

TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

KALITIMDA DENGESİZ AKTARIM TESTİ

Umut ÖZBEK

BİYOİSTATİSTİK ANABİLİM DALI
YÜKSEK LİSANS TEZİ

DANIŞMAN
Prof. Dr. Ersöz TÜCCAR

2007-ANKARA

Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı

Uzmanlık eğitimi çerçevesinde yürütülmüş olan “Kalıtımda Dengesiz Aktarım Testi” başlıklı Umut ÖZBEK’e ait bu çalışma aşağıdaki jüri tarafından Uzmanlık tezi olarak kabul edilmiştir.

Tez Savunma Tarihi : 16/07/ 2007

Prof. Dr.Ersöz TÜCCAR
Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı Başkanı
Jüri Başkanı

Doç.Dr.Atilla Halil ELHAN
Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Üye

Yrd.Doç.Dr.Timur TUNCALI
Ankara Üniversitesi Tıp Fakültesi
Tıbbi Genetik Anabilim Dalı
Üye

Yrd.Doç.Dr.Yasemin GENÇ
Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Üye

Yrd.Doç.Dr.Kenan KÖSE
Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Üye

İÇİNDEKİLER

KABUL ve ONAY SAYFASI	ii
İÇİNDEKİLER DİZİNİ	iii
ÖNSÖZ	v
KISALTMALAR DİZİNİ	vi
ŞEKİLLER DİZİNİ	vii
TABLolar DİZİNİ	viii
GRAFİKLER DİZİNİ	ix
1. GİRİŞ	1
1.1 Genetik Bilimi	2
1.2 Genetik ile İlgili Temel Bilgiler	2
1.3 Genlerin Kalıtımı	4
1.4 Genetik Yapı ve Genlerin Konumlarının Belirlenmesi	6
1.5 Genetik Epidemiyoloji	9
1.6 Gen Haritalama	11
1.6.1 Bağlantı Analizi	12
1.6.1.1 Parametrik Bağlantı Analizi	12
1.6.1.2 Parametrik Olmayan Bağlantı Analizi	16
1.6.1.2.1 Hasta Kardeşler Yöntemi	19
1.6.1.2.2 Hasta Akraba Çiftleri Yöntemi	20
1.6.1.2.3 Kantitatif Karakterler için Kardeş Çifti Yöntemi	20
1.6.2 İlişki Analizi	21
1.6.2.1 Aile Bazlı İlişki Analizi Yöntemleri	28
1.6.2.1.1 Kalıtımda Dengesiz Aktarım Testi (TDT)	28
1.6.2.1.2 Haplotip Bazlı Haplotip Göreli Risk Testi (HHRR)	36
1.6.2.1.3 Çok Alelli Marker ile Kalıtımda Dengesiz Aktarım Testi	38

1.6.2.1.4 1-TDT: Sadece Bir Ebeveynin Genotipine Ulaşılabildiğinde TDT	43
1.6.2.1.5 Ailesel Bilgi Olmadığı Durumlarda Test	48
1.6.2.1.6 İlişki Olduğunda Kardeşlerle Yapılan Bağlantı Testi (S-TDT)	48
1.6.2.1.7 Uyumsuz Kardeşlik Testi (SDT)	50
1.6.2.1.8 Genişletilmiş Ailelerde İlişki Testi	52
1.6.2.1.9 Farklı Tipteki Ailelerden Alınan Bilginin Birleştirilmesi	55
1.6.2.1.10 TDT'nin İstatistiksel Gücü	58
2. GEREÇ VE YÖNTEM	60
2.1 Çok Alelli Marker İçin TDT Programı	60
2.2 TDT _i Uygulaması	66
2.3 ETDT Uygulaması	66
2.4 1-TDT Uygulaması	66
3. BULGULAR	67
3.1 TDT _a	67
3.2 TDT _i	67
3.3 ETDT	68
3.4 1-TDT	69
4. TARTIŞMA	71
5. SONUÇ VE ÖNERİLER	73
ÖZET	74
SUMMARY	75
KAYNAKLAR	76
EK 1	80
ÖZGEÇMİŞ	85

ÖNSÖZ

Gerek bölüm başkanımız gerekse tez danışmanım olarak hiçbir desteğini ve bilgisini esirgemeyen değerli hocam Prof. Dr. Ersöz Tüccar'a, yüksek lisans öğrenimim boyunca bana destek olan, öğrencisi olduğumuz için kendimizi çok şanslı hissettiğimiz sevgili hocam Doç. Dr. Atilla Halil Elhan'a, genetik alanında çalışmamız için bizi destekleyen ve değerli bilgilerini paylaşan, tezimin veri setini sağlayan hocam Doç. Dr. Nurten Akarsu'ya,

Her türlü destek ve anlayışları için Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı'nda çalışan hocalarıma ve arkadaşlarıma,

Beni her konuda yüreklendiren ve hep yanımda olan anneme, babama ve kardeşim Gizem'e,

Tez çalışmam süresince, sınırsız desteğiyle yanımda olan Alper'e,

Hayatımdaki katkılarından ötürü teşekkürü bir borç bilirim.

KISALTMALAR DİZİNİ

DNA	: Deoksribonükleik Asit
LOD	: Odds oranının logaritması
IBD	: Ebeveynlerden Dolayı (Alel paylaşımı açısından) Aynı Olma Durumu
IBS	: (Alel paylaşımı açısından) Aynı Olma Durumu
TDT	: Kalıtımda Dengesiz Aktarım Testi
HHRR	: Haplotip Bazlı Haplotip Göreli Risk Testi
ETDT	: Genişletilmiş Kalıtımda Dengesiz Aktarım Testi
1-TDT	: Sadece Bir Ebeveynin Genotipine Ulaşılabildiğinde TDT
S-TDT	: İlişki Olduğunda Kardeşlerle Yapılan Bağlantı Testi
SDT	: Uyumsuz Kardeşlik Testi
PDT	: Genişletilmiş Ailelerde İlişki Testi

ŞEKİLLER DİZİNİ

Şekil1.1 Çapraz değişim	4
Şekil 1.2 Otozomal dominant (a) ve resesif (b) kalıtım	7
Şekil 1.3 Genotiplerle aile ağacı	8
Şekil 1.4 Alel Paylaşımı	16
Şekil 1.5 TDT-PC Programı	59
Şekil 2.1 Program görüntüsü 1	62
Şekil 2.2 Program görüntüsü 2	63
Şekil 2.3 Program görüntüsü 3	64
Şekil 2.4 Program görüntüsü 4	65

TABLOLAR DİZİNİ

Tablo 1.1 IBD Olarak Alel Paylaşım Olasılıkları	18
Tablo 1.2 Beklenen ve Gözlenen Alel Paylaşım Dağılımları	19
Tablo 1.3 3 alt popülasyonun alel ve haplotip göreceli sıklıkları	25
Tablo 1.4 2n ebeveynin n hasta çocuğa geçirdiği alel sayıları a,b,c ve d	29
Tablo 1.5 Geçirilen ve geçirilmeyen alellerin kombinasyonlarının olasılıkları	30
Tablo 1.6 Çok alelli markerda geçirilen ve geçirilmeyen allel sayıları	38
Tablo 1.7 Çok alelli marker için geçirilen ve geçirilmeyen alellerin Olasılıkları	40
Tablo 1.8 5 alelli marker için lojistik regresyon analizinde kullanılacak veri düzenlemesi	42
Tablo 1.9 Bir ebeveyn mevcutken vaka-ebeveyn kontrol çalışma tasarımı	43
Tablo 1.10 Birleşik alel dağılımları	47
Tablo 1.11 Genotiplerin olasılıkları	48
Tablo 3.1 TDT _i değerleri	68
Tablo 3.2 9 alel için parametre kestirimleri ve standart hataları	68
Tablo 3.3 1-TDT için standart normal dağılıma göre analiz sonuçları	69
Tablo 3.4 1-TDT için binom dağılıma göre sonuçlar	70

GRAFİKLER DİZİNİ

Grafik 1.1 LOD Skor Grafikleri	15
Grafik 1.2 Kuşaklar arası bağlantı dengesizliği	24

1. GİRİŞ

Her alanda olduğu gibi genetik alanında da son yüzyılda pek çok gelişme kaydedilmiştir. Bunun sonucunda günümüzde genetik olarak çok fazla bilgiye ulaşılabilmektedir. Genetik çalışmalardan elde edilen sonuçların anlamlı bir hal alabilmesi istatistiğin etkili kullanımıyla mümkün olmaktadır. Tüm bu bilgilerin kullanılması ve değerlendirilmesi için ise istatistiksel yöntemlere gereksinim duyulmaktadır. Bu nedenle özel olarak genetik alanında kullanılabilecek çeşitli istatistiksel yöntemler geliştirilmiş ve halen geliştirilmektedir.

İleri gelen insan genetiği dergilerinde çıkan makalelerin çoğu yeni istatistiksel analiz yöntemleri geliştirmek üzerine yazılmaktadır (Sham,1998). İstatistik ile genetik arasında geçmişe dayanan bir bağ bulunmaktadır. Modern istatistiğin iki öncü ismi Karl Pearson ve Ronald Fisher, çalışma hayatlarının belli bir döneminde genetikle ilgilenmişlerdir. Karl Pearson, Francis Galton ve William Weldon'un beraber kurdukları Biometrika dergisinin ilk sayısında (1901) başmakalede biyologların, matematikçilerin ve istatistikçilerin farklı alanlarda çalıştıkları ancak birbirlerinin konularını anlayarak beraber çalışacakları ve böylece de biyometri alanının geliyeceği ve yeni çalışma sahalarının ortaya çıkacağı belirtilmiştir.

Genetik ve istatistik arasında eskiye dayanan bir bağ olmasına rağmen genetik istatistiği hem genetikçilere hem de istatistikçilere zor gelmiştir. Ancak günümüzde İstatistik, Genetik, Matematik, Bilgisayar Mühendisliği gibi birden fazla disiplinin bir araya getirilmesiyle genetik verilerin analizi için gereken yöntemler geliştirilmektedir.

Bu çalışmanın amacı, hastalıkların genetik ilişkilerini ortaya koyup, onları önlemek ya da tedavi etmek için son yıllarda geliştirilen aile bazı istatistiksel genetik analiz yöntemlerinden sık kullanılanlar hakkında bilgi

vermek ve uygun olan analiz yöntemlerini elimizde bulunan veriye uygulayarak göstermektir.

1.1 Genetik Bilimi

Genetik, canlı organizmalarda kalıtım ve değişimin incelendiği bir bilim dalıdır (Griffiths ve ark. 2000). İstenen özelliklerin kalıtılabildiği bilgisi tarih öncesi çağlardan beri bitki ve hayvan kalitesinin artırılması için kullanılmıştır. Ancak kalıtımın bilimsel boyutunu araştıran modern genetik, 1800'lü yılların ortalarında Gregor Mendel'in yaptığı çalışmalarla ortaya çıkmıştır (Weiling, 1991). Mendel kalıtımın, belli özelliklerin bağımsız olarak aktarıldığı bir süreç olarak gözlemlemiştir. Kalıtımın bu temel birimi bugün 'gen' olarak adlandırılmaktadır. Gen yaşayan canlılardaki kalıtım birimidir. Genler, canlı organizma hücrelerinde DNA, deoksiribonükleik asit, molekülünün içinde fiziksel olarak yer alırlar. Genlerde taşınan bilgi hücrenin bölümlerinin oluşturulması ve kontrol edilmesi için kullanılır. Genetik bilgi canlının dış görünüşüne ve davranışlarına büyük ölçüde etki etse de, asıl olarak organizmanın maruz kaldığı dış çevre ile genetiğinin etkileşimi canlının yapısını belirlemektedir. Örneğin; genler kişinin boyunu büyük ölçüde belirlese de, beslenme şekli, çocukken maruz kaldığı hastalıklar gibi dış faktörlerin de kişinin boy uzunluğuna büyük etkisi vardır.

1.2 Genetik ile İlgili Temel Bilgiler

Bireyin genetik bilgisi hücre çekirdeğinin içinde bulunan 23 çift kromozomdadır. Bunların 22 çifti otozom, vücut kromozomları, 1 çifti de cinsiyet kromozomlarıdır, gonozom veya eşey kromozomları olarak da adlandırılırlar.

Kromozomların kimyasal yapısında deoksiribonükleik asit, DNA bulunur. DNA, DNA ipliği olarak adlandırılan ve birbiri etrafında dönerek

sarmal bir yapı oluşturan bir çift molekülden oluşur. Bu ipliklerden her biri nükleotid zinciridir. Nükleotidler; şeker, fosfat ve azotlu bazdan oluşur. Dört çeşit azotlu baz bulunmaktadır: Adenin (A), guanin (G), sitozin (C) ve timin (T). DNA ipliği üzerinde bulunan gen içindeki nükleotid dizisi canlının yaşamı boyunca üretmek zorunda olduğu proteinleri tanımlar. Nükleotid dizisi ile proteinleri oluşturan amino asit dizisi arasındaki ilişki genetik kodlarla belirlenir. Genetik kod 'kodon' adı verilen üçlü nükleotid dizisinden oluşur (Örn: ACA, TTG gibi). Her kodon bir amino asidi belirlemektedir. Dört çeşit baz bulunduğu için $4^3 = 64$ farklı kodon vardır. Ancak 20 çeşit amino asit bulunduğu için bir amino asidin birden fazla belirleyici kodonu bulunmaktadır.

Watson ve Crick 1953'te birbirine bağlı iki DNA ipliğiyle birlikte kromozomların çift sarmal yapıda olduklarını öne sürmüşlerdir. Bir iplikte bulunan her bir baz, diğer iplikteki baza hidrojen bağıyla bağlıdır. A bazı T'ye, G bazı da C'ye bağlanmaktadır. Bu nedenle iki iplik de aynı genetik bilgiyi taşımaktadır. 23 kromozomdaki toplam baz çiftleri sayısı yaklaşık olarak 3×10^9 kadardır.

Gen adı verilen kalıtım birimlerinin varlığı 1865 yılında Mendel tarafından ortaya atılmıştır. Genler, kromozomlar üzerinde yer almaktadır. Şu anda insanda 30000 gen olduğu tahmin edilmektedir. Kromozom üzerinde bulunan ve DNA'nın bir parçası olan gen, proteinin yapı ve işlevlerini gösteren amino asit dizilimini belirler. Genin değişik biçimlerine alel adı verilir. Genler alel çiftleri halinde bulunurlar. Her bir alel, aynı özelliğin farklı kodlarını taşıdığı için farklı özelliklerin ortaya çıkmasını sağlar. Her eş kromozomda her karakter için genin bulunduğu belli yere lokus denir. Organizmanın genetik yapısı olan genotip, aynı lokusta biri anneden diğeri babadan kalıtılmış olan bir çift alel içerir. Örneğin A ve a alellerinden oluşan bir gen için olası genotipler aa, AA, Aa dır. Genotipi oluşturan aleller aynı ise (Örn. AA veya aa) birey homozigot, farklı ise

(Örn: Aa) birey heterozigottur denir. Bir ebeveynden kalıtılan farklı lokuslardaki alel dizisine haplotip denir.

1.3 Genlerin Kalıtımı

Her bir hücrede bulunan 46 kromozomun 23'ü anneden, 23'ü de babadan kalıtılmaktadır. Her bir ebeveynden aktarılan tüm kromozomlar hem o ebeveynin annesinden hem de babasından parçalar taşımaktadır. İki kromozom arasındaki DNA parçalarının değişimine çapraz değişim (crossover) adı verilir. (Şekil 1.1)



Şekil1.1 Çapraz değişim

Yumurta oluşurken sadece annenin nükleotidlerinin yarısı geçirilir. Annenin ve babanın ebeveynlerine ait olan parçaların karışması işlemine mayoz adı verilir. Aynı şekilde sperm hücreleri oluşurken babanın ebeveynlerinin DNA'larına ait parçalar mayoz yoluyla karışır.

DNA ipliğinin bir kısmının ayrılıp başka bir DNA molekülünün sonuna birleşmesi olayına rekombinasyon denir. Rekombinasyon sayesinde ebeveynlerden gelen aleller ileri kuşaklarda yeni kombinasyonlar oluşturabilirler. Aynı kromozomda iki lokusu ele alalım. İlk lokus için olası aleller A ve a, ikinci lokus için olası aleller B ve b olsun. Birey anneden AB, babadan da ab haplotiplerini almış olsun. Bu durumda

söz konusu bireyin lokuslarındaki genotipler Aa ve Bb dir. Eğer gamet (üreme hücresi) adı verilen yumurta veya sperm mayoz sırasında Ab haplotipini almışsa, bu haplotip oluşurken rekombinasyon meydana gelmiştir. Ab haplotipini oluşturan iki alel A ve b farklı ebeveynlerden gelmiştir. Aynı şekilde aB haplotipi de rekombinanttır. AB ve ab ise rekombinant olmayan haplotiplerdir. Eğer iki lokus arasında tek sayıda çapraz değişim olmuşsa iki lokus rekombinanttır. Çift sayıda çapraz değişim söz konusu ise iki lokus rekombinant değildir (Almgren ve ark. 2002).

Çapraz değişim her bir kromozom boyunca rasgele olarak gerçekleşir. İki lokus arasında 1 Morgan (100 santiMorgan, cM) mesafe vardır. 1 cM, istatistiksel olarak düzeltilmiş %1'lik rekombinasyon sıklığına sahip olan iki lokus arasındaki genetik mesafe olarak tanımlanmaktadır. cM olarak ifade edilen genetik uzaklık sayısal olarak yüzde şeklinde ifade edilen rekombinasyon sıklığına eşittir. Kadınlar ve erkeklerde birbirlerinden farklıdır. Kromozomların uzunlukları birbirlerinden çok farklı olabilirken, otozomların ortalama genetik mesafesi 1.6 Morgandır (Almgren ve ark. 2002).

Rekombinasyon fraksiyonu θ , mayoz sırasında iki lokusun rekombinant olma olasılığını verir. Rekombinasyon fraksiyonu, iki lokus arasındaki genetik mesafe olan x 'in fonksiyonu olarak ifade edilmelidir. Çünkü birbirine çok yakın iki lokusun rekombinant olma olasılığı çok düşüktür.

Çapraz değişimi gösteren pek çok olasılık modeli vardır. En anlaşılır olan ve en sık kullanılan Haldane (1919) tarafından geliştirilmiştir. Bu model, çapraz değişimin kromozom boyunca rasgele olarak, Poisson dağılıma göre meydana geldiği varsayımına dayanır. Haldane'nin uzaklık fonksiyonu, θ rekombinasyon fraksiyonu cM cinsinden genetik uzaklığı gösteren x 'in fonksiyonu olarak; Denklem 1.1'de gösterildiği gibidir.

$$\theta(x) = 0.5(1 - \exp(-0.02x))$$

Denklem 1.1

$\theta < 0.5$ olduğunda aynı kromozom üzerinde bulunan iki lokus birbirleriyle bağlantılıdır denir. Farklı kromozomlar üzerinde bulunan lokuslar birbirleriyle bağlantılı değildir ve rekombinasyon fraksiyonu $\theta = 0.5$ 'tir. Farklı kromozomlar üzerinde bulunan lokuslar arasındaki genetik mesafe sonsuz olarak tanımlanabilir. Bu durumda bu iki lokusun kalıtımı birbirinden bağımsız olaylardır.

