APPENDIX

Let N_{ij} be the count of heterozygotic parents with haplotypes i, j transmitting haplotype i to their child, and N_{ji} the count of parents with the same genotype but transmitting haplotype j to their child. Let n_{ii} be the count of homozygotic parents for haplotype i. Let n_{ij} be $N_{ij} + N_{ji}$. Let consider n_{ij} a realization of the random variable X_{ij}, $i = 1, \ldots, k$, $j = i + 1, \ldots, k$, n_{ji} a realization of the random variable X_{ji} and n_{ii} a realization of the random variable X_{ii}. Thus, $X_{ji} = N_{ij} - X_{ij}$ holds. Thus, the total counts of transmitted and non-transmitted haplotypes in group g_1 n_{g_1T} and n_{g_1U} are then realizations of the random variables

$$X_{g_1T} = \sum_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} \left[\sum_{\substack{j=1 \\ n_{jT} > n_{jU}}}^{k} X_{ij} - X_{ii} \right]$$

and

$$X_{g_1U} = \sum_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} \left[\sum_{\substack{j=1 \\ n_{jT} > n_{jU}}}^{k} X_{ji} - X_{ii} \right].$$

The variance of TDT_{2G} is therefore

$$Var(mTDT_{2G}) = \frac{1}{4} Var \left[\frac{(X_{g1T} - X_{g1U})^2}{X_{g1T} + X_{g1U}} + \frac{(X_{g2T} - X_{g2U})^2}{X_{g2T} + X_{g2U}} \right]$$

.

The random variable Y_{g_1} defined as the first summand can be rewritten as:

$$Y_{g_1} = \frac{\left[\sum\limits_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} \left[\sum\limits_{\substack{j \neq i \\ n_{jT} > n_{jU}}} (2X_{ij} - n_{ij}) \right] \right]^2}{\sum\limits_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} \sum\limits_{\substack{j \neq i \\ n_{jT} > n_{jU}}} n_{ij}} = \frac{\left[\sum\limits_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} \left[\sum\limits_{\substack{j \neq i \\ n_{jT} > n_{jU}}} 2X_{ij} - n_i \right] \right]^2}{\sum\limits_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} n_i}$$

An equivalent rewritten can be done for Y_{g_2}.

Thus,

$$Y_{g_1} = \frac{\left[\sum\limits_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} \sum\limits_{\substack{j \neq i \\ n_{jT} > n_{jU}}}^{k} 2X_{ij} - n_{g_1} \right]^2}{n_{g_1}}$$

is asymptotically χ_1^2 under the null hypothesis, as its square root is a standard normal, if we take into account that under the null X_{ij} is binomial with parameters $(n_{ij}, 1/2)$, for large samples $2X_{ij}$ is asymptotically normal with mean and variance n_{ij} and thus, as X_{ij} and X_{ik} are independent for any $k \neq j$, $\sum_{\substack{i=1 \\ n_{iT} > n_{iU}}}^{k} 2X_{ij}$ is normal with mean and variance n_{g_1}.

The same follows for Y_{g_2} and thus, the variance

$$Var(mTDT_{2G}) = \frac{1}{4}\left[VarY_{g_1} + VarY_{g_2} + 2Cov(Y_{g_1}, Y_{g_2})\right] = 2$$

taking into accoung that $Y_{g_1} = Y_{g_2} n_{g_1}/n_{g_2}$ and thus $Cov(Y_{g_1}, Y_{g_2}) = Var(Y_{g1}) n_{g_1}/n_{g_2}$ and $n_{g_1} = n_{g_2} = n/2$ under the null hypothesis.