The Transmission Disequilibrium Test and Imprinting Effects Test Based on Case-Parent Pairs

Statistical Genetics Journal Club

Jingyuan Yang

Department of Statistics
The Ohio State University

Jan. 15th 2008

Transmission Disequilibrium Test and Its Extensions

- **Spielman et al. (1993)** Original TDT tests linkage disequilibrium when linkage is present and tests linkage when linkage disequilibrium is present.

Transmission Disequilibrium Test and Its Extensions

- **Spielman et al. (1993)** Original TDT tests linkage disequilibrium when linkage is present and tests linkage when linkage disequilibrium is present.
- **Sun et al. (1999)** 1-TDT tests linkage/association when only one parent's genotype is available.

Transmission Disequilibrium Test and Its Extensions

- **Spielman et al. (1993)** Original TDT tests linkage disequilibrium when linkage is present and tests linkage when linkage disequilibrium is present.
- **Sun et al. (1999)** 1-TDT tests linkage/association when only one parent's genotype is available.
- **Hu et al. (2007)** TDTI tests linkage/association when imprinting effect is present and POET tests existence of imprinting effect.

Transmission Disequilibrium Test and Its Extensions

- **Spielman et al. (1993)** Original TDT tests linkage disequilibrium when linkage is present and tests linkage when linkage disequilibrium is present.
- **Sun et al. (1999)** 1-TDT tests linkage/association when only one parent's genotype is available.
- **Hu et al. (2007)** TDTI tests linkage/association when imprinting effect is present and POET tests existence of imprinting effect.
- **Hu et al. (2007)** 1-TDTI tests linkage/association when imprinting effect is present and only one parent's genotype is availbale. 1-POET tests existence of imprinting effect with one parent's genotype missing.

Data

Suppose M_1 and M_2 are the two marker alleles with population frequencies g and $g' = 1 - g$. Use 0, 1 and 2 to represent the marker genotypes M_2M_2, M_1M_2 and M_1M_1, so F, M and C take possible values of 0, 1 or 2.

- We collect n_m case-mother pairs and n_p case-father pairs.
- $N_{M>C}$ is the number of case-mother pairs with the mother carrying more copies of M_2 than the case, i.e. number of (M=1, C=0) and (M=2, C=0 or 1).
- Define $N_{M<C}$, $N_{F>C}$ and $N_{F<C}$ similarly.

Notations

- Suppose D and d are the mutant and normal alleles with freqencies p and $q = 1 - p$ at a DSL.
- The four risks associated to four ordered genotypes at DSL are denoted by $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$ and $\phi_{d/d}$.
- Genotype relative risks: $\gamma_2 = \phi_{D/D}/\phi_{d/d}$, $\gamma_{1p} = \phi_{D/d}/\phi_{d/d}$, $\gamma_{1m} = \phi_{d/D}/\phi_{d/d}$,
- $\gamma_1 = (\gamma_{1p} + \gamma_{1m})/2$ is the average risk for heterozygotes.
- Degree of imprinting: $I = (\phi_{D/d} - \phi_{d/D})/2$.
- Coefficient of linkage disequilibrium is $\delta = P_{M_1 D} - gp$
- Female and male recombination fractions are denoted by θ_f and θ_m.

Method

15 conditional probabilities $P(F, M, C|\text{the child is a case})$ are derived by Zhou et al. (2007). For example,

$$P(F = 1, M = 0, C = 1|\text{the child is a case})$$
$$=(\phi_{D/D}\omega_3\omega_5 + \phi_{D/d}\omega_4\omega_5 + \phi_{d/D}\omega_3\omega_7 + \phi_{d/d}\omega_4\omega_7)/\phi,$$

where $\phi = p^2\phi_{D/D} + pq\phi_{D/d} + pq\phi_{d/D} + q^2\phi_{d/d}$ is the prevalence of the disease, ω's are related to hyplotype frequencies, which are functions of allele frequencies(p, q, g, g'), LD coefficient(δ) and recombination fractions(θ_f, θ_m).