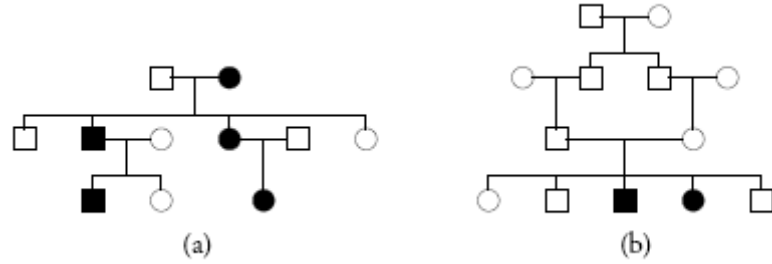
1.4 Genetik Yapı ve Genlerin Konumlarının Belirlenmesi

Genotiplerin kesin olarak belirlenmesi her zaman mümkün değildir. Fenotip, genotipin ve dış etkenlerin canlının dış görünüşündeki yansımasıdır. Genotipin ve çevresel faktörlerin fenotipteki etkileri genetik model yoluyla açıklanabilir. Genetik model bir veya daha fazla gen içerebilir. Tek gen hastalıklarında, hastalığın ortaya çıkmasından yalnız bir gen sorumludur. Örnek olarak Huntington hastalığı tek gen hastalığıdır (Gusella ve ark., 1983). Hastalığın genetik kısmı birden fazla genin bir araya gelmesiyle oluşuyorsa, bu tip hastalıklara çok gen hastalıkları adı verilir. Tip II diyabet çok gen hastalıklarına örnektir (Horikawa ve ark., 2000). Farklı alt popülasyonlarda farklı gen ya da genler aynı hastalıktan sorumlu olabilirler. Bu durumda bu hastalıklar heterogenik hastalıklardır denir. Kalıtsal meme kanseri bu tür hastalıklardandır. Farklı popülasyonlarda iki farklı genin bu hastalıktan sorumlu olduğu bulunmuştur (Hall ve ark., 1990, Wooster ve ark., 1995).

Gen veya genlerin fenotipte yarattığı etki penetrans parametreleriyle açıklanır. Penetrans, genotiple fenotip arasındaki ilişkidir. Başka bir deyişle genotipteki özelliklerin fenotipte ifade bulmasıdır. Genetik modelin penetrans parametrelerini ve çevresel etki parametrelerini, hastalığın sık görüldüğü

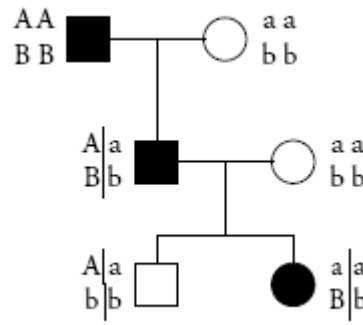
ailelerin fenotiplerinden oluşan verilerle bulunmasına segregasyon (ayırma) analizi denir.

Tek gen hastalıklarında, genotipte tek bir hastalık alelinin bulunması hastalığın ortaya çıkmasına yetiyorsa hastalık dominant (baskın), hastalığın ortaya çıkması için iki hastalık aleli gerekiyorsa hastalık resesif (çekinik) tir denir.



Şekil 1.2 Otozomal dominant (a) ve resesif (b) kalıtım

Şekil 1.2’de dominant (a) ve resesif (b) hastalıkların görüldüğü aile ağaçları verilmiştir. Kadınlar daire, erkekler de karelerle gösterilmektedir. Şekillerin renkleri bireylerin fenotiplerini belirtmektedir. Siyah olanlar hasta bireyler, beyazla gösterilenler sağlıklı bireylerdir. Aile ağacının başladığı bireyler kurucu olarak adlandırılır. a ailesinde kadın kurucu bireyde görülen hastalık, üç kuşakta da ortaya çıkmıştır. b ailesinde kuruculardan biri hastalık alelini taşımaktadır. Bu alel 3 kuşak boyunca kalıtılmıştır. Hastalık resesif özellik gösterdiği için hastalık aleli 3 kuşakta kendini göstermemiştir. 4. kuşakta iki kuzen arasında gerçekleşen evlilik sonucu doğan çocuklar, taşıyıcı olan hem anne hem babadan hastalık alelini almışlardır. Böylece resesif özellik gösteren hastalık fenotipe yansımıştır.



Şekil 1.3 Genotiplerle Aile Ağacı

Şekil 1.3'te iki lokustaki genotiplerin bilindiği aile ağacı verilmiştir. Aleller arasındaki dikey çizgiler haplotiplerin hangi ebeveynden kalıtıldığına bilindiği durumu anlatmaktadır. İkinci kuşaktaki erkek bireyde olduğu gibi bazen anneden ve babadan gelen haplotipler açık olarak görülmektedir. Burada A-B haplotipi babadan alınırken a-b haplotipi anneden alınmıştır. Birey her iki lokusta da heterozigot olduğu için durum açıktır. 3. kuşaktaki erkek birey için de A-b haplotipini babadan, a-b haplotipini anneden aldığı söylenir. Bu bilgiye ek olarak babadan gelen haplotipin rekombinant olduğu da görülmektedir. Çünkü A ve b alellerinden biri büyükanneden diğeri de büyükbabadan alınmıştır.

Genetik modelde gen ya da genlerin lokuslarının belirlenmesi diğer bir önemli husustur. Bunun için marker genotipleri bilinmelidir. Genetik marker, belirlenmiş DNA dizisine verilen addır. Kromozom üzerindeki yeri tam olarak bilinen, en az iki olası alelli lokuslardır. Gen olması gerekmez. Kalıtılan hastalık ve genetik nedeni arasındaki ilişkiyi ortaya çıkarmak için kullanılır. Kromozomda birbirine yakın olan DNA parçaları beraber kalıtılma eğilimi gösterirler. Bu bilgiden yola çıkarak, marker, kromozom üzerindeki yeri tam olarak bulunmamış genlerin kalıtımı çalışmalarında kullanılır. Ailedeki bireylerin marker genotiplerinin bilinmesi kalıtım deseninin çıkarılmasına olanak sağlar. Bağlantı analizi, markerlardan elde edilen kalıtım deseninin fenotiplerden elde edilen kalıtım deseniyle korelasyonunun yüksek olduğu bölgeleri analiz eder. Bağlantı analizi, birbirine yakın olan iki lokus arasında

 apraz deęiřim olasılıęının  ok d ř k olduęu dolayısıyla iki lokusun bereber kalıtılacakları varsayımına dayanır. İliřki analizinde ise, hastalık lokusuna yakın olan marker, g sterge, lokusunda bazı alellerin beklenenden daha sık kalıtılması durumu kullanılır.

1.5 Genetik Epidemiyoloji

Genetik Epidemiyoloji; insan pop lasyonunda hastalıęa neden olan genlerle  evresel etmenler arasındaki iliřkiyi ve bu hastalıkların ailelerdeki kalıtım desenlerini  alıřan bilim dalıdır. (Thomas, 2004) Genetik epidemiyoloji  alıřmaları aile bazlı ve pop lasyon bazlı olarak yapılır. Aile  alıřmaları genetik epidemiyologların kullandıkları bir yaklařımdır. Kohort ve vaka-kontrol  alıřmaları da alternatif yaklařımlardır.

Genetik epidemiyoloji  alıřmalarında ama  hastalıęın genetik dayanaęının arařtırılmasıdır. Bu arařtırma yapılırken řu sıra izlenir:

1. Tanımlayıcı epidemiyoloji adı verilen ilk safhada hastalıęın g sterdięi uluslar arası deęiřiklik deseni arařtırılır. G   eden bireylerde hastalık riskindeki deęiřiklik g zlemlenir. Aynı zamanda ırkın, sosyal stat n n, yařın ve cinsiyetin yarattıęı farklılıklar da izlenir. Bu g zlemlerden yola  ıkılarak hastalıęın genetik ya da  evresel fakt rlerden etkilenip etkilenmedięi arařtırılır.
2. Genetik nedenler arařtırılırken ilk adım hastalıęın aile i inde g r lme sıklıęının incelenmesidir. Eęer hastalık ailede, rasgele olarak beklenenden daha sık g r l yorsa, hastalıęın oluřumunda genetik etkilerden s z edilebilir. Daha sonra, ailesel yatkınlıęın akrabalıęın derecesiyle, yařla ve  evresel fakt rlerle ne kadar iliřkili olduęu arařtırılır. Bu arařtırmalar genellikle aile  yk lerini karřılařtıran vaka-kontrol  alıřmalarına dayalı olarak ya da ikiz veya evlatlık  alıřmaları baz alınarak y r t l r.

3. Bu adımda akrabalar arasındaki hastalık deseninin bir veya birden fazla genle veya ortak olan çevre faktörleriyle uyumu araştırılır. Eğer uyumlu gen bulunursa, ilgili genetik model parametreleri tahmin edilir. Bu çalışmaya Segregasyon Analizi adı verilir. Popülasyon bazlı aile çalışmaları bu amaçla kullanılır.
4. Daha sonraki aşama Bağlantı Analizidir. Bilgi veren aile üyelerinden kan örnekleri alınır ve bilinen bölgelerdeki genetik markerlara bakılır. Hastalıkla paralel olarak aile boyunca aktarılan markerlar (cosegregation), genin kromozomal bölgesi hakkında bilgi verirler. Genişletilmiş aileler bu amaçla, popülasyon bazlı olmasına gerek olmadan kullanılabilirler. Ayrıca çok sayıda etkilenmiş kardeş çiftleri de aynı amaçla çalışılabilir (Thomas, 2004). Bu çalışma belli bir kromozom seçilmeden, tüm genomun taranmasıyla ve çalışma ilerledikçe ilgilenilen bölgenin daralmasıyla gerçekleştirilebilir.
5. Araştırılan bölge daraldıkça, daha da özele inmek için başka yöntemler kullanılır. Bu çalışmaya ince haritalama adı verilir. Birbiriyle ilişkisi olmayan vaka ve kontroller aday bölgedeki markerlar açısından ilişki bulmak için karşılaştırılırlar. Bulunan ilişkiler herhangi bir genle bağlantı dengesizliği oluşturabilir. Bağlantı dengesizliği alellerin veya markerların bazı kombinasyonlarının popülasyonda beklenenden daha az ya da daha fazla görülmesidir.
6. Bu aşamada aday genlerle ilişkilendirme yapılır. Bağlantılı bölge hastalığın nedeni ile ilgili fonksiyonları bilinen bazı genler içerebilir. Popülasyon tabanlı veya aile bazlı vakalar ve kontroller arasında aday bölgedeki genotiplerin karşılaştırılmasıyla, hastalıkla gerçekten ilişkili olup olmadığı test edilebilir. Olası ilişki, ince haritalamada olduğu gibi, rasgele bir genle bağlantı dengesizliğini yansıtan ve rasgele olmayan bir ilişki olabilir (Thomas, 2004).

7. Sonraki aşama, gen klonlanması ve mutasyonların belirlenmesi aşamasıdır. Aday bölge yeterince daraldığında ve bu bölgede aday herhangi bir gen bulunmadığında, bölgedeki DNA polimorfizimler için ayrıntılı olarak incelenir. Polimorfizm; bir popülasyonda çok alelli genin birden fazla fenotip göstermesidir. En az popülasyonun %1'inde görülmesi gerekmektedir. Polimorfizm incelemesi pek çok moleküler tekniğin yanı sıra, 1990-2003 yılları arasında Birleşik Devletler Enerji Bölümü ve Ulusal Sağlık Enstitüsü'nün yürüttüğü İnsan Genomu Projesi'nden elde edilen veri tabanı yardımıyla yapılabilir. Kontrollerde çok az görülen vakalardaki polimorfizimler olası mutasyonlar olarak değerlendirilirler.
8. Gen belirlendiğinde, genin yapısı, her bir egzonun işlevi moleküler yöntemlerle incelenir. Daha sonra araştırmacı çeşitli mutasyonların görülme sıklığını, her bir mutasyonun hastalık üzerindeki etkisini yaşla veya çevreyle etkileşimlerini kestirir.

1.6 Gen Haritalama

Gen haritalama, genlerin belirli kromozomlar üzerindeki yerlerinin belirlenmesidir. Genetik Haritalama ve Fiziksel Haritalama olmak üzere iki farklı gen haritalama yöntemi kullanılmaktadır. Genetik haritalama kromozomdaki genlerin, birbirlerine bağlantılı olup olmamasından yola çıkarak, görel konumlarını inceler. Eğer iki gen birlikte kalıtılıyorsa, bu iki genin bağlantılı olduğu düşünülür. Böylece bağlantılı genlerin görel konumları belirlenebilir. Fiziksel haritalama, belirli bir genin kromozom üzerindeki yerinin kesin olarak bulunmasıdır (Gene Therapy Solutions, 2004).

1.6.1 Bağlantı Analizi

Bağlantı analizi, yaklaşık kromozomal bölgeyi bulmak için kullanılan istatistiksel bir yöntemdir. Hastalıkla yeri bilinen markerlar arasında kosegregasyon (cosegregation) bulunmasına dayanır. Kosegregasyon, iki veya daha fazla genin beraber kalıtılma eğilimini tanımlamaktadır. Aynı zamanda hastalık fenotipini gösteren bireylerin marker lokusun yakınlarında bulunan alelleri paylaşmasıdır.

Bağlantı analizinde sık kullanılan iki yöntem göreceli çift ve LOD skor yöntemleridir.

1.6.1.1 Parametrik Bağlantı Analizi

İki gen arasındaki bağlantı analizinde (hastalık geni ile marker gen olabilir) ilgilenilen parametre rekombinasyon fraksiyonu θ dır. Birbirine tam olarak bağlantılı olan iki gen mayoz sırasında beraber aktarılır ve burada $\theta=0$ dır. Buna karşılık bağlantılı olmayan, farklı kromozomlarda bulunan iki genin aktarım süreci birbirinden bağımsızdır ve $\theta=0.5$ tir. Bir ailedeki hastalık ve marker alellerinin kosegregasyonu (cosegregation) farklı θ değerleri için olabilirlik fonksiyonunda özetlenebilir.

İnsan haritalama çalışmalarında karşılaşılan bağlantı verilerinin olabilirlik fonksiyonları halinde ifade edilmesi çok zordur. 1971'de Elston ve Stewart tarafından geniş ailelerden elde edilen bağlantı verilerinin analizi için geliştirilen algorithmadan önce, bağlantı analizini ortaya atan ilk bilim adamlarından olan Morton (1956) bağlantı verilerinin olabilirlik fonksiyonu şeklindeki ifadesinin çok karmaşık olduğunu göstermiştir. 5 kuşaktan oluşan ailenin olabilirlik ifadesini 60 terimli 20. dereceden polinom (θ) halinde göstermiştir. Bağlantı problemlerini analitik olarak çözmek her zaman

mümkün olmamaktadır. Ancak bağlantı analizini anlayabilmek için analizin matematik temelini özümsemek gereklidir.

Bir ailede tüm mayozlar rekombinant olanlar ve olmayanlar şeklinde ayrılabilirse, olabilirlik fonksiyonu binomial olasılık fonksiyonudur. n gözlenen toplam mayoz sayısı ve r rekombinant sayısı, $n-r$ de rekombinant olmayanların sayısı olsun. Olabilirlik fonksiyonu;

$$L(\theta) = \binom{n}{r} \theta^r (1 - \theta)^{n-r} \quad \text{Denklem 1.2}$$

şeklindedir. Bu fonksiyonu en büyük yapan θ değeri, başka bir ifadeyle gözlenen verilere en iyi uyum gösteren değer, rekombinantların görece sıklıklarıdır. Ancak bağlantı analizinde $\theta > 0.5$, yani iki lokus arasındaki rekombinasyonun rasgele olarak gerçekleştiği durum, biyolojik açıdan konu dışı olarak ele alınır. Bu nedenle tahmin edilen rekombinasyon fraksiyonu şu şekilde tanımlanır:

$$\hat{\theta} = \begin{cases} \frac{r}{n} & \text{eger } \frac{r}{n} \leq \frac{1}{2} \\ 0.5 & \text{eger } \frac{r}{n} > \frac{1}{2} \end{cases} \quad \text{Denklem 1.3}$$

θ tahmin edildikten sonra bağlantı olmadığı ($\theta=0.5$) hipotezi test edilir. %50'den farklı olarak gözlenen rekombinasyon oranının istatistiksel olarak anlamlı olup olmadığı incelenir. Bunun için denklem 1.4'teki test istatistiği kullanılmalıdır:

$$Z(\theta) = \log \frac{L(\theta)}{L(0.5)} \quad \text{Denklem 1.4}$$

LOD skoru olarak bilinen bu test istatistiği bağlantı analizi içindeki en önemli kavramdır. Haldane ve Smith (1947) tarafından insan çalışmaları için

ortaya atılmıştır. Pozitif LOD skorları bağlantı için kanıt oluştururken, negatif skorlar bağlantı olmadığını gösterir. $Z(\theta) > 3$ bağlantı sonucuna ulaşmak için yeterli görülürken, $Z(\theta) < -2$ bağlantı olmadığını, yokluk hipotezini, kabul etmek için yeterlidir (Morton, 1955).

$\theta = \hat{\theta}$ varsayımı altında 3 olarak bulunan LOD skoru, gözlenen veriden elde edilen rekombinasyon fraksiyonunun, yokluk hipotezi olan $\theta=0.5$ 'ten $10^3 = 1000$ kez daha olası olduğunu göstermektedir. Yani bağlantı için odds 1000:1 dir. Bu sonucun p değeri 0.0001 dir. Çok düşük anlamlılık düzeyindeki bağlantı çalışmaları için çoklu testler önerilmektedir. X kromozomu üzerindeki herhangi bir lokustaki anlamlı bağlantı için LOD skorunun 2 olması yeterlidir.

Denklem 1.4 açılırsa;

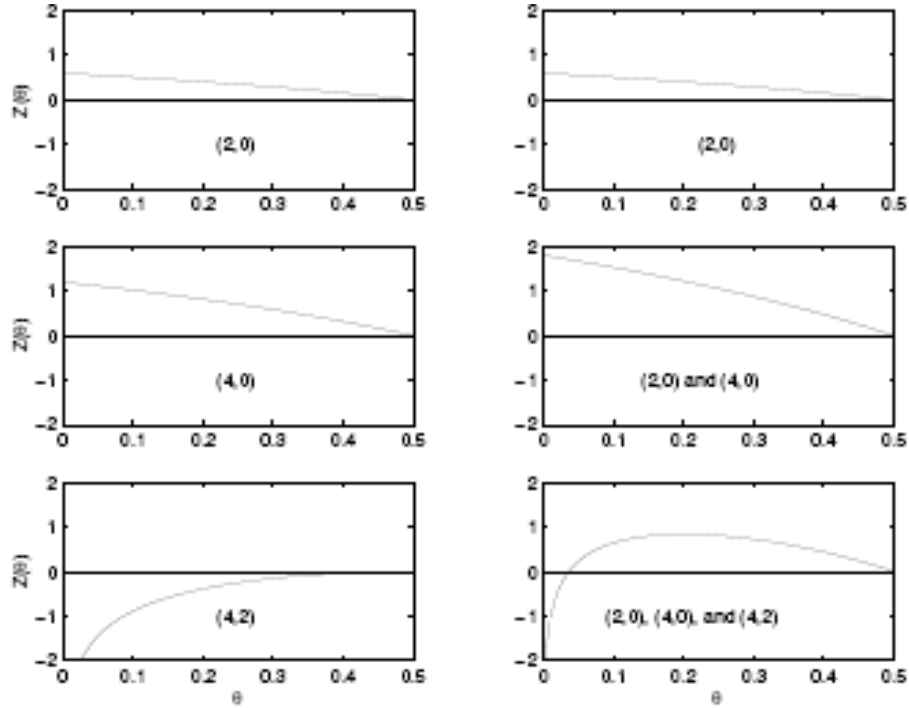
$$\begin{aligned}
 Z(\theta) &= \log\left(\frac{\binom{n}{r} \theta^r (1-\theta)^{n-r}}{\binom{n}{r} 0.5^r (1-0.5)^{n-r}}\right) \\
 &= \log\left(\frac{\theta^r (1-\theta)^{n-r}}{0.5^n}\right) \\
 &= \log \theta^r + \log(1-\theta)^{n-r} - \log 0.5^n \\
 &= r \log \theta + (n-r) \log(1-\theta) + n \log 2
 \end{aligned}$$

Denklem 1.5

elde edilir. LOD skor fonksiyonunun elverişli yanı bağımsız aileler için toplamsal olmasıdır. Aile kümelerinde sabit bir θ için toplam LOD skoru ailelerin LOD skorlarının toplanmasıyla elde edilir. Bu durum Grafik 1.1'de gösterilmiştir.

(n,r) her bir aileden elde edilen n tane mayozu ve r tane rekombinasyonu göstermektedir. Aileler bazında elde edilen mayoz ve rekombinasyon sayılarına göre, sol taraftaki grafikler her bir ailenin LOD skor

fonksiyonlarını verirken, sağ tarafta sırasıyla 1., 1. ve 2., 1., 2. ve 3. ailelerin toplam LOD skor grafikleri çizdirilmiştir.



