Dunnett, C. W.: A Multiple Comparison Procedure For Comparing Several Treatments With A Control. In: *Journal Of The American Statistical Association* 50 (1955), Nr. 272, S. 1096–1121

[FZLG02] Freidlin, B ; Zheng, G ; Li, ZH ; Gastwirth, JL: Trend tests for case-control studies of genetic markers: Power, sample size and robustness. In: *Human Heredity* 53 (2002), Nr. 3, S. 146–152

[Has08] Hasler, M.: *Multivariate*, Leibniz University of Hannover, Diplomarbeit, 2008

[HH13] Herberich, Esther ; Hothorn, Ludwig A.: A maximum-type association test for censored time-to-event data. In: *tba* xx (2013), OCT, Nr. 1, S. in prep.

[KLH12] Konietschke, F. ; Libiger, O. ; Hothorn, L. A.: Nonparametric Evaluation of Quantitative Traits in Population-Based Association Studies when the Genetic Model is Unknown. In: *Plos One* 7 (2012), Februar, Nr. 2, S. e31242. http://dx.doi.org/10.1371/journal.pone.0031242. – DOI 10.1371/journal.pone.0031242

[Kon09] Konietschke, F.: *Simultane Konfidenzintervalle für nichtparametrische relative Kontrasteffekte*, Georg-August Universität Göttingen, Diss., 2009

[SCK10] Smith, E. N. ; Chen, W. ; Kahonen, M. et a.: Longitudinal Genome-Wide Association of Cardiovascular Disease Risk Factors in the Bogalusa Heart Study. In: *Plos Genetics* 6 (2010), September, Nr. 9, S. e1001094. http://dx.doi.org/10.1371/journal.pgen.1001094. – DOI 10.1371/journal.pgen.1001094