Method

15 conditional probabilities $P(F, M, C|\text{the child is a case})$ are derived by Zhou et al. (2007). For example,

$$P(F = 1, M = 0, C = 1|\text{the child is a case})$$
$$=(\phi_{D/D}\omega_3\omega_5 + \phi_{D/d}\omega_4\omega_5 + \phi_{d/D}\omega_3\omega_7 + \phi_{d/d}\omega_4\omega_7)/\phi,$$

where $\phi = p^2\phi_{D/D} + pq\phi_{D/d} + pq\phi_{d/D} + q^2\phi_{d/d}$ is the prevalence of the disease, ω's are related to hyplotype frequencies, which are functions of allele frequencies(p, q, g, g'), LD coefficient(δ) and recombination fractions(θ_f, θ_m).

Numbers of those 15 F/M/C categories form a multinomial distribution, whose nice asymptotic properties of MLEs are crucial to the validity of 1-TDT1 and 1-POET.

Method-Cont.

By using the asymptotic property of $\hat{P}_{MLE}$ vector of the multinomial distribution, tests among the following class are proved to be valid to test linkage/association when only one parent's genotype is available.

$$T_\omega = \frac{\omega(N_{M<C} - N_{M>C}) + (1-\omega)(N_{F<C} - N_{F>C})}{\sqrt{\omega^2 N_{M\neq C} + (1-\omega)^2 N_{F\neq C}}}$$

Method-Cont.

By using the asymptotic property of $\hat{P}_{MLE}$ vector of the multinomial distribution, tests among the following class are proved to be valid to test linkage/association when only one parent's genotype is available.

$$T_\omega = \frac{\omega(N_{M<C} - N_{M>C}) + (1-\omega)(N_{F<C} - N_{F>C})}{\sqrt{\omega^2 N_{M\neq C} + (1-\omega)^2 N_{F\neq C}}}$$

1-TDT is a special case of T_ω with $\omega = 0.5$, which is shown to have an inflated type I error rate when imprinting effect exists.

Method-Cont.

By using the asymptotic property of $\hat{P}_{MLE}$ vector of the multinomial distribution, tests among the following class are proved to be valid to test linkage/association when only one parent's genotype is available.

$$T_\omega = \frac{\omega(N_{M<C} - N_{M>C}) + (1-\omega)(N_{F<C} - N_{F>C})}{\sqrt{\omega^2 N_{M\neq C} + (1-\omega)^2 N_{F\neq C}}}$$

1-TDT is a special case of T_ω with $\omega = 0.5$, which is shown to have an inflated type I error rate when imprinting effect exists.

By taking $\omega = \omega_0 = n_p/(n_m + n_p)$, the authors get the 1-TDTI test, which is good for testing linkage/association when imprinting effect is present. Its power and size are evaluated thorugh simulation.

Method-Cont.

To test the existence of imprinting effect, the authors showed that, when the population mating is symmetric and the female and male recombination fractions are the same ($\theta_f = \theta_m$), the following test statistic

$$\text{1-POET} = \frac{\omega_0(N_{M<C} - N_{M>C}) - (1-\omega_0)(N_{F<C} - N_{F>C})}{\sqrt{\omega_0^2 N_{M\neq C} - (1-\omega_0)^2 N_{F\neq C}}}$$

has a zero expectation and asymptotically follows a N(0,1) distribution under the hull hypothesis of no imprinting.

Method-Cont.

To test the existence of imprinting effect, the authors showed that, when the population mating is symmetric and the female and male recombination fractions are the same ($\theta_f = \theta_m$), the following test statistic

$$\text{1-POET} = \frac{\omega_0(N_{M<C} - N_{M>C}) - (1-\omega_0)(N_{F<C} - N_{F>C})}{\sqrt{\omega_0^2 N_{M\neq C} - (1-\omega_0)^2 N_{F\neq C}}}$$

has a zero expectation and asymptotically follows a N(0,1) distribution under the hull hypothesis of no imprinting.

Curves in Fig. 3 "the acutal powers of the 1-POET against the number of case-mother pairs $n_m \in [50, 150]$ in increments of 10 (pairs total is 200)" all have a convex shape with the peak at $n_m = 100$, which indicates the best scheme of data collection is $n_m : n_p = 1$.

Discussion

- In theory, $\theta_f = \theta_m$ is required to guarantee the expectstion of 1-POET being 0 under the null hypothesis of no imprinting. In the presence of the marker locus and a DSL, it's plausible to assume both θ_f and θ_m are small, so their difference is consequently small.
- In practice, it is common to have two kinds of data, case-parents trios and case-parent(mother or father) pairs, statistcs to combine those two type of data are discussed.
- Compare to Weinberg's PAT method