Grafik 1.1 LOD Skor Grafikleri

Marker lokusuyla hastalık lokusu arasında bir rekombinasyon gözlemlendiği zaman, iki lokus arasındaki uzaklığın 0'dan büyük olduğu söylenebilir. Grafiklerden anlaşılacağı gibi, mayozda rekombinasyon gözlemlendiği zaman $\theta=0$ 'da fonksiyon $-\infty$ 'a gitmektedir.

Buraya kadar analizler mayozların rekombinasyon olanlar ve olmayanlar şeklinde ayrılabilirdiği varsayımına dayandırılmıştı. Ancak her zaman bu ayırım yapılamamaktadır. Marker ve/veya hastalık lokusunun homozigotluğu, fenokopiler (aynı çevre koşulları altında bir bireyin fenotipinin, genotipleri farklı olduğu halde diğer bir bireyin fenotipiyle aynı olması durumu), tamamlanmamış penetrans, tanının kesin olmaması, bilinmeyen kalıtım kalıbı (genetik karakterin ya da hastalığın kuşaktan kuşağa aktarılma şekli. Örneğin, otozomal dominant, otozomal resesif), rekombinasyon

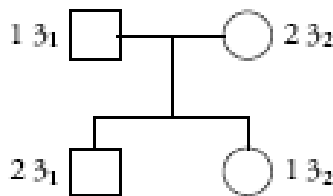
fraksiyonunun cinsiyete göre farklılık göstermesi, kayıp marker ya da fenotip verisi gibi durumlarla karşılaşıldığında analiz yapmak zorlaşmaktadır.

1.6.1.2 Parametrik Olmayan Bağlantı Analizi

Parametrik bağlantı analizi alel sıklıklarının ve penetrans parametrelerinin bilindiği durumda uygulanabilir. Ancak, daha karmaşık, çok genli hastalıklarda, sadece genler değil çevre de etkilidir. Bu nedenle genetik model oluşturmak zordur. Bu tip durumlar için parametrik olmayan bağlantı analizi yöntemleri geliştirilmiştir. Birden fazla hasta bireyi bulunan aile ele alınır. Hasta veya sağlıklı olmak üzere 2 çeşit fenotip gözlenmektedir. Çekirdek ailelerden elde edilen genotiplerde, çocukların paylaştıkları aleller iki durumda incelenir. Hasta bireyler aynı kurucu bireyden aldıkları hastalık alelini paylaşıyorsa buna ebeveynlerden dolayı aynı (identical by descent), IBD olma durumu denir. Test edilen lokusun yakınlarındaki markerlardan elde edilen bilgiler doğrultusunda o lokustaki kalıtım deseni tahmin edilebilir ve hasta bireylerdeki IBD deseninin rasgele olarak ortaya çıkma olasılığı test edilebilir.

Hasta kardeşlerden elde edilen veriler parametrik olmayan bağlantı analizinde sıklıkla kullanılır. Çünkü bu türdeki veriyi toplamak ve parametrik olmayan bağlantı analizi yöntemini iki kişiden oluşan küçük aileye uygulamak kolaydır. Bir kuşakta aynı alel hangi ebeveynden geldiğine bakılmaksızın paylaşılıyorsa, aynı olma (identical by state) (IBS) durumu denir.

Örneğin;



Şekil 1.4 Alel Paylaşımı

Şekil 1.4'te 3 alelli bir sistemde aile bireylerinin genotipleri görülmektedir. Çocukların bir ortak aleli vardır. Ancak 3 numaralı alel olan bu ortak alel erkek çocuğa babasından geçmişken, kız çocuğa annesinden geçmiştir. Bu durumda IBD, yani ebeveynlerden dolayı aynı olan alel sayısı 0 iken, IBS sayısı 1 dir.

IBD alel paylaşımı bağlantı analizi için önemi bir hususu oluştururken, IBS alel paylaşımı aynı şekilde bilgi vermemektedir. Çünkü iki alel aynı atadan gelmesi de IBS olabilmektedir.

Akraba çiftlerinden elde edilen genotip bilgisine dayanan yöntemler, benzer fenotipe sahip çiftlerin marker alelleri paylaşma sıklıklarının rasgele olarak beklenenden daha fazla olup olmadığını araştırır. Bu yöntemlerle rekombinasyon fraksiyonu tahmin edilemez. Ancak hipotez testleri uygulaması oldukça kolaydır. Bu nedenle bu yöntemlerle çok sayıda marker lokusunu bağlantı için kısa sürede taramak mümkündür. Ancak ince haritalama için önerilmez. Çünkü bu yöntemler penetransla ilgili bir varsayım içermezler ve parametrik değildirler. Bu nedenle LOD skoru yönteminden daha zayıf testlerdir. Bağlantı analizinde uygulanacak testin seçimi hastalık modelinin bilinmesine bağlıdır. Eğer hastalık modeli bilinmiyorsa parametrik olmayan testler kullanılır. Bu tür testler modelden bağımsız testler olarak adlandırılır (Whittemore, 1996). Dikotom (iki sonuçlu) hastalık fenotipleri için, mutant (ışınımlara maruz kalmış ve bozulmuş) alellerin çok ender görüldüğü varsayımına dayanarak, sadece hasta çiftler kullanılır. Hasta-sağlıklı ya da sağlıklı-sağlıklı bireylerden oluşan çiftler çalışmaya alınmaz. Eğer ebeveynlerden biri homozigotsa, çocukların paylaştığı alellerin IBD olup olmadığına karar verilemez.

Aleller arasındaki IBD ilişkisinin bilinmesi; $H_0 : \theta = 1/2$ altında iki akrabanın, akrabalık ilişkilerine göre bir veya iki alellerini paylaşma

olasılıklarının sabit olmasını sağlar. Tablo 1.1’de akrabalık dereceleri bilinen çiftlerin alellerini IBD olarak paylaşma olasılıkları verilmiştir.

Tablo 1.1 IBD Olarak Alel Paylaşım Olasılıkları

Akrabalık Derecesi	IBD Alel Paylaşım Olasılıkları		
	π_0	π_1	π_2
Tek yumurta ikizleri	0	0	1
Kardeşler	1/4	1/2	1/4
Ebeveyn-Çocuk	0	1	0
1. derece kuzenler	3/4	1/4	0
Büyük ebeveyn-torun	1/2	1/2	0

Alellerin IBS olarak paylaşımında kurulan yokluk hipotezi popülasyondaki alel sıklıklarına dayandığı için IBD’den daha karmaşıktır. Sadece a ve b alellerinin bulunduğunu varsayalım. Alel sıklıları $p=(p_a, p_b)$ olsun. Eğer $p_a > p_b$ ise, iki kardeşin a alelini paylaşma olasılıkları b alelinden daha yüksektir. Ebeveynlerin genotipleri bilinmiyorsa, IBS olarak paylaşılan alellerin yokluk olasılık dağılımları hangi alelin paylaşıldığına bağlıdır ve ebeveynlerin olası tüm genotip kombinasyonlarının toplanmasıyla hesaplanır. M_p ebeveyne bağlı marker genotipini, M_o da çocuğun marker genotipini versin. O halde, IBS olarak n tane alelin paylaşılması olasılığı Denklem 1.6’daki gibi hesaplanır. (Thomas, 2004)

$$\begin{aligned}
 P(IBM = n) &= \sum_{M_p} P(M_p | p) P(IBM = n | M_p) \\
 &= \sum_{M_p} P(M_p | p) \sum_{M_o} P(M_o | M_p) I(IBM = n | M_o)
 \end{aligned}$$

Denklem 1.6

1.6.1.2.1 Hasta Kardeşler Yöntemi

Bu yöntem fenotipleri benzer olan, dolayısıyla hastalık lokusundaki bir veya iki aleli paylaşma olasılıkları yüksek olan iki kardeşin benzer marker alellerini paylaşmaya eğilimli olduklarını belirten yokluk hipotezini test eder. Eğer hastalık ender görülüyorsa, hasta bireyin mutant alele sahip olma olasılığı herhangi bir bireyden fazladır. Ancak sağlıklı bireyin mutant aleli olması olasılığı herhangi bir bireyinkinden çok da farklı değildir. Bu nedenle dikotom hastalıklar için en çok kullanılan akraba çiftleri hasta kardeşlerdir. Rasgele olarak seçilen kardeş çiftlerinin IBD olarak hiç alel paylaşmama, yalnız bir alel paylaşma ve 2 alel paylaşma olasılıkları sırasıyla 1/4, 1/2 ve 1/4 olarak verilmiştir. Eğer çalışılan markerla hastalık lokusu arasında bir bağlantı yoksa, bu olasılıklar bireylerin hastalıklarının konumlarına bağlı olmaksızın doğrudur. Eğer iki kardeş de hastaysa ve marker lokusu ile hastalık lokusu bağlantılıysa, şans eseri olarak beklenen IBD alel paylaşımından daha fazla alel paylaşımı beklenir. Tablo 1.2’de beklenen ve gözlenen alel paylaşım dağılımları verilmiştir.

Tablo 1.2 Beklenen ve Gözlenen Alel Paylaşım Dağılımları

	IBD olarak paylaşılan alel sayısı			
	0	1	2	Toplam
Gözlenen	A	B	C	N
Beklenen	N/4	N/2	N/4	N

Hastalık lokusunun belirlenen bir x genomu ile bağlantılı olup (H_0) olmadığını (H_1) test etmek için, Tablo 1.2’deki verilere dayanarak hesaplanan test istatistiği;

$$\chi^2 = (A - C)^2 / (A + C) \sim \chi^2_{1,\alpha} \quad \text{Denklem 1.7}$$

dir. Alternatif olarak Blackwelder ve Elston (1985) tarafından geliştirilen test, IBD olarak paylaşılan alellerin ortalama oranlarını karşılaştıran t testidir. $H_0: \bar{\pi} = 1/2$ olarak belirlenir. $\bar{\pi} = \pi_1 / 2 + \pi_2$ dir. Kestirimi; $\hat{\bar{\pi}} = (B + 2C) / 2N$ standart hatası; $SE(\hat{\bar{\pi}}) = \sqrt{(AB + 4AC + BC) / 4N^3}$. Serbestlik derecesi N-1 dir. Ki kare testinden biraz daha güçlü olduğu bilinmektedir. (Thomas, 2004)

1.6.1.2.2 Hasta Akraba Çiftleri Yöntemi

Teorik olarak hasta kardeşler yöntemi, hasta her akraba çiftine genellenebilir. Tablo 1.2'deki beklenen değerler hesaplanırken, akrabalık ilişkisine göre Tablo 1.1'deki katsayılar kullanılır. Ve bir önceki bölümde sözü geçen ortalamalar testi kullanılır.

Uygulamada uzak akrabaların alelleri arasındaki IBS ilişkisinin tespiti IBD ilişkisinin tespitinden daha kolaydır. Popülasyona ait marker alel sıklıkları biliniyorsa her iki türdeki ilişki için de olasılık tabloları oluşturulabilir. Ancak genellikle popülasyonda görülen alel sıklıkları bilinmez ve uygulanacak olan verilerden kestirilmeye çalışılır. Sadece ikili hasta akrabaları kullanmak yerine, daha geniş ailelerdeki hasta tüm akrabaları çalışmaya dahil etmek daha güçlü sonuçlar verecektir. (Whittemore ve Halpern, 1994)

1.6.1.2.3 Kantitatif Karakterler için Kardeş Çifti Yöntemi

Tek genin sorumluğu olduğu karakterler aktarılırken Mendel Kalıtım deseni gösterirler. Ancak birden fazla genin sorumlu olduğu karakterler bu kalıtıma bağlı olarak aktarılmazlar. Bu genlerin birbirleriyle ve çevreyle olan ilişkileri kalıtım desenini belirler. Bu tür karakterlere kantitatif karakterler adı verilir.

Kantitatif karakterler için test edilecek hipotez şu şekilde kurulur: Ortak alellerini IBD olarak paylaşan kardeş çiftlerinin (veya diğer akraba çiftlerinin) fenotipleri paylaşmayanlara göre daha fazla benzerlik gösterir. Eğer her iki alelleri de ortak ve ikisi de IBD olarak paylaşılıyorsa, bu durumda fenotiplerindeki benzerlik çok daha fazladır. Bu test yöntemi Haseman ve Elston (1979) tarafından geliştirilmiştir. Hasta kardeş çiftleri yönteminde olduğu gibi IBD alel paylaşımını belirlemek her zaman mümkün olmamaktadır. Bu nedenle Haseman ve Elston denklem 1.8'deki regresyon denklemini geliştirmişlerdir.

$$E(D_i) = \beta \bar{\pi}_i + \gamma(\pi_{i1} - 1/2) \quad \text{Denklem 1.8}$$

$D_i = (Y_{i1} - Y_{i2})^2$ dir. Bu denklemdeki katsayıların beklenen değerleri ise;

$$\begin{aligned} E(\beta) &= 2\sigma_A^2(1 - 2\theta)^2 \\ E(\gamma) &= \sigma_D^2(1 - 2\theta)^4 \end{aligned} \quad \text{Denklem 1.9}$$

şeklinde hesaplanır.

Genetik varyans; σ_A^2 toplamsal genetik varyans (her bir alelin toplamsal etkisi), σ_D^2 aynı lokusta bulunan alellerin etkileşimi ve σ_I^2 farklı lokuslardaki etkileşimi bileşenlerine ayrılabilir. Denklem 1.9'da kullanılan varyanslar genetik varyansın bileşenlerindendir.

1.6.2 İlişki Analizi

Genetik bağlantı, kısa kromozomal parçaların bozulmamış şekilde ebeveynden çocuğa aktarılma eğilimidir. Bağlantı sonucunda bu kısa kromozomal parçalardaki alel kombinasyonları nesiller boyunca taşınarak korunabilirler. Alellerin genetik kosegregasyonu, lokuslar arasındaki genetik uzaklık ne kadar az ise o kadar fazla olmaktadır. Belli bazı haplotiplerin

nesilden nesile çok fazla tekrar etmesi sıkı bağlantıdan ileri gelebilir. Bu duruma alelik ilişki adı verilir. Bağlantı analizi, daha az markerla karakterden sorumlu lokusu bulmak için tüm genom boyunca yapılabilir. Ancak ilişki analizi, ilgilenilen karakterden sorumlu olabileceği düşünülen alel için yapılan analizdir. Bağlantı analizinin aksine, ilişki analizinde genetik lokusların haritalanması önemlidir. İlişkilendirme analizi gelecek için çok şey vaat etmektedir. Gelecekte genomdaki her lokusun hastalık karakteri taşıyıp taşımadığının ilişki analiz yöntemleriyle test edilebileceği öngörülmüştür (Risch ve Merikangas, 1996).

Her birinde sırasıyla $A_1, A_2, \dots, A_m$, ve $B_1, B_2, \dots, B_m$, alelleri bulunan A ve B lokuslarını ele alalım. Alellerin göreceli olarak popülasyonda görülme sıklıkları sırasıyla $p_1, p_2, \dots, p_m$, ve $q_1, q_2, \dots, q_m$, olsun. Bu lokuslar için toplam $m \times m$ tane olası haplotip yazılabilir ($A_1B_1, A_1B_2, \dots, A_mB_m$). Her bir haplotipin göreceli sıklıkları da $h_{11}, h_{12}, \dots, h_{mm}$, olsun. Eğer bir haplotipte A_i alelinin görülmesiyle B_j alelinin görülmesi birbirinden bağımsız olaylarsa, aynı haplotip içinde görülme sıklıkları marjinal sıklıklarının çarpımına eşittir.

$$h_{ij} = p_i \times q_j \quad \text{Denklem 1.10}$$

Eğer bu denklem sağlanamıyorsa, aleller ilişkilidir denir.

θ , A ve B lokusları arasındaki rekombinasyon fraksiyonu, h_{ij0} ise A_iB_j haplotipinin şimdiki kuşakta görülme sıklığı olsun. Gelecek kuşaktaki A ve B lokusları için her bir haplotip ya rekombinanttır (θ olasılığıyla) yada rekombinant değildir ($1-\theta$ olasılığıyla). Eğer haplotip rekombinantsa, A_iB_j olma olasılığı h_{ij0} iken, eğer rekombinant değilse bu olasılık rasgele eşleşme varsayımı altında marjinal frekansların çarpımına eşittir. Bu nedenle A_iB_j haplotipinin bir sonraki kuşağa aktarılma olasılığı;

$$h_{ij1} = (1 - \theta)h_{ij0} + \theta p_i q_j \quad \text{Denklem 1.11}$$

O halde gelecek kuşak ile şimdiki kuşak arasında bir haplotipin görülme olasılığındaki değişim;

$$h_{ij1} - h_{ijo} = \theta(p_i q_j - h_{ijo}) \quad \text{Denklem 1.12}$$

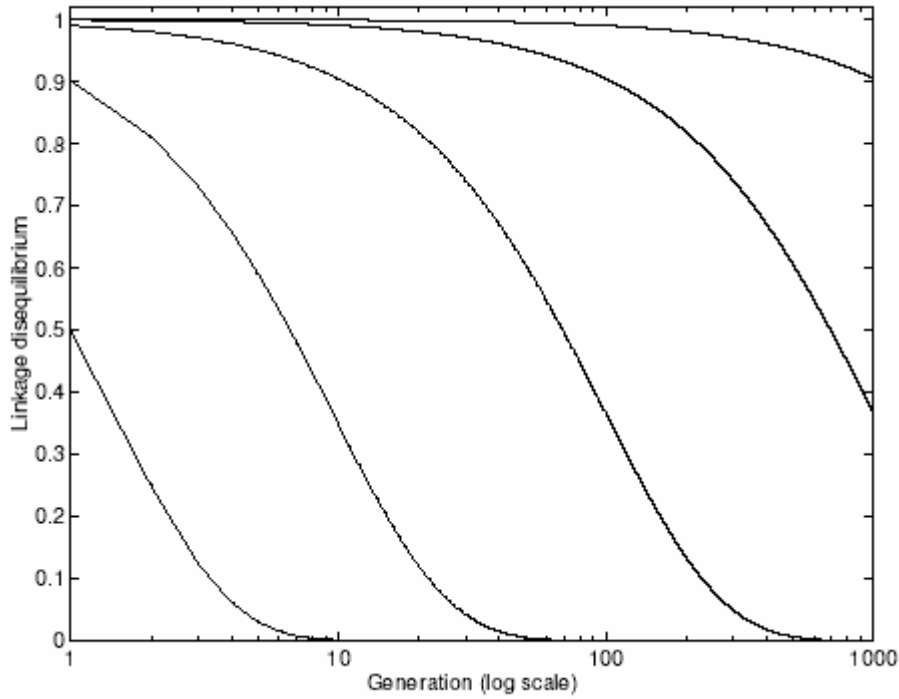
Şimdiki kuşakta alelik ilişki yoksa haplotip sıklığı değişmeyecektir. Eğer ilişki varsa, değişim rekombinasyon fraksiyonu ile orantılıdır. Kuşaklar arası haplotip sıklıklarında değişim gözlenmesi, iki lokusun bağlantı dengesizliğinde olduğunu gösterir. Rasgele eşleşmenin olduğu popülasyonlarda bağlantı dengesine ulaşma hızı rekombinasyon fraksiyonuna bağlıdır. Denklem 1.11'de gerekli matematiksel işlemler yapıldıktan sonra;

$$h_{ij1} - p_i q_j = (1 - \theta)(h_{ij0} - p_i q_j) \quad \text{Denklem 1.13}$$

eşitliği elde edilir. Buradan görülmektedir ki her bir kuşakta haplotip sıklığı ve onun dangedeki değeri arasındaki fark $(1-\theta)$ kadar azalmaktadır. O halde k kuşak sonra;

$$h_{ijk} - p_i q_j = (1 - \theta)^k (h_{ij0} - p_i q_j) \quad \text{Denklem 1.14}$$

Görelî haplotip sıklığı ve dangedeki değeri arasındaki fark aleller arasındaki ilişkinin büyüklüğünü ölçmek için kullanılabilir. Kuşaklar arasında bağlantı dengesizliğinin düşüşü Grafik 1.2'de görülmektedir.



Grafik 1.2 Kuşaklar arası bağlantı dengesizliği

Grafik sola alttan sağ üste doğru sırasıyla 0.5, 0.1, 0.01, 0.001 ve 0.0001 rekombinasyon fraksiyonlarına göre çizilmiştir.

İlk hastalık mutasyonu belirli bir kromozom üzerinde yer alır ve aynı kromozom üzerinde yakın lokuslardaki alellerle ilişkilidir. Bu ilişki kuşaklar boyunca meydana gelen rekombinasyonlar sonucu bozulur. İlişki çalışmalarında, çoğunlukla hastalık lokuslarıyla sıkı bağlantı içinde olduğu, yani ilişkili olduğu, düşünülen markerlar kullanılır. Ancak alelik ilişkisinin pek çok nedeni olabilir. Bağlantı dengesizliği ve alelik ilişkisi; rasgele genetik sürüklenme (genetic drift), mutasyon, seçilim, popülasyon karışımı (population admixture) ve tabakalaşma sonucu ortaya çıkar. Seçilim; bir bireyin genotipi sağlıklı üreme üzerinde etkili olduğunda ortaya çıkar. Popülasyon karışımı ve tabakalaşma ise popülasyon çeşitli kültürel ve coğrafi nedenlerden dolayı iki veya daha fazla alt gruba ayrılmışsa görülen bir durumdur. Her alt grupta birbiriyle bağlantı dengesizliğinde olan iki lokus popülasyonun tamamıyla da dengesizlik halinde olabilir.

Rasgele genetik sürüklenme, sonlu bir popülasyonda kuşaktan kuşağa, alel ve haplotip dağılımlarında meydana gelen rasgele değişimi anlatır. Bir kuşağın gen havuzu, bir önceki kuşağın gen havuzundan alınan örneklem gibidir. Bu nedenle alel ve haplotiplerin göreceli sıklıkları örnekleme varyansına bağlıdır. Aynı zamanda, meydana gelen mutasyonlar alel ve haplotip sıklıklarını kuşaktan kuşağa değiştirebilirler. Popülasyon ne kadar küçük olursa, oluşan mutasyonların ve örnekleme varyansının etkisi o kadar büyük olur. Sabit bir popülasyonda iki lokus arasındaki bağlantı dengesizliğinin büyüklüğü popülasyonun büyüklüğüne, rekombinasyon fraksiyonuna ve mutasyon hızına bağlıdır. Seçilim, popülasyon karışımı ve tabakalaşma popülasyon büyüklüğünden bağımsız olarak bağlantı dengesizliğine yol açabilirler.

Bağlantı dengesinde bulunan 3 alt popülasyonu ele alalım. Tablo 1.3'te popülasyonların büyüklükleri, alel ve haplotip göreceli sıklıkları verilmiştir:

Tablo 1.3 3 alt popülasyonun alel ve haplotip göreceli sıklıkları

Büyüklik (N)	A_i alelinin görülme olasılıkları	B_j alelinin görülme olasılıkları	$A_i B_j$ haplotipinin görülme olasılıkları
N_1	p_1	q_1	$p_1 q_1$
N_2	p_2	q_2	$p_2 q_2$
N_3	p_3	q_3	$p_3 q_3$

Bu 3 alt popülasyon birleştiğinde, alt popülasyonlar arası üreme olmadığı varsayılarak, alel ve haplotip sıklıkları Denklem 1.15'teki gibi olacaktır:

$$P(A'_i) = \sum_{i=1}^3 A_i N_i / \sum_{i=1}^3 N_i$$

$$P(B'_i) = \sum_{i=1}^3 B_i N_i / \sum_{i=1}^3 N_i$$

Denklem 1.15

$$P(A'_i B'_i) = \sum_{i=1}^3 A_i B_i N_i / \sum_{i=1}^3 N_i$$

Eğer dengedeki görelî sıklık $P(A'_i B'_i)$ alel olasılıklarının çarpımına eşit değilse, birleştirilen popülasyonda A ve B alelleri ilişkilidir denir.

İki lokus arasında bulunan ilişki, bir lokustaki alelin diğer lokustaki alele bağlı olarak görülmesinden kaynaklanır. Marker lokus ile şüphelenilen hastalık lokusu arasındaki bağlantı, ilişkiye yol açar. Hasta bireylerde marker alellerinin sağlıklı bireylerden daha fazla görülmesi popülasyon düzeyindeki alelik ilişkiyi gösterir. Bu ilişki bağlantıyla alakalı olmayabilir. Bağlantı dengesizliği sıklıkla alelik ilişkiyle eş anlamda kullanılmaktadır. Ancak bu durumda ilişkiye neden olacak etmenler gözden kaçabilmektedir. Bu nedenle gamet aşaması dengesizliği terimi önerilmiş (Almgren ve ark. 2002). Bağlantı dengesizliği sadece bağlantı durumunda, rekombinasyon olmadığı zaman, alelik ilişkilendirme anlamına gelir. Bağlantı dengesizliğinden kaynaklanmayan alelik ilişki lokus haritalaması için kullanılamaz.

Bağlantı analizi kromozomal bölgede hastalık lokusunun tespiti için önemli bir analizdir. Ancak ince haritalama için çok kullanışlı değildir. Çünkü rekombinasyon fraksiyonundaki küçük farklılıkların ayırt edilebilmesi için çok sayıda bilgi veren gametten oluşan veri gerekmektedir. İnce haritalama çalışmalarında ilişki analizinden sonra bağlantı analizi yapılabilir. Çoğu popülasyonda güçlü bağlantıya bağlı olan alelik ilişki, lokuslar arasındaki rekombinasyon fraksiyonları %1'den az olduğunda ortaya çıkar. Çoğu ilişki yöntemi ya bağlantı olduğu durumda ilişki testi ya da ilişki olduğu durumda bağlantı testi şeklinde kullanılır. Alelik ilişki olmadığında, ince haritalama için

çok büyük örneklem genişlikleri gerekirken, bağlantı dengesizliğinde daha küçük örneklem genişlikleri yeterlidir.

İki lokus arasındaki alelik ilişki, bu lokuslar arasında güçlü bir bağlantı olduğunun göstergesidir. Ancak alelik ilişki, popülasyon karışımı ve tabakalaşma olduğunda bağlantı olmasa da bulunabilir. Bu nedenle alelik ilişki çalışmalarının tasarımında, analizinde ve sonuçlarının yorumlanmasında bu hususlar göz önüne alınmalıdır.

En eski ve en basit ilişki analizi yöntemi vaka-kontrol çalışmalarıdır. Hasta olan ve olmayan bireylerden, iki örneklem rasgele olarak alınır. Marker alelin hasta bireylerde, sağlıklı olanlara nazaran daha sık görülüp görülmediği test edilir. Eğer hastalık, marker aleli taşıyan kromozomda mutasyon olarak ortaya çıkmışsa, istatistiksel olarak anlamlı bulunan ilişki marker ve hastalık lokusları arasındaki bağlantı dengesizliğinden kaynaklanabilir. Ancak örneklem heterojen, birden fazla tabakadan oluşan bir popülasyondan alınmışsa ve tabakalar hastalık-marker alellerinin dağılımlarına göre farklılık gösteriyorsa, tabakalarda ilişki olmadığı halde genel popülasyonda hastalık-marker ilişkisi bulunabilir. Bu tür ilişkinin biyolojik olarak bir anlamı olmadığı halde, popülasyondaki heterojenlikten kaynaklanan ilişki vardır.

Heterojenlikten kaynaklanan ilişkiyi engellemenin üç yolu vardır. İlki; uygulaması zor olsa da homojen popülasyonlardan örnek alınmalıdır. İkinci yol; etniklik gibi birlikte değişen değişkenleri de eklemektir, ancak uygun olan genetik ya da çevresel değişkenlerin tespit edilmesi her zaman mümkün olmamaktadır. Üçüncü yol ise eşleşmiş kontroller kullanmaktır. Eğer diğer genetik faktörler ilişkiye neden oluyorsa, etnik eşleme yapılması gereklidir. Eşleme yapabilmenin bir yolu kontrolleri aileden seçmektir. En çok kullanılan aile bazlı ilişki analizi yöntemi Kalıtımda Dengesiz Aktarım Testi (TDT) 'dir.

1.6.2.1 Aile Bazlı İlişki Analizi Yöntemleri

Aile bazlı ilişki çalışmaları bağlantı çalışmaları ve vaka-kontrol ilişki çalışmalarının birleştirilmesiyle oluşmuştur. Aile bazlı ilişki analizinde kullanılan tüm yöntemler ebeveynden geçirilen alellerle, ebeveynden hasta çocuğa geçirilmeyen veya sağlıklı çocuğa geçirilen alellerin karşılaştırılması esasına dayanır.

1.6.2.1.1 Kalıtımda Dengesiz Aktarım Testi (TDT)

Kalıtımda dengesiz aktarım testi, (Spielman ve ark., 1993) marker ve hastalık lokusları arasında ilişkinin olduğu bilindiğinde, markerla yakınlarında bulunan aday gen arasındaki bağlantıyı test eder. TDT yöntemi, aileler arasında alellerin rasgele olarak dağılımdan ayrılmasını incelerken, IBD ve parametrik bağlantı analizi yöntemleri alellerin rasgele olarak dağılımdan ayrılmasını aile içinde inceler. Diğer bir deyişle; TDT, belli bir marker aleli ile hastalık aleli arasındaki bağlantıya, diğer bağlantı testleri ise belli bir marker lokusu ile hastalık lokusu arasındaki bağlantıya odaklanır.

Aile üçlüsü, anne-baba ve hasta çocuktan oluşur ve bütün bireylerin marker genotipleri bilinir. Bir aile üçlüsü ele alalım. Segregasyon bozulması olmadığını varsayalım. Yani belli bir alelin çocuğa geçirilmesi rasgele olarak gerçekleşsin. Hastalık lokusunda D_1 hastalık aleli ve D_2 normal alel olmak üzere iki alel olsun. Burada D_1 ve D_2 birer alel değil, birer alel grubu da olabilir. Marker lokustaki alellerin ebeveynlerden çocuğa geçişlerinin farkı popülasyonda var olan alelik ilişkinin ve bağlantının göstergesidir.

Marker lokusta sadece iki alel varsa; $2n$ ebeveynin n tane hasta çocuğa marker lokusundaki M_1 ve M_2 alellerini geçirme sayıları Tablo 1.4'te verilmiştir.

Tablo 1.4 2n ebeveynin n hasta çocuğa geçirdiği alel sayıları a,b,c ve d

	<i>Geçirilmeyen Alel</i>		
<i>Geçirilen Alel</i>	M ₁	M ₂	Toplam
M ₁	a	b	a+b
M ₂	c	d	c+d
Toplam	a+c	b+d	2n

Sadece marker lokuslarında heterozigot olan, M₁M₂ alellerini taşıyan, ebeveynlerin rekombinasyon fraksiyonu ile ilgili bilgi sağlayabilecekleri düşünülmektedir (Almgren ve ark. 2002). Test istatistiği;

$$TDT = \frac{(b - c)^2}{(b + c)} \sim \chi_1^2 \quad \text{Denklem 1.16}$$

p D₁ hastalık alelinin görülme olasılığı, q da M₁ marker alelinin görülme olasılığı olsun. M_iD_j haplotipinin görülme olasılığı; $P(M_i D_j) = h_{ij}$ olarak verilsin. D₁ ve M₁ alelleri arasındaki ilişkinin ölçütü, δ ;

$$\delta = P(M_1 D_1) - P(M_1)P(D_1) = P(M_1 D_1) - pq \quad \text{Denklem 1.17}$$

p, q ve δ 'ya bağlı haplotip olasılıklarının fonksiyonları;

$$\begin{aligned} h_{11} &= P(M_1 D_1) = pq + \delta \\ h_{12} &= P(M_1) - h_{11} = q(1 - p) - \delta \\ h_{21} &= P(D_1) - h_{11} = p(1 - q) - \delta \\ h_{22} &= P(M_2) - h_{21} = (1 - q)(1 - p) + \delta \end{aligned} \quad \text{Denklem 1.18}$$

Sırasıyla D₁D₁, D₁D₂ ve D₂D₂ genotiplerine sahip olan bireylerin göreceli sıklıkları, penetransları, f_{11} , f_{12} ve f_{22} 'dir. Eşleşmelerin rasgele olduğu varsayılır. Tablo 1.4'teki verilerin olasılıkları Tablo 1.5'te verilmiştir.

Tablo 1.5 Geçirilen ve geçirilmeyen alellerin kombinasyonlarının olasılıkları

Geçirilen Alel	Geçirilmeyen Alel	
	M ₁	M ₂
M ₁	$q^2 + Bq\delta / p$	$q(1 - q) + B(1 - \theta - q)\delta / p$
M ₂	$q(1 - q) + B(\theta - q)\delta / p$	$(1 - q)^2 - B(1 - q)\delta / p$

Tablo 1.5'te, iki alelli sistemde, hasta çocuklara ebeveynlerden geçen ve geçmeyen alellerin tüm kombinasyonlarının olasılıkları verilmiştir. θ , marker ve hastalık lokusları arasındaki rekombinasyon fraksiyonu ve $B = p[p(f_{11} - f_{12}) + (1 - p)(f_{12} - f_{22})][p^2 f_{11} + 2p(1 - p)f_{12} + (1 - p)^2 f_{22}]^{-1}$ dir.

Olasılıkların Çıkarımı

Ebeveynler F_1 ve F_2 ile çocuk C ile gösterilmektedir. A çocuğun hasta olduğunu belirtir. F_1 ebeveyninin genotipinin $M_i M_j$ olma ve M_i alelini hasta çocuğa geçirme olayı $T_{ij} = \{F_1 = M_i M_j, F_1 \rightarrow M_i\}$ şeklinde gösterilir.

Bayes teoremine göre;

$$P(T_{ij}|A) = P(A|T_{ij})P(T_{ij}) / P(A) \quad \text{Denklem 1.19}$$

yazılır.

Bayes Teoremi, olasılık teorisinin bir sonucudur. Rasegele değişkenlerin koşullu ve marjinal olasılık dağılımlarının ilişkilendirilmesi sonucu ortaya çıkmıştır. A olayının B olayına koşullu olasılığı, başka bir deyişle B olayı biliniyorken A'nın olma olasılığı, $P(B) \neq 0$ iken;

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{Denklem 1.20}$$

olarak hesaplanır. Bayes kuralı ise B olayı olduğunda A olayının olma olasılığını, farklı bir olasılığa sahip olan A olayı olduğunda B olayının olma olasılığı cinsinden hesaplamaya olanak verir (Armitage ve Colton, 1998).

$$P(A|B) = P(B|A)P(A) / P(B) \quad \text{Denklem 1.21}$$

A_k ; birlikte bir örneklem uzayını oluşturan, ayrışık $A_1, A_2, \dots, A_n$ olaylarını temsil etsin. B de aynı uzayda bir olay ve $P(B) > 0$ olsun. Bu durumda Bayes kuralı;

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)} \quad \text{Denklem 1.22}$$

Denklem 1.19'da verilen koşulsuz olasılıklar, $P(T_{ij})$;

$$P(T_{ij}) = P(M_i M_j)P(F_1 \rightarrow M_i | M_i M_j) = \begin{cases} q^2 & i = j = 1 \\ q(1-q) & (i, j) = (1,2) \text{ veya } (2,1) \\ (1-q)^2 & i = j = 2 \end{cases}$$

Denklem 1.23

Çocukların hastalık genotiplerine göre;

$$\begin{aligned} P(A) &= P(A|D_1 D_1)p^2 + P(A|D_1 D_2)2p(1-p) + P(A|D_2 D_2)(1-p)^2 \\ &= f_{11}p^2 + 2f_{12}p(1-p) + f_{22}(1-p)^2 \end{aligned} \quad \text{Denklem 1.24}$$

$$\begin{aligned}
P(A|T_{ij}) &= P(A|C = D_1 D_1, T_{ij})P(C = D_1 D_1|T_{ij}) \\
&\quad + P(A|C = D_1 D_2, T_{ij})P(C = D_1 D_2|T_{ij}) \\
&\quad + P(A|C = D_2 D_2, T_{ij})P(C = D_2 D_2|T_{ij}) \\
&= f_{11}P(C = D_1 D_1|T_{ij}) + f_{12}P(C = D_1 D_2|T_{ij}) \\
&\quad + f_{22}P(C = D_2 D_2|T_{ij})
\end{aligned}$$

Denklem 1.25

α_{ij} , F_1 ebeveyninin T_{ij} olayına bağlı olarak hastalık alelini geçirme olasılığı ise;

$$\begin{aligned}
P(C = D_1 D_1|T_{ij}) &= P(F_2 \rightarrow D_1)P(F_1 \rightarrow D_1|T_{ij}) \\
&= p\alpha_{ij} \\
P(C = D_1 D_2|T_{ij}) &= P(F_2 \rightarrow D_2)P(F_1 \rightarrow D_1|T_{ij}) \\
&\quad + P(F_2 \rightarrow D_1)P(F_1 \rightarrow D_2|T_{ij}) \\
&= (1-p)\alpha_{ij} + p(1-\alpha_{ij}) \\
P(C = D_2 D_2|T_{ij}) &= P(F_2 \rightarrow D_2)P(F_1 \rightarrow D_2|T_{ij}) \\
&= (1-p)(1-\alpha_{ij})
\end{aligned}$$

Denklem 1.26

Denklem 1.25 ve 1.26'dan;

$$\begin{aligned}
P(A|T_{ij}) &= f_{11}p\alpha_{ij} + f_{12}[(1-p)\alpha_{ij} + p(1-\alpha_{ij})] + f_{22}(1-p)(1-\alpha_{ij}) \\
&= \alpha_{ij}[p(f_{11} - f_{12}) + (1-p)(f_{12} - f_{22})] + pf_{12} + (1-p)f_{22} \\
&= \alpha_{ij}Z + pf_{12} + (1-p)f_{22}
\end{aligned}$$

Denklem 1.27

$Z = p(f_{11} - f_{12}) + (1-p)(f_{12} - f_{22})$ dir. F_1 'in normal olan D_2 alelini çocuğa geçirdiği bilindiğinde, çocuğun hasta olma olasılığı; $P(A|F_1 \rightarrow D_2) = pf_{12} + (1-p)f_{22}$ koşullu olasılığı şeklinde hesaplanır. 1-B, hasta çocuğun kromozomundaki hastalık lokusunda normal alel bulunması olasılığının, popülasyondan rasgele olarak seçilen bir kromozomdaki hastalık lokusunda normal alelin bulunması olasılığına oranını gösterebilirsin:

$$\begin{aligned}
1 - B &= \frac{P(F_1 \rightarrow D_2 | A)}{P(F_1 \rightarrow D_2)} = \\
&= \frac{P(F_1 \rightarrow D_2, A) / P(A)}{P(F_1 \rightarrow D_2)} \\
&= \frac{P(A | F_1 \rightarrow D_2) P(F_1 \rightarrow D_2) / P(A)}{P(F_1 \rightarrow D_2)} \\
&= \frac{P(A | F_1 \rightarrow D_2)}{P(A)}
\end{aligned}$$

Denklem 1.28

Denklem 1.28, Bayes kuralı kullanılarak son halini almıştır. Gerekli işlemlerden sonra B çekilerek Denklem 1.29 elde edilir:

$$B = \frac{pZ}{P(A)} = \frac{p[p(f_{11} - f_{12}) + (1 - p)(f_{12} - f_{22})]}{p^2 f_{11} + 2p(1 - p)f_{12} + (1 - p)^2 f_{22}}$$

Denklem 1.29

Denklem 1.19 ve 1.28'den:

$$P(T_{ij} | A) = P(T_{ij})[1 + (\alpha_{ij} - p)B / p]$$

Denklem 1.30

F_1 'in $M_i M_j$ marker genotipine sahip olduğu bilindiğinde, hastalık alelini geçirme koşullu olasılığı α_{ij} ;

$$\begin{aligned}
\alpha_{ij}P(T_{ij}) &= P(F_1 \rightarrow D_1, T_{ij}) \\
&= P(F_1 \rightarrow D_1, T_{ij}, \bar{R}) + P(F_1 \rightarrow D_1, T_{ij}, R) \\
&= P(F_1 \rightarrow M_i D_1, F_1 = M_i M_j, \bar{R}) \\
&\quad + P(F_1 \rightarrow M_i D_1, F_1 = M_i M_j, R) \\
&= P(F_1 = M_i D_1 / M_j D_1, F_1 \rightarrow M_i D_1, \bar{R}) \\
&\quad + P(F_1 = M_i D_1 / M_j D_1, F_1 \rightarrow M_i D_1, R) \\
&\quad + P(F_1 = M_i D_1 / M_j D_2, F_1 \rightarrow M_i D_1, \bar{R}) \\
&\quad + P(F_1 = M_i D_2 / M_j D_1, F_1 \rightarrow M_i D_1, R) \\
&= P(F_1 = M_i D_1 / M_j D_1, F_1 \rightarrow M_i D_1) \\
&\quad + [P(M_i D_1 / M_j D_2)(1 - \theta) + P(M_i D_2 / M_j D_1)\theta] / 2 \\
&= [(1 + I\{i = j\})P(M_i D_1 / M_j D_1) + P(M_i D_1 / M_j D_2)] / 2 \\
&\quad + [\theta\{P(M_i D_2 / M_j D_1) - P(M_i D_1 / M_j D_2)\}] / 2
\end{aligned}$$

Denklem 1.31

Denklemin son satırına dikkat edilirse rekombinasyon katsayısı olan θ $(i,j)=(1,2)$ ve $(i,j)=(2,1)$ olan α_{ij} ler için eklenmiştir. Rasgele eşleşme varsayımı altında Denklem 1.18 ve Denklem 1.31'den aşağıdaki gibi bulunur:

$$\alpha_{ij} = \begin{cases} (pq + \delta) / q & i = j = 1 \\ (pq + \delta) / q - \theta\delta / q(1 - q) & (i, j) = (1, 2) \\ (p(1 - q) - \delta) / (1 - q) + \theta\delta / q(1 - q) & (i, j) = (2, 1) \\ (p(1 - q) - \delta) / (1 - q) & i = j = 2 \end{cases}$$

Denklem 1.32

Denklem 1.32, 1.30 ve 1.23 birleştirilirse; Tablo 1.5'teki değerlere ulaşılır.

b ve c'nin farkının beklenen değeri;

$$E(b - c) = 2nB\delta(1 - 2\theta) / p$$

Denklem 1.33

Eğer penetranslar aynı ise $B=0$ 'dır. Dolayısıyla hastalığın ortaya çıkmasında hastalığın bir etkisi yoktur. b-c'nin beklenen değeri, alelik ilişkisi olmadığında, yani $\delta=0$, veya bağlantı olmadığında, yani $\theta=1/2$ olduğunda sıfırdır. Bu nedenle TDT istatistiği sadece şu yokluk hipotezlerini test eder:

Bağlantı olduğunda, $\theta < 1/2$, ilişki yoktur, $\delta = 0$. İlişki olduğunda, $\delta \neq 0$, bağlantı yoktur, $\theta = 1/2$. İlişkinin neden kaynaklandığının bir önemi yoktur.

TDT istatistiği çıkarılırken, hasta çocuğun ebeveynlerinin katkısının birbirinden bağımsız olduğu varsayılmıştır. Bu varsayım ilişkinin yokluğunda, $\delta = 0$, doğrudur. Çünkü ebeveynlerin seçimi hastalık aleline bağlıdır. İlişki olmadığında hastalık alelleriyle marker alellerinin ebeveynlerde görülmesi arasında korelasyon yoktur. Ancak, bağlantı olmadığında, alelik ilişki varsa, alellerin ebeveynlerden geçişlerinin bağımsızlığı hakkında kesin bir şey söylenemez. Bu durumda, bağımsızlık varsayımı, eğer marker alellerinin toplumda görülme sıklıkları eşitse ya da penetranslar çarpımsalsa; $f_{12}^2 = f_{11} \times f_{22}$, sağlanır.

Ancak, b-c'nin ortalamasının ve varyansının bağlantıdan etkilenmediği ispatlanmıştır (Whittaker ve ark., 1998). Bundan ötürü χ^2 testi büyük örneklem genişliklerinde her iki ebeveyn de analize dahil edilmişse geçerlidir.

Ebeveynlerin alelleri birbirlerinden bağımsız olarak aktardığı varsayımını yapmayan, daha az güce sahip başka bir test Schaid ve Sommer (1993) ve Knapp ve ark. (1995) tarafından geliştirilmiştir.

TDT istatistiği heterozigot ebeveynlerin M_1 ve M_2 alellerini hasta çocuklarına geçirme sayılarını karşılaştırır. Geçirilen aleller vakalar, geçirilmeyen aleller kontroller olarak alındığında, iki grup etnik açıdan mükemmel uyum gösterir, çünkü gruplar arasındaki eşler aynı kişiden alınmıştır. TDT, bağlantı olmadığı ya da alelik ilişki olmadığı yokluk hipotezini test eder. Bu testten elde edilecek anlamlı sonuç, şans eseri değilse, hem bağlantının hem de ilişkinin varlığından ileri gelmiştir. Anlamlı sonucun başka bir nedeni de belirlenen alelin, hasta çocuğa geçirilmesinin beklenenden daha fazla olmasıdır. Bu nedenle testin geçerli olması için mayotik seçilimin olmadığı varsayılır. İlişki olduğu zaman, TDT diğer bağlantı testlerinden daha fazla güce sahiptir. Ancak TDT aile içinde karşılaştırma yaptığı için, bağlantı

olmadığında ilişkiye yol açan toplumdaki tabaklaşmadan etkilenmez. (Ewens ve Spielman, 1995)

1.6.2.1.2 Haplotip Bazlı Haplotip Görelî Risk Testi (HHRR)

TDT'de bir ebeveynin geçirdiği ve geçirmediği aleller eş gözlemler olarak değerlendirilir. Birbirinden bağımsız aile üçlülerinden elde edilen geçirilen ve geçirilmeyen alel toplamlarından oluşan veriler, iki bağımsız vaka-kontrol örnekleme gibi ele alınır. Bu test haplotip bazlı haplotip görelî risk testi (HHRR) olarak adlandırılmıştır (Terwilliger ve Ott, 1992). İlişki yoktur hipotezi test edilir. Tablo 1.4'teki veriler kullanılarak yazılan test istatistiği;

$$HHRR = \frac{(b - c)^2}{(2a + b + c)(b + c + 2d) / 4n} \sim \chi^2_1 \quad \text{Denklem 1.34}$$

TDT'nin aksine HHRR bağlantıyı test etmek için kullanılamaz. Ve popülasyonda var olabilecek tabakalaşmadan etkilenir. Ancak, rekombinasyon katsayısı 0'a yakın olduğu zaman alelik ilişki için kullanılabilecek güçlü bir testtir.

TDT'nin Geçerliliği

TDT ilişkinin varlığında bağlantı testi ya da bağlantının olduğu bilindiğinde ilişki testidir. Bağlantı testi olduğunda bir ailede birden fazla hasta çocuk analize alınabilir. Bu durumda ebeveynlerin genotiplerine göre geçirilen ve geçirilmeyen alel sıklıkları her çocuk için ayrı ayrı yazılabilir. Bağlantı analizi olarak kullanıldığında genişletilmiş aile ağaçları da teste alınabilir. Bunun nedeni bağlantı olmadığı yokluk hipotezi altında, Mendel kalıtım kuralına göre (rasgele karışım) ebeveynlerden çocuklara alellerin geçirilmesi ya da geçirilmemesi rasgele olarak gerçekleşir. İlişki analizi olarak kullanıldığında, ne kadar hasta çocuk olursa olsun, ebeveynlerden alınan genotipik bilgi her

ebeveyn için sadece bir kez kullanılabilir. Bunun nedeni, ilişki olmasa da eğer bağlantı varsa bir ebeveynin her bir hasta çocuğuna doğru gerçekleşen alel aktarımı arasında korelasyon vardır (Spielman ve Ewens, 1996).

T_{sp} , hasta iki kardeşin ve ebeveynlerinin genotip verilerinin kullanıldığı, hem bağlantı hem de ilişki için geçerli olan testtir (Martin ve ark. 1997). TDT ile benzerdir. Farkları, TDT'de ilişkiyi test etmek için iki hasta kardeşin kullanılamamasıdır. Bunun nedeni Wicks tarafından T_{sp} ve TDT'nin karşılaştırılması ile gösterilmiştir (2000).

M_1M_2 heterozigot marker genotipine sahip olan ebeveynler için n_{11} ; M_1 alelini her iki hasta çocuğa da geçiren ebeveynlerin sayısını, n_{22} ; M_2 alelini her iki hasta çocuğa da geçiren ebeveynlerin sayısını ve n_{12} ; M_1 alelini bir hasta çocuğa M_2 alelini de diğerine geçiren ebeveynlerin sayısını gösterebilir. T_{sp} test istatistiği;

$$T_{sp} = \frac{(n_{11} - n_{22})^2}{n_{11} + n_{22}} \quad \text{Denklem 1.35}$$

TDT istatistiği;

$$TDT = \frac{(n_{11} - n_{22})^2}{(n_{11} + n_{22} + n_{12})/2} \quad \text{Denklem 1.36}$$

TDT, IBD alel geçişlerinin fazla olduğu hasta kardeşlerden oluşan verileri kullandığı için bağlantı testi olarak T_{sp} 'den daha güçlüdür (Wicks, 2000). Çünkü, bağlantı olduğunda her iki hasta çocuğa M_1 alelini geçirenlerin ve her iki hasta çocuğa M_2 alelini geçirenlerin toplam sayısı, bir hasta çocuğa M_1 alelini, diğerine M_2 alelini geçiren ebeveynlerin sayısından fazladır.

$$TDT = T_{sp} \times \frac{(n_{11} + n_{22})}{(n_{11} + n_{22} + n_{12})/2} \quad \text{Denklem 1.37}$$

Denklem 1.37'deki çarpım, fazla paylaşımı göstermektedir. İlişki olmadan, sadece bağlantının varlığı, alellerin beklenenden fazla paylaşımına neden olmaktadır. T_{sp} 'nin tek başına bağlantının yanında ilişkinin varlığını ortaya çıkaradığı zaman, TDT'den elde edilecek ret sonucu alellerin fazla paylaşıldığı şeklinde yorumlanabilir.

1.6.2.1.3 Çok Alelli Marker ile Kalıtımda Dengesiz Aktarım Testi:

Marker lokusunda yer alabilecek alel çeşidi $m > 2$ tane ise; Tablo 1.4, Tablo 1.6'daki gibi oluşturulur:

Tablo 1.6 Çok alelli markerda geçirilen ve geçirilmeyen alel sayıları

	Geçirilmeyen				
Geçirilen	M_1	M_2	...	M_m	Toplam
M_1	t_{11}	t_{12}		t_{1m}	$t_{1.}$
M_2	t_{21}	t_{22}		t_{2m}	$t_{2.}$
...					
M_m	t_{m1}	t_{m2}		t_{mm}	$t_{m.}$
Toplam	$t_{.1}$	$t_{.2}$		$t_{.m}$	$t_{..}$

İkiden fazla alel söz konusu ise, hastalıkla ilişkili olabilecek birden fazla alel bulunabilir.

Marjinal homojenliği test eden HHRR test istatistiğinin çok alelli marker için genelleştirilmiş hali;

$$HHRR = \sum_i (t_{i.} - t_{.i})^2 / (t_{i.} + t_{.i}) \quad \text{Denklem 1.38}$$

olarak verilmiştir.

Bu istatistik, bağlantı dengesinin sağlandığı varsayımı altında asimptotik olarak χ^2_{m-1} dağılımı gösterir.

Geçirilmiş ve geçirilmemiş alellerin tablodaki simetrisini test eden çok alelli marker için genişletilmiş TDT istatistiği aşağıdaki gibidir (Bickemöller ve Clerget-Darpoux, 1995).

$$TDT_a = \sum_{i < j} (t_{ij} - t_{ji})^2 / (t_{ij} + t_{ji}) \quad \text{Denklem 1.39}$$

Denklemden anlaşılacağı gibi ebeveynlere ait $m(m-1)/2$ tane olası tüm heterozigot marker genotiplerinin toplamıdır. Rekombinasyon fraksiyonu $1/2$ olduğunda asimptotik olarak $\chi^2_{m(m-1)/2}$ dağılımı gösterir.

HHRR istatistiğinin bir dezavantajı bağlantı dengesizliği bilindiğinde, $\theta = 1/2$ olduğunda geçirilen ve geçirilmeyen aleller birbirinden bağımsız değildir. (Ott, 1989) Bu durumda test istatistiğinin varsayılan örneklem dağılımı geçersiz olur ve tip I hata olasılığı artar. HHRR $\theta = 0$ alternatif hipotezi altında tercih edilebilir bir test iken, $\theta = 1/2$ yokluk hipotezi altında uygun değildir. Ancak HHRR testi bağlantı dengesizliği testi olarak kullanılırsa bu sorunlarla karşılaşmaz, çünkü geçirilen ve geçirilmeyen aleller bağlantı dengesinde bağımsızdır.

Sadece bir marker alelin ilişkili olup olmadığına karar vermek için, o alele karşılık diğer tüm aleller alınarak TDT_i hesaplanır:

$$TDT_i = \left[\sum_{i \neq j} (t_{ij} - t_{ji}) \right]^2 / \sum_{i \neq j} (t_{ij} + t_{ji}) \quad \text{Denklem 1.40}$$

Hesaplanan TDT_i değerleri bağımsız dağılım göstermediği için Bonferroni düzeltme faktörü kullanılmalıdır.

Sham ve Curtis (1995) bağlantı ve bağlantı dengesizliği için geçerli bir test geliştirmişlerdir. Bu test her bir marker alelinin farklı derecede hastalıkla ilişkili olduğu durumunu dikkate alır. Bir ebeveynde iki farklı marker aleli varsa, hastalıkla daha çok ilişkili olan alelin geçirilme olasılığı yüksek olacaktır. Başka bir deyişle, ebeveynden gelen her marker genotipi için ayrı parametrenin olduğu modele karşı kurulan yokluk hipotezi yerine, her marker alel için bir parametrenin bulunduğu modele karşı kurulan yokluk hipotezi dikkate alınmıştır. (Sham ve Curtis, 1995)

Tek lokuslu hastalık modeli için alellerin geçirilme ve geçirilmeme durumuna göre olasılıkları Tablo 1.7’de gösterilmiştir.

Tablo 1.7 Çok alelli marker için geçirilen ve geçirilmeyen alellerin olasılıkları

	Geçirilmeyen				
Geçirilen	M ₁	M ₂	...	M _m	Toplam
M ₁	$p_1^2 d_{11}$	$p_1 p_2 d_{12}$	...	$p_1 p_m d_{1m}$	$p_1 [1+B(1-\theta)(e_{11}-1)]$
M ₂	$p_2 p_1 d_{21}$	$p_2^2 d_{22}$	...	$p_2 p_m d_{2m}$	$p_2 [1+B(1-\theta)(e_{12}-1)]$
...	...	...	...	...	...
M _m	$p_m p_1 d_{m1}$	$p_m p_2 d_{m2}$	...	$p_m^2 d_{mm}$	$p_m [1+B(1-\theta)(e_{1m}-1)]$
Toplam	$p_1 [1+B\theta(e_{11}-1)]$	$p_2 [1+B\theta(e_{12}-1)]$	...	$p_m [1+B\theta(e_{1m}-1)]$	1

Genetik model hastalık lokusunu ve marker lokusunu içerir. Hastalık lokusundaki aleller 1 ve 2, görülme olasılıkları q_1 ve q_2 dir. Benzer şekilde marker lokusundaki aleller 1,2,...,m ve görülme olasılıkları p_1 , p_2 ,..., p_m olarak gösterilmiştir. s hastalık lokusunu i de marker lokusunu temsil etmek üzere, s_i haplotipinin görülme sıklığı h_{si} ve s_i haplotipi için bağlantı dengesizliği parametresi haplotip frekansının bağlantı dengesinde beklenen frekansa oranıdır, $e_{si} = h_{si} / (q_s p_i)$ şeklinde ifade edilir. Hastalıkla marker lokusu arasındaki rekombinasyon fraksiyonu θ dir. st genotipinin penetransı

f_{st} şeklinde ifade edilir ve $f_{11} \geq f_{12} = f_{21} \geq f_{22}$ olduğu varsayılır. Ayrıca popülasyondaki eşleşmelerin rasgele olduğu ve verilerin hasta çocuklar arasından rasgele olarak seçildiği de yapılan varsayımlardır.

ij genotipindeki ebeveynlerden i ve j alellerinin eşit olmayan bir şekilde aktarılması $d_{ij} \neq d_{ji}$ ise gerçekleşir. Bu nedenle $d_{ij} = 1 + B[(e_{1i} - 1) + \theta(e_{1j} - e_{1i})]$ aktarım eşitsizliği parametresi olarak ele alınır. $B = \{q_1[q_1(f_{11} - f_{12}) + q_2(f_{12} - f_{22})]\} / \{q_1^2 f_{11} + 2q_1 q_2 f_{12} + q_2^2 f_{22}\}$ değeri, çocuk hastayken normal alelin aktarılması olasılığının, koşul olmadan normal alelin aktarılması olasılığına oranının 1'den çıkarılmasıyla bulunur.

ETDT, TDT'nin çok alelli marker için bağlantı dengesizliğinde güçlü bağlantıyı test etmek için lojistik regresyon yaklaşımını kullanan genişletilmiş halidir. ij genotipine sahip ebeveynlerin i alelini geçirme oddslarının logaritması $b_i - b_j$ ile belirlenir. Burada b_i ve b_j sırasıyla i ve j alellerine bağlı parametreleri vermektedir. p_{ij} ve p_{ji} ij genotipine sahip ebeveynlerin sırasıyla i ve j alellerini geçirme olasılıkları olarak belirlenmiştir. Böylelikle model denklemi i ve j'nin $m(m-1)/2$ olası kombinasyonu için aşağıdaki gibi yazılır:

$$g_{ij} = \ln(p_{ij} / p_{ji}) = b_i - b_j \quad \text{Denklem 1.41}$$

Rasgele olarak $b_m=0$ alınmıştır. $\theta=0$ olduğunda $b_i = \ln(d_{ii}) - \ln(d_{mm})$ olması için $\ln(p_{ij} / p_{ji}) = \ln(d_{ii}) - \ln(d_{jj})$ olarak ifade edilmiştir. Matris gösteriminde genel model $g=Xb$ olarak ifade edilmiştir. g vektörü, $m(m-1)/2$ tane g_{ij} 'in log-odds değerlerinden oluşur. X; $m(m-1)/2 \times m-1$ 'lik tasarım matrisidir. Verilerin, her bir sıranın ilgili heterozigot ebeveyn genotipi ile ilgili tüm verileri barındırabilmesi için, $m(m-1)/2$ sıra için tekrar düzenlenmesi gerekmektedir. Bu nedenle $m-1$ tane kukla değişkeni tanımlanmalıdır. Örneğin; i aleli için, eğer ebeveynin genotipinin ilk aleli i ise kukla değişkeninin değeri 1, ikinci aleli i ise kukla değişkeninin değeri -1, i

aleli o genotipte yer almıyorsa değer 0'dır. m=5 için hazırlanmış olan örnek Tablo 1.8'de verilmiştir.

Tablo 1.8 5 alelli marker için lojistik regresyon analizinde kullanılacak veri düzenlemesi

Ebeveyn Genotipi	Frekans Toplamı	İlk Alelin Geçirilmesi	Kukla Değişken i=1	Kukla Değişken i=2	Kukla Değişken i=3	Kukla Değişken i=4
1 2	$t_{12}+t_{21}$	t_{12}	1	-1	0	0
1 3	$t_{13}+t_{31}$	t_{13}	1	0	-1	0
1 4	$t_{14}+t_{41}$	t_{14}	1	0	0	-1
1 5	$t_{15}+t_{51}$	t_{15}	1	0	0	0
2 3	$t_{23}+t_{32}$	t_{23}	0	1	-1	0
2 4	$t_{24}+t_{42}$	t_{24}	0	1	0	-1
2 5	$t_{25}+t_{52}$	t_{25}	0	1	0	0
3 4	$t_{34}+t_{43}$	t_{34}	0	0	1	-1
3 5	$t_{35}+t_{53}$	t_{35}	0	0	1	0
4 5	$t_{45}+t_{54}$	t_{45}	0	0	0	1

Modelin Log-olabilirlik denklemi aşağıdaki gibidir:

$$L = \sum_{i < j} \{t_{ij} \ln(p_{ij}) + t_{ji} \ln(p_{ji})\} \quad \text{Denklem 1.42}$$

Burada toplam, farklı alellerin olası tüm kombinasyonlarından oluşur ve p aktarım olasılıkları m-1 tane b regresyon katsayılarının fonksiyonudur. Bağlantıyı ve bağlantı dengesizliğini alel bazında test etmek için kullanılan olabilirlik oran istatistiği $G_1 = 2(L_1 - L_0) \sim \chi_{m-1}$ dir. L_0 , $\theta=1/2$ yokluk hipotezi altında b değerlerinin 0 olduğu durumda hesaplanan log-olabilirlik değeri, L_1 ise m-1 parametreye göre en büyük değeri hesaplanan log-olabilirlik değeridir.

1.6.2.1.4 1-TDT: Sadece Bir Ebeveynin Genotipine Ulaşılabildiğinde TDT

Sun ve ark. (1998) tarafından, marker lokusla hastalık lokusu arasındaki bağlantıyı, hasta çocuğun ve bir ebeveyninin genotipleri kullanılarak test eden 1-TDT testi geliştirilmiştir. Bu testin bağlantı ve ilişki olmadığı yokluk hipotezi altında yansız olduğu belirtilmiştir. Alkolizm gibi bazı hastalıklarda karşılaşılabilen, hasta bireylerin ebeveynlerinden sadece birine ulaşılabiliriyorsa kullanılır.

M_1 ve M_1 alellerinden oluşan iki alelli marker lokusunu ele alalım. Heterozigot genotipe sahip bir ebeveyn (M_1M_2) ve homozigot genotipe (M_1M_1 veya M_2M_2) sahip hasta çocuğu çalışmaya alındığında çocuğun genotipinde bulunan hangi alelin genotipine ulaşılan ebeveynden geçtiği anlaşılır. Ancak eğer çocuk da ebeveyn de heterozigot genotipe sahipse (M_1M_2), ebeveynin hangi alelini hasta çocuğa geçirdiği bilinemez. Bu tip ebeveyn-hasta çocuk çiftleri çalışmadan çıkarılmalıdır (Curtis ve Sham, 1995). Alellerin geçirilme olasılıkları aynı olsa bile bu yöntemle alel sıklıklarına bağlı yanlılık ortaya çıkacaktır. 1-TDT yöntemi bu yanlılığı ortadan kaldırıp, sadece bir ebeveyni olan hasta çocuklarla çalışma yapılabilmesini sağlar.

Vaka (hasta çocuk) - ebeveyn kontrolden oluşan çalışma tasarımı Tablo 1.8'deki gibi özetlenir:

Tablo 1.9 Bir ebeveyn mevcutken vaka-ebeveyn kontrol çalışma tasarımı

	Ebeveynin Genotipi		
Vakanın Genotipi	M_1M_1 (0)	M_1M_2 (1)	M_2M_2 (2)
M_1M_1 (0)	A_{00}	A_{01}	0
M_1M_2 (1)	A_{10}	A_{11}	A_{12}
M_2M_2 (2)	0	A_{21}	A_{22}

Tabloda parantez içindeki değerler bireyin genotipinde bulunan M_2 alelinin sayısını vermektedir. A_{ij} ise hasta çocuğun genotipinin i , ebeveyninin genotipinin j olduğu ikililerin sayısını gösterir. i ve j , iki alelli marker söz konusu olduğu için 0,1 ve 2 değerlerini almaktadır.

Bu testte anne ve babanın ulaşılabilir olmasının genotiplerinden bağımsız olduğu düşünülmektedir. Bunun yanı sıra kestirimlerin yansız olabilmesi için iki varsayımdan birinin sağlanması gerekmektedir:

1. Marker lokusta aynı genotipe sahip kadın ve erkek bireylerin aynı eşleşme tercihinə sahiptir.
2. Her bir çekirdek ailedeki anne ya da baba aynı kayıp olma olasılığına (1/2) sahiptir.

Bu varsayımlardan biri sağlandığı takdirde çalışılan popülasyonda tabakalaşma ve rasgele olmayan eşleşmeler görülse bile kestirimler yansızdır.

M_1M_2 genotipine sahip bireylerle M_1M_1 genotipine sahip bireylerin risk oranı λ_1 , M_1M_2 genotipine sahip bireylerle M_2M_2 genotipine sahip bireylerin risk oranı λ_{-1} olarak tanımlanmıştır. (Sun ve ark., 1998)

$$\hat{\lambda}_{-1} = \frac{(A_{11} + A_{12} - A_{10})}{(2A_{21})}$$

$$\hat{\lambda}_1 = \frac{(A_{11} + A_{10} - A_{12})}{(2A_{01})}$$

Denklem 1.43

Eğer hastalık marker lokusuyla ilişkili değilse $\lambda_1 = \lambda_{-1} = 1$ 'dir. Eğer hastalık M_2 aleliyle ilişkili ise $\lambda_1 > 1$ ve /veya $\lambda_{-1} < 1$ olurken, eğer hastalık M_1 aleliyle ilişkili ise $\lambda_1 < 1$ ve /veya $\lambda_{-1} > 1$ dir. $\hat{\lambda}_1$ 'nin payıyla $\hat{\lambda}_{-1}$ 'nin paydasının toplamı $\hat{\lambda}_1$ 'nin paydasıyla $\hat{\lambda}_{-1}$ 'nin payının toplamı karşılaştırılır. Buradan yola çıkılarak bir takım matematiksel işlemler sonucunda $b_1 = A_{01} + A_{12}$ ile

$c_1 = A_{10} + A_{21}$ karşılaştırılması gerektiği bulunmuştur (Sun ve ark., 1998). Eğer hastalık marker lokusu ile ilişkiliyse $b_1 - c_1 \neq 0$ 'dır.

Eğer her bir çekirdek aileden çalışmaya yalnızca bir hasta çocuğun ve bir ebeveynin genotipi alınmışsa $b_1 - c_1$ 'in varyans kestirimi $V = b_1 + c_1$ 'dir. Hastalık ve marker lokusu arasındaki ilişkiyi test etmek için kullanılan test istatistiği;

$$T_1 = \frac{A_{01} + A_{12} - A_{10} - A_{21}}{\sqrt{V}} = \frac{b_1 - c_1}{\sqrt{V}} \quad \text{Denklem 1. 44}$$

yokluk hipotezi altında $b_1 + c_1$ büyük olduğunda, T_1 yaklaşık olarak standart normal dağılır. p değeri de normallik yaklaşımı kullanılarak hesaplanır.

$b_1 + c_1$ büyük olmadığı zaman normallik yaklaşımı kullanılamaz. İlişki olmadığı yokluk hipotezi altında b_1 ve c_1 'in aynı olması beklenir. $b_1 + c_1$ 'e bağlı olarak b_1 binom dağılımdan gelen rasgele bir değişkendir. ($\text{Bin}(b_1 + c_1, 1/2)$) Böylece normal yaklaşım kullanılamadığı zaman binom dağılımı kullanılarak p değeri hesaplanır.

Birden fazla hasta çocuğun olduğu ailelerde veya birden fazla kuşak bulunuyorsa, aile içindeki bireylerin genotipleri birbirine bağımlı olacağı için $b_1 - c_1$ 'in varyansı $b_1 + c_1$ ile kestirilemez. $b_1 - c_1$ 'in varyansını kestirebilmek için her bir aile ayrı olarak ele alınmalıdır. Örneğin, i. aileyi ele alırsak, b_{1i} ve c_{1i} hesaplanır. $b_{1i} - c_{1i}$ 'nin beklenen değeri, yokluk hipotezi altında 0 olduğu için $b_{1i} - c_{1i}$ 'nin varyansı $(b_{1i} - c_{1i})^2$ ile kestirilebilir. T_1 istatistiğinin payının varyansı $V_1 = \sum (b_{1i} - c_{1i})^2$ ile kestirilir. Burada toplam sembolü çalışmaya alınan tüm ailelerden elde edilen sonuçların toplandığını göstermektedir.

Eğer iki varsayım da sağlanmazsa, marker lokusla hastalık arasında ilişki yoksa, T_1 istatistiği normal dağılım göstermez. Varsayımların sağlanmadığı durumlarda marker lokusla hastalık arasındaki ilişkiyi test edebilmek için T_2 adı verilen test istatistiği geliştirilmiştir (Sun ve ark., 1998). Geliştirilen bu istatistik, varsayımlardan herhangi biri sağlandığında T_1 'den daha az güce sahiptir.

Geliştirilen testte, yaklaşık olarak yansız olan kestirimler için denklem 1.40'ta kullanılan A_{ij} 'ler A'_{ij} olarak değiştirilir:

$$A'_{ij} = M \times P_{ij} + P \times M_{ij} \quad \text{Denklem 1.45}$$

$i, j=0,1,2$ değerlerini alır. P ve M sırasıyla baba ve anne olan ailelerin sayısını vermektedir. P_{ij} kendisi i genotipine sahip babası j genotipine sahip hasta çocuk sayısını, M_{ij} ise kendisi i genotipine sahip annesi ise j genotipinde olan hasta çocuk sayısını vermektedir. Hastalıkla marker aleli ilişki olmadığı yokluk hipotezi altında;

$$E(M(pb_1 - pc_1) + P(mb_1 - mc_1)) = 0 \quad \text{Denklem 1.46}$$

$$pb_1 = P_{01} + P_{12}, \quad pc_1 = P_{10} + P_{21}, \quad mb_1 = M_{01} + M_{12} \text{ ve } mc_1 = M_{10} + M_{21}.$$

Test istatistiği;

$$T_2 = \frac{M(pb_1 - pc_1) + P(mb_1 - mc_1)}{\sqrt{M^2 \times pV + P^2 \times mV}} \quad \text{Denklem 1. 47}$$

pV ve mV , ilk testteki V gibi hesaplanır yalnız pV 'de baba olan aileleri, mV 'de ise anne olan aileler kullanılır.

Bağlantı dengesizliğinde 1-TDT, bağlantı testi olarak da kullanılabilir. Hastalık resesif olduğunda, rasgele eşleşme varsayımı altında ve Hardy-Weinberg dengesinde Ott bu yaklaşımı ispatlamıştır.

Hardy-Weinberg prensibi popülasyondaki alel frekansları ve genotipler arasındaki ilişkidir. Eğer popülasyon büyükse, (rasgele örenkleme hatasının önlenmesi için) eşleşmeler rasgele olarak yapılıyorsa, mutasyon yoksa, (aleller değişmiyorsa), popülasyonlar arasında göç olmuyorsa ve doğal seçim yoksa o zaman toplum Hardy-Weinberg dengesindedir. Bu denge, bu şartlar bozulmadıkça kuşaktan kuşağa değişmez. Popülasyon gen havuzundaki A ve a alelleri için frekanslar sırasıyla p ve q olsun. $p+q=1$ olmalıdır. A ve a alleleriyle oluşturulabilecek olası tüm genotipler AA, Aa ve aa dır. O halde popülasyonda olası tüm genotip kombinasyonları $p^2+2pq+q^2$ dir (Hardy, 1908).

D hastalık alelinin frekansı p, marker lokustaki M_1 alelinin frekansı da q olsun. Hastalık lokusu ile marker lokus arasındaki ilişki; $\delta = p(DM) - pq$ ve rekombinasyon fraksiyonu da θ dır. Tablo 1.9'da bir ebeveynin bilgisine ulaşılabılırken geçirilen ve geçirilmeyen alellerin birleşik dağılımları verilmiştir:

Tablo 1.10 Birleşik alel dağılımları

	Geçirilmeyen		
Geçirilen	M_1	M_2	Toplam
M_1	r	s	r+s
M_2	t	u	t+u
Toplam	r+t	s+u	1

Tabloda

$$r = (q + \delta / p)q, \quad s = (q + \delta / p)(1 - q) - \theta \delta / p, \quad t = (1 - q - \delta / p)q + \theta \delta / p \text{ ve}$$

$u = (1 - q - \delta / p)(1 - q)$ dir. Tablo 1.10'dan genotiplerin olasılıkları Tablo 1.11'deki gibi elde edilir.

Tablo 1.11 Genotiplerin olasılıkları

	Ebeveynin Genotipi		
Vakanın Genotipi	M ₁ M ₁ (0)	M ₁ M ₂ (1)	M ₂ M ₂ (2)
M ₁ M ₁ (0)	r(r+s)	s(r+s)	0
M ₁ M ₂ (1)	r(t+u)	s(t+u)+t(r+s)	u(r+s)
M ₂ M ₂ (2)	0	t(t+u)	u(s+t)

n tane hasta çocuk ve her birinin bir ebeveyni olsun. Tablo 1.9 ve 1.11'den $b_{1i} - c_{i1}$ 'in beklenen değeri aşağıdaki gibidir:

$$E(b_1 - c_1) = E(A_{01} + A_{12} - A_{10} - A_{21}) = n((s+u)(r+s) - (r+t)(t+u)) = n(s-t) = n(1-2\theta)\delta / p$$

Denklem 1.48

$\theta=1/2$ veya $\delta=0$ olduğunda beklenen değer 0'a eşittir. Bu nedenle bağlantı dengesizliği olduğunda bu test bağlantı testi olarak kullanılabilir.

1.6.2.1.5 Ailesel Bilgi Olmadığı Durumlarda Test

TDT, ebeveynlerden hasta çocuğa geçen alellerin dengesiz geçişini incelediği için, ebeveynlerin bilgisine ulaşamadığı zaman kullanılamaz. Yetişkin ya da ileri yaştaki bireylerde ortaya çıkan hastalıklar çalışılırken, hasta bireylerin ebeveynlerinin bilgisine ulaşmak zordur. Bu durumda hasta bireylerin kardeşlerinin bilgileriyle test yapılır. Kardeşlerin marker genotipleri karşılaştırılır.

1.6.2.1.6 İlişki Olduğunda Kardeşlerle Yapılan Bağlantı Testi (S-TDT)

Ebeveynlerin bilgisine ulaşamadığı, sağlıklı kardeşlerin genotiplerinin bilindiği durumda sib TDT (S-TDT) kullanılır. (Spielman ve Ewens, 1998) Bu yöntemde hasta ile sağlıklı kardeşlerin marker genotipleri karşılaştırılır. S-

TDT’de, ebeveynlerin genotip verilerine gerek duyulmadığı gibi alel sıklıkları da kullanılmaz.

Çalışmaya kardeş genotipleri alınacaksa; en az bir hasta ve bir sağlıklı kardeşin çalışmaya dahil edilmesi ve kardeşlerin farklı genotiplere sahip olmaları gerekir. S-TDT hasta bireylerin marker alel sıklıklarının sağlıklı kardeşlerinin alel sıklıklarından anlamlı derecede farklı olup olmadığını test eder. Bağlantı olmadığı zaman hastalıkla olan ilişki bu şekilde ortaya çıkmaz. Rasgele örnekleme hataları olmazsa, bağlantı yoksa, hasta ve sağlıklı kardeşlerin marker alel frekansları aynı olacaktır. Bu nedenle S-TDT’nin yokluk hipotezi hastalık ve marker lokusu arasında bağlantı yoktur olmalıdır (Spielman ve Ewens, 1998).

a tane hasta birey ve bu a bireyin u tane sağlıklı kardeşi olsun. Toplam birey sayısı $t=a+u$. M_1M_1 genotipine sahip birey sayısı r, M_1M_2 genotipine sahip birey sayısı s olsun. Marker lokusu ikiden fazla farklı alel içeriyorsa M_2 , M_1 olmayan diğer alelleri göstermektedir. Hastalık ile marker lokusu arasında bağlantı yoktur yokluk hipotezi altında her ailedeki hasta bireylerde bulunan M_1 alel sayısının ortalaması $(2r + s)a / t$ ve varyansı $au[4r(t - r - s) + s(t - s)] / [t^2(t - 1)]$ dir. Bu formüller hipergeometrik dağılımdan çıkarılmıştır.

A, tüm ailelerdeki hasta bireylerden elde edilen M_1 alelin sayısının ortalaması ve V de varyansı olma üzere;

$$\begin{aligned} A &= \sum (2r + s)a / t \\ V &= \sum au[4r(t - r - s) + s(t - s)] / [t^2(t - 1)] \end{aligned} \quad \text{Denklem 1.49}$$

Ortalama ve varyans, her bir aileden elde edilen ortalama ve varyansların toplanmasıyla hesaplanır.

Y, hasta bireylerde gözlenen toplam M_1 alel sayısı olmak üzere test istatistiği;

$$z = (Y - A) / \sqrt{V} \quad \text{Denklem 1.50}$$

olarak bulunur.

Testin p değeri normal dağılım yaklaşımıyla hesaplanır.

Kesikli dağılımlardan normal yaklaşım kullanılmak istendiğinde süreklilik düzeltmesi yapılır. Süreklilik düzeltmesi yapılmış test istatistiği;

$$z' = (|Y - A| - 1/2) / \sqrt{V} \quad \text{Denklem 1.51}$$

olarak hesaplanır.

1.6.2.1.7 Uyumsuz Kardeşlik Testi (SDT)

SDT; ilişki varlığında bağlantıyı, bağlantı dengesizliği olduğunda ise ilişkiyi test eder (Horvath ve Laird, 1998). Testin uygulanabilmesi için en az bir hasta ve en az bir sağlıklı bireyden oluşan kardeşlerin genotipik verileri gereklidir.

M_1 ve M_2 alellerinden oluşan iki alelli marker için a hasta birey sayısını, u da sağlıklı birey sayısını versin. Her bir kardeşlikte, hasta bireylerin taşıdığı M_2 alel sayısı ortalaması m_a^1 ve sağlıklı bireylerin taşıdığı M_2 alel sayısı ortalaması m_u^1 ile gösterilsin.

SDT, $d^1 = m_a^1 - m_u^1$ farklarının parametrik olmayan işaret testi olarak tanımlanmıştır (Horvath ve Laird, 1998).

Eğer $d^1 = 0$ ise bu kardeşler analizden çıkarılır.

b $d^1 > 0$ olan kardeşlerin sayısını, c ise $d^1 < 0$ olan kardeşlerin sayısını versin. SDT istatistiği;

$$T = (b-c)^2 / (b+c) \quad \text{Denklem 1.52}$$

olarak tanımlanır. Bağlantı (dengesizliği) olmadığı yokluk hipotezi, $H_0 : \Delta(\theta - 1/2) = 0$, altında $\chi^2_{(1)}$ dağılımı gösterir. İki yönlü p değeri aşağıdaki gibi hesaplanır:

$$p = 2 \min \left[\sum_{i=0}^b \binom{b+c}{i} \left(\frac{1}{2}\right)^{b+c}, \sum_{i=b}^{b+c} \binom{b+c}{i} \left(\frac{1}{2}\right)^{b+c} \right] \quad \text{Denklem 1.53}$$

Çok Alelli Marker ile SDT

Eğer m tane alel söz konusu ise, SDT $d^j = m_a^j - m_u^j$ farklarına dayanan çok değişkenli işaret testi halini alır. m_a^j ve m_u^j sırasıyla bir kardeşlikte hasta ve sağlıklı bireylerin sahip oldukları j alel sayısının ortalamasıdır.

Çok alelli SDT, en çok kullanılan çok değişkenli işaret testlerinden bileşen işaret testi kullanılarak tanımlanır (Bickel, 1965; Randles, 1989).

$S' = (S^1, \dots, S^{m-1})'$, $S^j = \sum_{i=1}^N \text{sgn}(d_i^j)$ ve d_i^j , j aleli için i. kardeşlikteki fark olarak tanımlanır. Ayrıca işaret fonksiyonu $\text{sgn}(d)$;

$$\text{sgn}(d) = \begin{cases} -1 & d < 0 \\ 0 & d = 0 \\ 1 & d > 0 \end{cases} \quad \text{Denklem 1.54}$$

dır. $H_0: E(S) = 0$ olan yokluk hipotezi altında $\chi^2_{(m-1)}$ dağılımı gösteren test istatistiği; $T = S'W^{-1}S$ dir. Burada W matrisinin elemanları;

$$W_{jk} = \sum_{i=1}^N \text{sgn}(d_i^j) \text{sgn}(d_i^k) \quad j, k = 1, \dots, m-1 \quad \text{Denklem 1.55}$$

1.6.2.1.8 Genişletilmiş Ailelerde İlişki Testi

Yukarıda bahsedilen testler geniş aile ağaçlarında akraba aileler bulunduğu zaman, bağlantı testi olarak kullanılsalar da, ilişkiyi ortaya çıkarmak için kullanılamazlar. Bu nedenle Martin ve ark. (2000, 2001) tarafından, akraba aile üçlülerinin ve/veya geniş ailelerdeki uyumsuz kardeşlerin kullanılabileceği bir ilişki testi olan PDT'yi geliştirmişlerdir. TDT gibi PDT de popülasyonda tabakalaşma bulunsa bile geçerli sonuçlar verir.

Eğer bağlantı varsa, alelik ilişki olmasa bile, akraba ailelerden alınan genotipler ilişkili (korele) olduğundan dolayı ilişki testi yapılamaz. Bu sorunu ortadan kaldırmak için aile üçlülerini birbirinden bağımsız gibi düşünüp değerlendirmek yerine tüm geniş ailede ilişkiye bakılır (Martin ve ark., 2000). Geniş aile içindeki her aile üçlüsü ve kardeşler arasındaki ilişki ölçümünün ortalaması alınarak genel aile için ilişki ölçüsü çıkarılır. Farklı aileler birbirlerinden bağımsız olarak ele alınırlar.

İlişki ölçümü için bilgi sağlayabilecek iki tür aile vardır: 1. En az bir hasta çocuk ve marker genotipleri bilinen ve en az biri heterozigot olan iki ebeveynden oluşan çekirdek aileler. 2. En az bir sağlıklı ve bir hasta bireyden oluşan, farklı marker genotiplerine sahip kardeşler.

M_1 ve M_2 alellerinden oluşan iki alelli marker lokusunu ele alalım. X_T her bir aile üçlüsünde geçirilen ve geçirilmeyen M_1 alel sayılarının farkını versin. Heterozigot bir ebeveynin X_T 'ye katkısı 1 veya -1'dir. Benzer şekilde, uyumsuz kardeşliklerdeki her kardeş çifti için hasta olan bireylerdeki M_1 alel sayısı ile sağlıklı bireylerdeki M_1 alel sayısının farkı olan X_S hesaplanır. X_S -2,-1,0,1 veya 2 olabilir.

Her aile için bilgi sağlayan en az bir çekirdek aile ve/veya bilgi sağlayan en az bir uyumsuz kardeşlikten elde edilen ölçüm:

$$D = \sum_j X_{T_j} + \sum_j X_{S_j} \quad \text{Denklem 1.56}$$

İlişkinin olmadığı yokluk hipotezi altında bütün aile üçlülere ve kardeş çiftleri için hesaplanan X_T ve X_S 'nin beklenen değerleri 0'dır. Bu nedenle her aile için $E(D)=0$ 'dır. N tane akrabalık barındırmayan, en az bir bilgi veren çekirdek aile veya en az bir bilgi veren uyumsuz kardeşlikten oluşan aile olsun. D_i , i. Aile için hesaplanan ölçüm ise, "İlişki yoktur." yokluk hipotezi altında;

$$E\left(\sum_{i=1}^N D_i\right) = 0 \quad \text{Denklem 1.57}$$

$$V\left(\sum_{i=1}^N D_i\right) = \sum_{i=1}^N V(D_i) = E\left(\sum_{i=1}^N D_i^2\right)$$

PDT için test istatistiği;

$$PDT = \frac{\left(\sum_{i=1}^N D_i\right)^2}{\sum_{i=1}^N D_i^2} \quad \text{Denklem 1.58}$$

Büyük örneklem genişliklerinde PDT χ_1^2 dağılımı gösterir.

Eğer veriler sadece bağımsız aile üçlülerinden oluşuyorsa, ilişki testi olarak TDT kullanılabilir. TDT heterozigot ebeveynlerden alınan bilgiyi bağımsız olarak kabul ederken, PDT'de her bir aile üçlüsü bağımsız birimler olarak alınır. İki alelli marker lokus için her bir heterozigot ebeveynden ve hasta çocukta elde edilen Y_i rasgele değişkeni hesaplanır. Y_i , geçirilen ve

geçirilmeyen M_1 alel sayılarının farkıdır. h tane heterozigot ebeveyn varsa, $i=1,2,...,h$ dir. Buradan TDT için test istatistiği;

$$TDT = \frac{\left(\sum_{i=1}^h Y_i \right)^2}{\sum_{i=1}^h Y_i^2} \quad \text{Denklem 1.59}$$

şeklinde hesaplanır. TDT ve PDT'nin payları aynı iken paydadaki varyans kestirimi farklıdır. Ailede tek heterozigot ebeveyn varsa, hesaplamalar aynı iken, iki heterozigot ebeveyn bulunduğu durumda istatistikler farklı hesaplanır.

$$\sum_{i=1}^h Y_i^2 = h$$

$$\sum_{i=1}^h D_i^2 = h + 2(n_c - n_d) \quad \text{Denklem 1.60}$$

Burada n_c ailedeki iki heterozigot ebeveynin hasta çocuğa aynı aleli geçirme sayısı, uyumlu kalıtım, n_d ise iki heterozigot ebeveynin hasta çocuğa farklı alelleri geçirme sayısı, uyumsuz kalıtım olarak tanımlanır. O halde,

$$\frac{TDT}{PDT} = 1 + \frac{2(n_c - n_d)}{h} \quad \text{Denklem 1.61}$$

“Bağlantı veya ilişki yoktur“ yokluk hipotezi altında $E(n_c - n_d) = 0$ 'dır. Bu nedenle iki test, yokluk hipotezi altında asimtotik olarak aynıdır. “Bağlantı ve ilişki vardır“ alternatif hipotezi altında iki testin sonuçları aynı olmayabilir. Farklı genetik modele sahip verilerde zaman zaman iki testin birbirlerinden daha güçlü sonuçlar verdiği gösterilmiştir (Martin ve ark., 2000). Ancak sağlanması muhtemel varsayımlara dayanarak her iki testten alınan sonuçlar hemen hemen aynıdır.

Eğer veriler uyumsuz kardeşlerden oluşuyorsa PDT, S-TDT ile aynıdır (Spielman ve Ewens, 1998; Curtis, 1997). Marker lokusu ikiden fazla alel çeşidinden oluşuyorsa, seçilen alele karşılık diğer tüm aleller alınır, her bir alel için PDT istatistiği hesaplanır. İstatistiksel anlamlılık test edilirken Bonferroni düzeltmesi yapılmalıdır.

1.6.2.1.9 Farklı Tipteki Ailelerden Alınan Bilginin Birleştirilmesi

Pratikte tek tip ailelerle değil, çeşitli bireylere sahip ailelerle karşılaşırız. Bireyler arasındaki akrabalık ilişkileri ne olursa olsun hasta bir çocuğun olması şarttır. Spielman ve Ewens (1998) aileleri dört gruba ayırmıştır:

- i. Her iki ebeveynin de genotipleri biliniyor
- ii. Ebeveynlerden birinin genotipi biliniyor, sağlıklı kardeşin genotipine ulaşılamıyor
- iii. Sağlıklı kardeşin genotipi biliniyor fakat ebeveynlerin genotiplerine ulaşılamıyor
- iv. Hem ebeveynlerden birinin genotipine ulaşılabilir hem de sağlıklı kardeşin genotipi biliniyor

IV. gruptaki veriler S-TDT ya da 1-TDT ile analiz edilebilir. Yapılan simülasyon çalışmalarında hem bir ebeveynin hem de sağlıklı kardeşin genotiplerine ulaşıldığı zaman S-TDT'nin 1-TDT'den daha güçlü sonuçlar verdiği ortaya konmuştur (Sun ve ark., 1998). Aynı zamanda TDT de 1-TDT'den daha güçlü bir testtir. Çünkü iki ebeveynin de genotipi bilinmektedir. Ancak her iki ebeveyn de homozigot genotipe sahipse testin gücü düşer.

Verilerin ilk üç gruptaki aile tiplerine uygun olduğunu varsayalım. Birinci türdeki ailelere TDT uygulanabilir. İki alelli bir sistemde b , M_1 alelini geçirip M_2 alelini geçirmeyen heterozigot ebeveynlerin sayısı, c ise M_2 alelini geçirip M_1 alelini geçirmeyen heterozigot ebeveynlerin sayısı olsun. Marker lokus ile hastalık lokusu arasında bağlantı yoksa b , $(b+c)/2$ ortalama ve $(b+c)/4$ varyansa sahip binom dağılımdan gelir.

İkinci türdeki ailelere 1-TDT uygulanabilir. 1-TDT konusunda bahsedilen varsayımlardan birisi sağlanıyorsa, T_1 test istatistiği kullanılır. b_1+c_1 verildiğinde, $S_1=b_1 (b_1+c_1)/2$ ortalama ve $V_1=(b_1+c_1)/4$ varyansa sahip binom dağılımdan gelmektedir. Eğer varsayımlar sağlanmıyorsa T_2 test istatistiği kullanılır. $S_1 = M \times pb_1 + P \times mb_1$ 'in ortalaması $M_1 = (M \times (pb_1 + pc_1) + P \times (mb_1 + mc_1)) / 2$ ve varyansı $V_1 / 4 = (M^2 \times pV + P^2 \times mV) / 4$ 'tür.

Üçüncü türdeki verilere de S-TDT uygulanabilir. Hasta bireylerdeki toplam M_1 alel sayısı olan Y 'nin ortalaması A ve varyansı V idi. Tüm ailelerden elde edilen ortalama ve varyansların toplanarak, genel ortalamanın A ve genel varyansın V hesaplandığı söylenmişti.

b , S_1 ve Y 'nin toplamı olan W istatistiği, yokluk hipotezi altında A_{com} ortalaması ve V_{com} varyansına sahiptir:

$$\begin{aligned} A_{com} &= (b + c) / 2 + M_1 / 2 + A \\ V_{com} &= (b + c) / 4 + V_1 / 4 + V \end{aligned} \quad \text{Denklem 1.62}$$

Üç testin birleştirilmesinden yola çıkılarak hesaplanan test istatistiği olan z , yokluk hipotezi altında yaklaşık normal dağılıma sahiptir.

$$z = (W - A_{com}) / \sqrt{V_{com}} \quad \text{Denklem 1.63}$$

S-TDT ve TDT'nin Birleştirilmesi

Verilerin bir kısmı TDT'ye bir kısmı da S-TDT'ye uygunsa, iki testin sonuçları birleştirilir.

TDT'de genellikle χ^2 test istatistiği kullanılmasına rağmen, kolaylık olması açısından, test istatistiği olarak M_1 alelinin n tane heterozigot ebeveynin hasta çocuklara geçirme sayısı olan X alınmıştır (Spielman ve Ewens, 1998). Marker ve hastalık arasında bağlantı yoksa, X , $n/2$ ortalamasına ve $n/4$ varyansına sahip Binom dağılımı gösterir.

S-TDT için de test istatistiği hasta çocuklar arasındaki M_1 alel sayısı olan Y 'dir. Marker ve hastalık arasında bağlantı yoksa Y , A ortalaması ve V varyansına sahip dağılımdan gelir.

İki testin birleştirilmesinden oluşan W istatistiği X ve Y 'nin toplamıdır. Hastalık ve marker lokusu arasında bağlantı olmadığı yokluk hipotezi altında W , A_{com} ortalaması ve V_{com} varyansına sahiptir:

$$\begin{aligned} A_{com} &= n/2 + A \\ V_{com} &= n/4 + V \end{aligned} \quad \text{Denklem 1.64}$$

Test istatistiği:

$$z = (W - A_{com}) / \sqrt{V_{com}} \quad \text{Denklem 1.65}$$

S-TDT'de olduğu gibi, burada da süreklilik düzeltmesinin kullanılması önerilir. (Spielman ve Ewens, 1998)

1.6.2.1.10 TDT'nin İstatistiksel Gücü

TDT analizinde hastalığın, anne ve babanın marker genotiplerinin ve en az bir hasta çocuğun olması gereklidir. Güç hesaplaması yapılırken popülasyonun Hardy-Weinberg dengesinde olduğu varsayılır. İlişki olduğu varsayımı altında yokluk hipotezi “Bağlantı vardır.” olacaktır (Chen ve Deng, 2001).

TDT Power Calculator (TDT-PC)

Chen ve Deng tarafından geliştirilen TDT-PC programı (2001), analitik olarak istatistiksel gücü veya gerekli olan örneklem genişliğini hesaplar. Aynı zamanda, analitik gücü desteklemesi için, simülasyona dayalı deneysel gücü ve Tip-I hata oranlarını da hesaplar. Deneysel hesaplamalarda, çekirdek aileler belirlenen parametreler doğrultusunda simüle edilirler ve hipotez testi istenen sayıda tekrarlanır. Reddedilen yokluk hipotezlerinin oranı deneysel gücü verir.

TDT'nin gücünü hesaplayabilmek için;

1. Popülasyonda görülen alel frekansları ve genotip penetransları,
2. Örneklemde yer alan farklı yapıdaki ailelerin sayısı, ailedeki toplam çocuk sayısı, hasta çocuk sayısı, ebeveynlerin hastalık durumları ve her bir farklı yapıya sahip kaç aile olduğu,
3. Önemlilik düzeyi ve tekrarlanan simülasyon sayısının bilinmesi gerekir.

TDT-PC programının görüntüsü Şekil 1.5'te görülmektedir.

```

TDT Power Calculator, Version 1.2.1
by Wei-Min Chen and Hong-Wen Deng, 9/18/2001
1<M> Parameters population-wide
2<F> Family structure (FA) Multiple family structures
3<P> Parameters related to statistical power

4<T> Sample size CALCULATION
5<S> Power CALCULATION
6<C> Combine power CALCULATION with sample size calculation

7<B> Batch job & file operating 8<A> Approach 9<H> Help 0<E> Exit
*****
Please make your choice:

```

Şekil 1.5 TDT-PC Programı

İlk bölüme gerekli bilgiler girildikten sonra ikinci bölümde örneklem genişliği ve güç hesaplamaları yapılır. 1<M> seçimi yapıldıktan sonra program, popülasyonda görülen A alel frekansının, AA, Aa ve aa genotiplerinin penetranslarının girilmesini istemektedir. Daha sonra 2<F> seçimi yapılır. Burada da ailedeki çocuk sayısı, hasta çocuk sayısı, hasta ebeveyn sayısı ve bu yapıdaki ailelerin sayısı istenmektedir. Son olarak 3<P> seçilir. Bu bölüme Tip I hata, güç ve tekrar sayısı girilir. Bu verilere göre 4, 5 ve 6. bölümlerde gerekli olan örneklem genişliği ve güç hem analitik hem de simülasyon sonucuna göre hesaplanır.

2. GEREÇ VE YÖNTEM

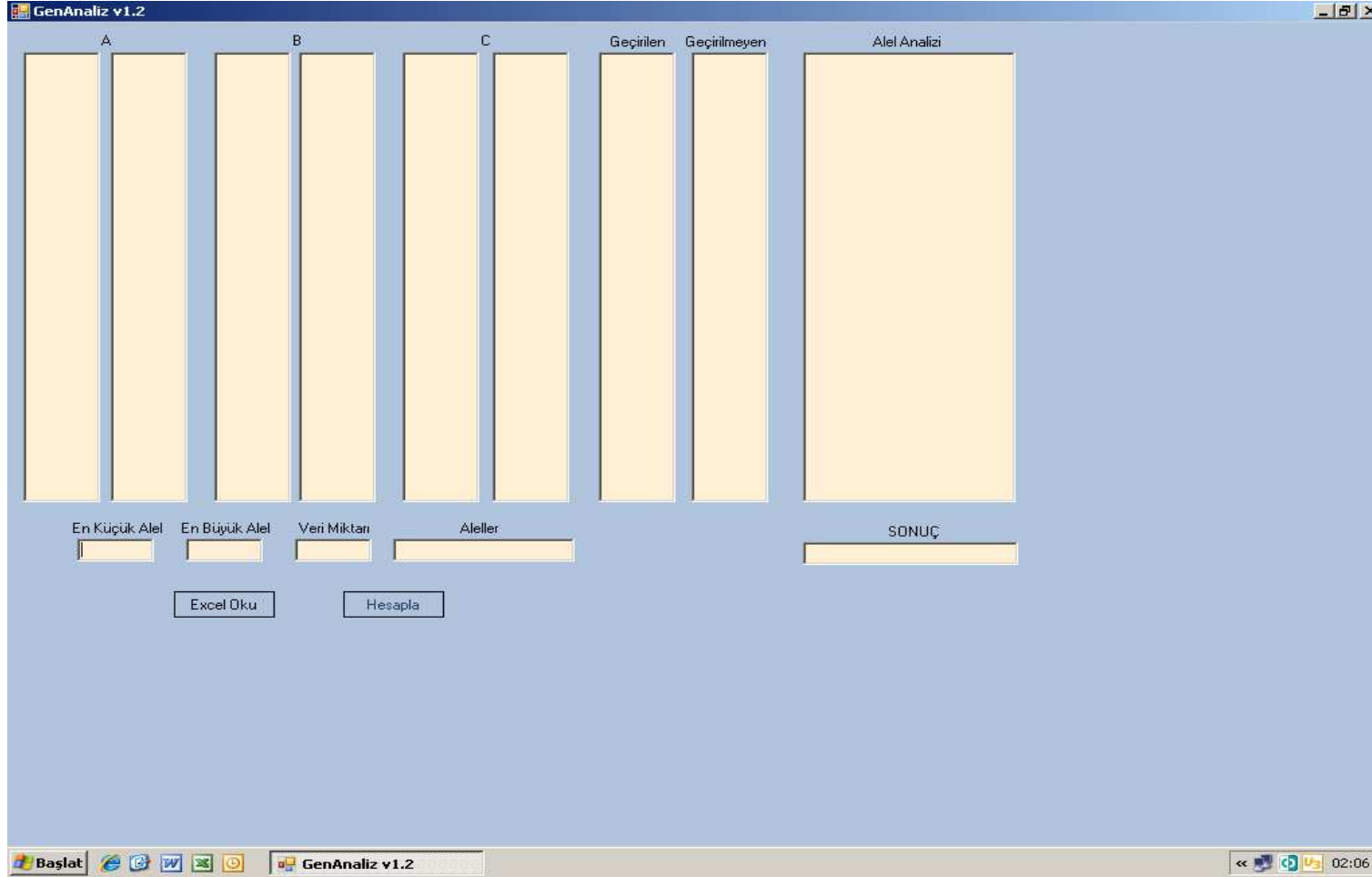
Bu çalışmada bipolar bozukluk tip 1 tanılı 93 ailenin anne-baba-hasta çocuk üçlülerinden oluşan veriye kalıtımda dengesiz aktarım testi uygulanmıştır. Bipolar bozukluk, manik depresif bozukluk olarak bilinen hastalıktır. Veriler Hacettepe Üniversitesi Tıp Fakültesi Tıbbi Genetik Anabilim Dalı Öğretim Üyesi Doç. Dr. Nurten Akarsu tarafından sağlanmıştır. Veriler 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 numaralı alellerden oluşan 10 alelli bir sistemde her bir bireyin marker genotiplerinden oluşmaktadır. Hasta bireylerin anne ve babalarının genotiplerine ulaşılabilirdiği için analiz için çok alelli TDT uygulanabilmektedir. Aynı zamanda TDT_i ve tek ebeveynin genotipine ulaşılabilirdiği zaman kullanılan 1-TDT analizi de uygulanmıştır.

2.1 Çok Alelli Marker İçin TDT_a Programı

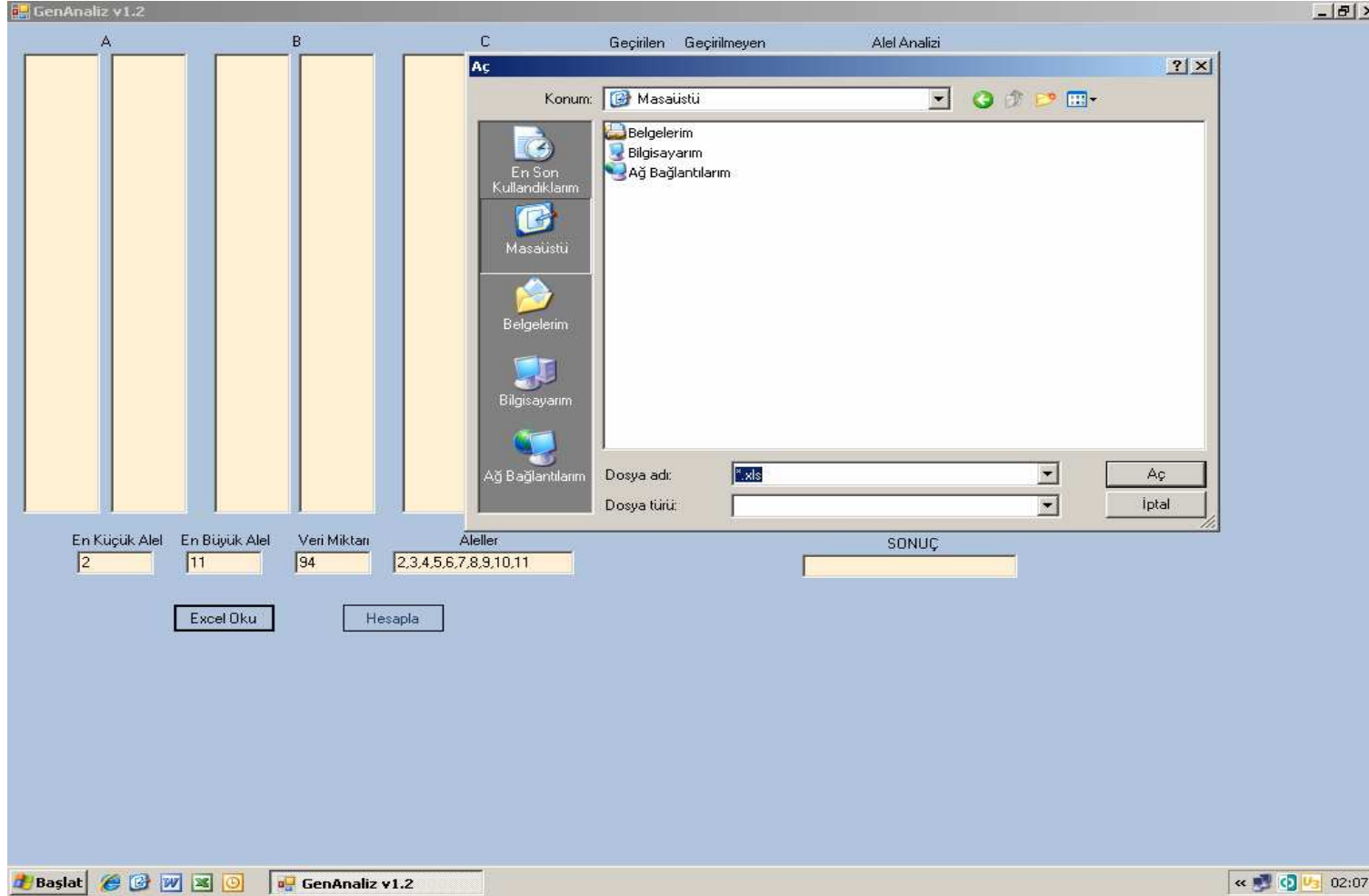
Çok alelli marker ile kalıtımda dengesiz aktarım testi yapmak için C# programlama dili kullanılarak GenAnaliz adı verilen program yazılmıştır. Yazılan kod Ek 1’de yer almaktadır.

Şekil 2.1’de program penceresinin görüntüsü mevcuttur. Programa “En küçük alel” yazısının altındaki kutucuğa verilerimizdeki sayısal değer olarak en küçük alel değeri girilerek başlanır. “En büyük alel” yazısının altında bulunan kutucuk da bezer şekilde doldurulduktan sonra “Veri miktarı” yazısının altındaki kutucuğa verilerimizin bulunduğu Excel sayfasındaki en son verimizin bulunduğu satır numarası yazılır. “Aleller” yazısının altındaki kutucuğa ise verilerimizde bulunan tüm alel değerleri aralarına virgül konulmak şartıyla girilir (örneğin 2,3,4,5). Artık programımız verilerin çağırılması için hazırdır. Sol tarafta yer alan A (anne), B (baba), ve C (çocuk) kolonları verilerin görüntülenmesi için konulmuştur. Verileri programa “Excel oku” düğmesi ile çağırırız. Veriler, programda “Excel oku” düğmesine bastığımız zaman açılan pencerede (Bakınız Şekil 2.2), verileri kaydettiğimiz

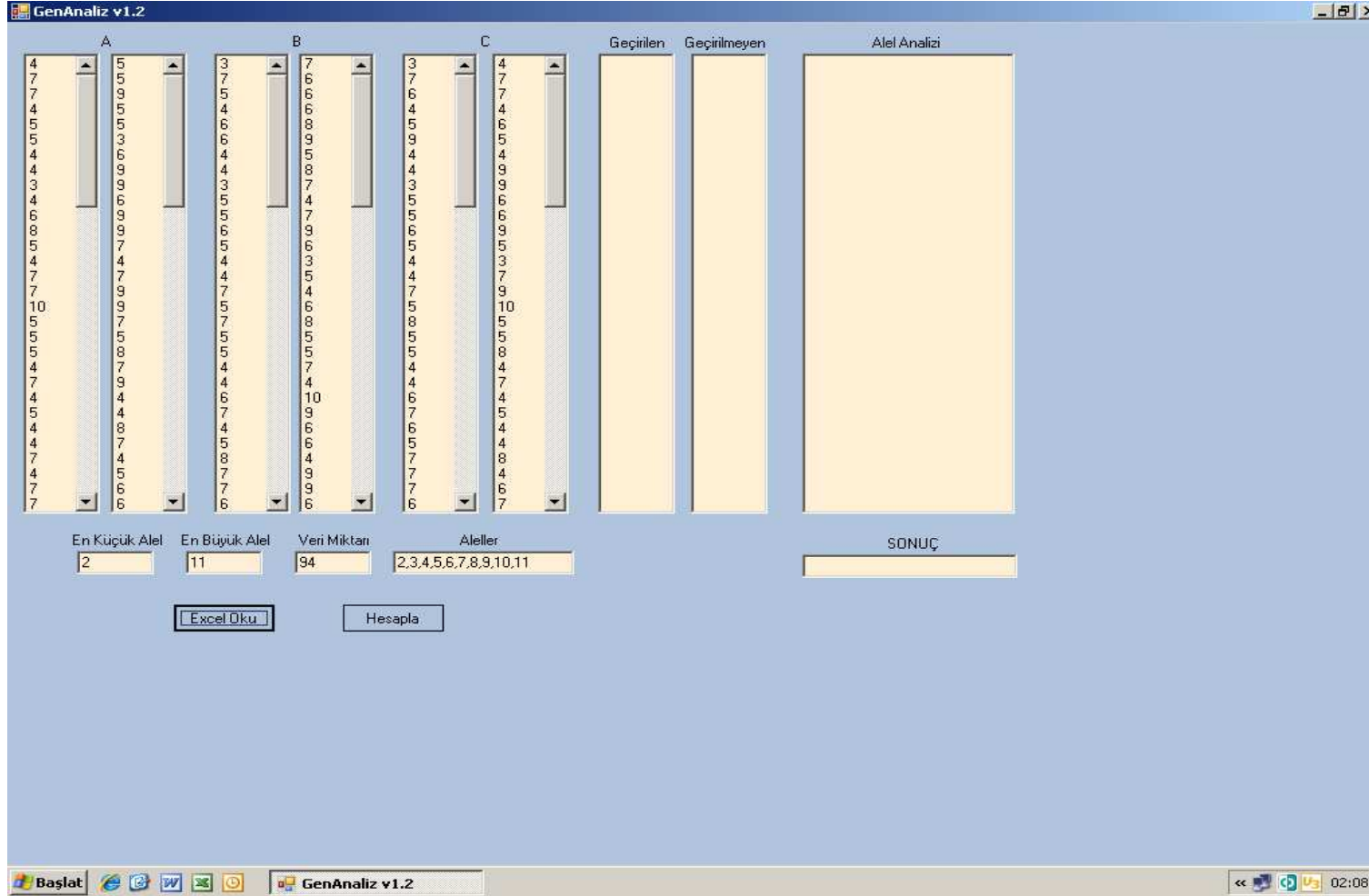
dosyanın yeri belirtildikten sonra “aç” veya “open” düğmesine basılınca yukarıda konusu geçen kolonlarda belirir (Şekil 2.3). Veriler 6 kolon (her bireye iki kolon ve bireyin her aleline bir kolon) ve aile sayısı kadar satır kullanılarak Excel’e girilmelidir. Her kolonda sırasıyla anne, baba ve hasta çocuğa ait marker genotipleri bulunmalıdır. Veriler ekrana geldikten sonra hesapla düğmesine basıldığında (Şekil 2.4) her veri satırına karşılık olarak geçirilen ve geçirilmeyen kolonlarında olmak üzere (anneden ve babadan şeklinde sıralı olarak) iki satır sonuç mevcuttur. Bu kolonların sağında bulunan “Alel analizi” kolonunda ise geçirilen ve geçirilmeyen alel çiftlerinden aynı olanların kaç kere tekrar ettikleri belirtilir. Burada x-x ifadesi sırası ile geçirilen-geçirilmeyen’i ifade etmektedir. “Sonuç” yazısının altında bulunan kutucukta ise program bize TDT değerini bildirir.



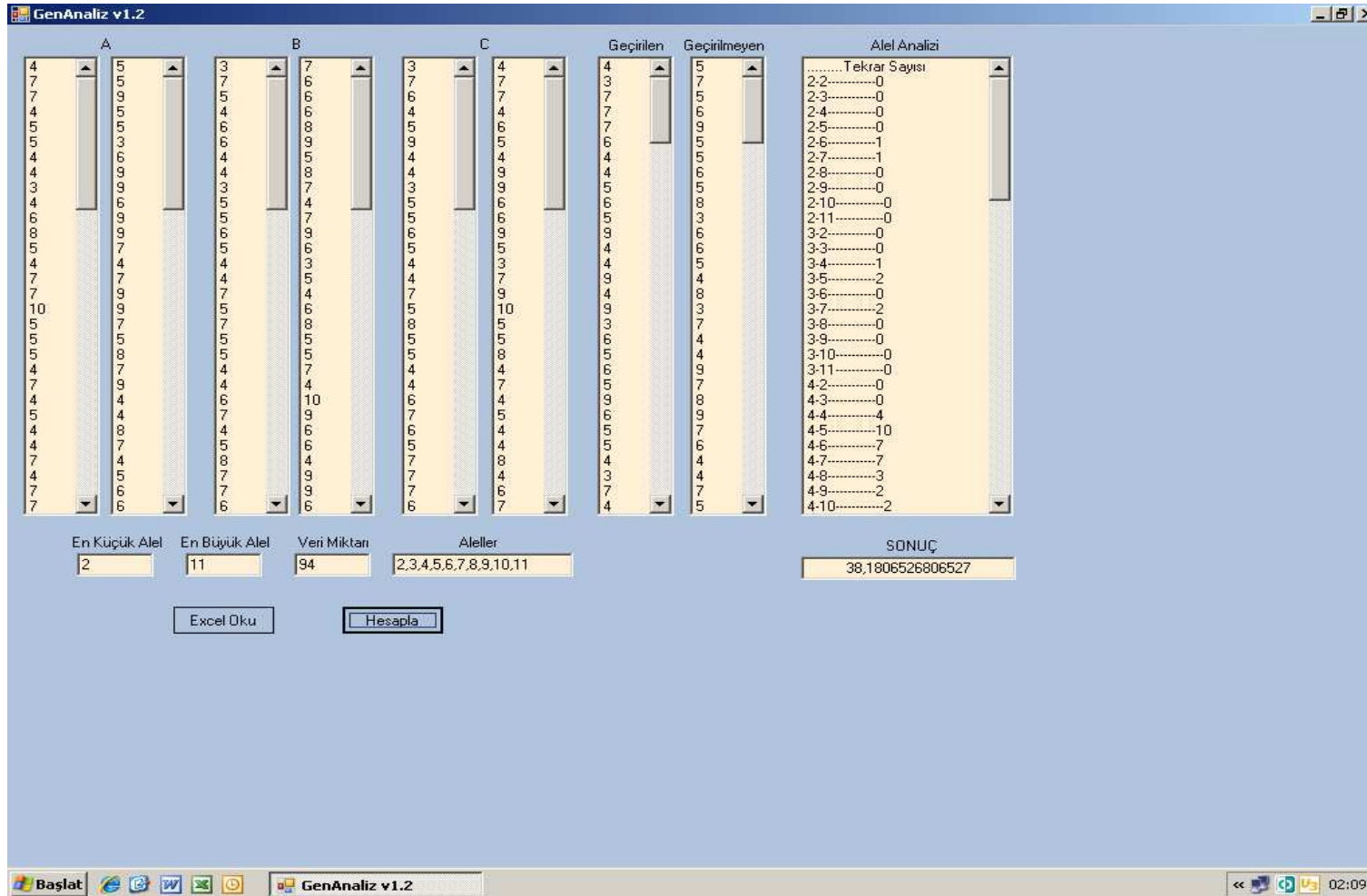
Şekil 2.1 Program görüntüsü 1



Şekil 2.2 Program görüntüsü 2



Şekil 2.3 Program görüntüsü 3



Şekil 2.4 Program görüntüsü 4

2.2 TDT_i Uygulaması

Veriler çok alelli sistemde oldukları için, her bir alel diğerlerine karşı incelenebilir. 10 alel, her biri diğerlerine karşı olmak üzere ayrı ayrı TDT_i değerleri hesaplanmış ve test değerleri bulunmuştur. Bu uygulamada da, yazılan, iki alelli sistemler için de geçerli olan GenAnaliz programı kullanılmıştır. Toplam 10 adet iki alelli TDT uygulanmıştır.

2.3 ETDT Uygulaması

Lojistik Regresyon yöntemine dayanan genişletilmiş TDT (ETDT) SPSS 11.5 paket programı kullanılarak uygulanmıştır. $m=10$ alelli marker genotiplerinden oluşan veriler 1.6.2.1.3 bölümünde anlatıldığı biçimde kukla değişkeniyle yeniden düzenlenmiştir. ETDT yönteminde ilgilenilen her bir genotip için $(m-1)=9$ farklı kukla değişkeni kullanılırken her bir kukla değişkeni $m(m-1)/2=45$ değerden oluşmuştur. Bu değerler daha önce anlatıldığı gibi ilgili alelin genotipte bulunma durumuna göre 1, -1 veya 0 olarak kodlanırlar.

Veriler SPSS'e şu şekilde girilmiştir: Bağımlı y değişkeni ilk 45 satırda 1, ikinci 45 satırda 0 değerini almıştır. Yani 9 sütundan oluşan kukla değişkenleri 45'er satır halinde 2 kez girilmiştir. 45 genotipin her biri için y bağımlı değişkeninin aldığı değere göre frekanslar hesaplanmış ve bu sütuna göre SPSS'e girilen veriler ağırlıklandırılmıştır.

2.4 1-TDT Uygulaması

93 anne-baba-çocuk genotipinden oluşan verilerden, rasgele olarak anne ya da babanın genotipleri seçilmiş ve 93 hasta çocuğun sadece bir ebeveynine ulaşılabildiği varsayılarak bu durumda kullanılabilecek 1-TDT uygulanmıştır. 1-TDT analizinde sözü geçen anne ya da babaya ait verilerin kayıp olma olasılığı aynı ($1/2$) olduğu için test istatistiğinin hesaplanmasında uygun formüller kullanılmıştır. Sonuçlar bulgular bölümünde yer almaktadır.

3. BULGULAR

3.1 TDT_a

Şekil 2.4'te verilerin bipolar bozukluk tip 1 tanılı 93 ailenin ana-baba-hasta çocuk üçlülerinden oluşan veriye çok alelli marker için yazılan program ile kalıtımda dengesiz aktarım testi uygulanmıştır. Kalıtımda aktarım dengeli ise, Tablo 1.6'daki değerlerin ortadaki verilere göre simetrik dağılması beklenir.

Denklem 1.39'a $TDT_a = \sum_{i < j} (t_{ij} - t_{ji})^2 / (t_{ij} + t_{ji})$, göre hesaplanmış olan

istatistik; $TDT_a = 38.18$ olarak bulunmuştur. Daha önce de belirtildiği gibi m alel sayısını vermek üzere, test istatistiği $TDT_a \sim \chi^2_{m(m-1)/2}$ dağılımına sahiptir. Uygulamamızda m=10 olduğu için %95'lik güvenilirlikle Ki-Kare tablosundan bulunan kritik değer; $\chi^2_{45} = 61.66$ dır.

$$TDT_a < \chi^2_{45}$$

Denklem 3.1

olduğu için hipotez kabul edilir (p=0.754). Başka bir deyişle, geçirilen ve geçirilmeyen marker alel dağılımlarının simetrik olmadığı kanıtlanamamıştır. Diğer bir deyişle marker ve hastalık lokusları arasında bağlantıdan söz edilememektedir.

3.2 TDT_i

Her bir alel ve geriye kalan diğer tüm aleller alınarak 10 ayrı TDT_i değeri hesaplanmıştır:

Tablo 3.1 TDT_i değerleri

	Alel 2	Alel 3	Alel 4	Alel 5	Alel 6	Alel 7	Alel 8	Alel 9	Alel 10	Alel 11
TDT _i	2	0.33	2.37	0.40	0.17	0.42	0.43	0	0.5	0.33

Bağımlılık söz konusu olduğu için Bonferroni düzeltmesi yapılmalıdır. O halde her bir TDT_i değeri $\chi^2_{1,0.05/10} = 7.88$ tablo değeriyle karşılaştırılmalıdır. Tablo 3.1’de alel bazında da anlamlı sonuca ulaşamadığı görülmektedir. Marker lokusla hastalık lokusu arasında ilişki bulunamamıştır.

3.3 ETD

Genişletilmiş TDT için SPSS 11.5 programı kullanılarak yapılan lojistik regresyon analizinden elde edilen parametre kestirimleri ve standart hataları Tablo 3.2’de verilmiştir.

Tablo 3.2 9 alel için parametre kestirimleri ve standart hataları

Alel No	b _i değeri	Standart Hata
2	20.116	28399.471
3	-1.194	1.382
4	-0.535	1.236
5	-1.039	1.257
6	-1.170	1.276
7	-1.008	1.270
8	-0.607	1.320
9	-0.589	1.304
10	-1.332	1.444

Ki-kare test istatistiği $\chi^2 = 2(L_1 - L_0) = 8.886$ olarak hesaplanmıştır. Test istatistiğinin dağılımı χ^2_9 'dir. $8.886 < \chi^2_{9,0.05} = 16.919$ ($p=0.448$) olduğu için marker lokusu ile hastalık lokusu arasında bağlantı yoktur yokluk hipotezini reddedecek yeterli kanıta ulaşamıyoruz.

3.4 1-TDT

Her bir alele karşılık diğer tüm aleller alınmış ve 1-TDT analizine anlatıldığı gibi Tablo 3.3'deki veriler elde edilmiştir.

Tablo 3.3 1-TDT için standart normal dağılıma göre analiz sonuçları

	Alel 2	Alel 3	Alel 4	Alel 5	Alel 6	Alel 7	Alel 8	Alel 9	Alel 10	Alel 11
b_1	1	0	14	10	12	10	7	12	3	2
c_1	0	4	10	0	16	15	4	9	3	0
V_1	1	4	24	10	28	25	11	21	6	2
T_1	1	-2	0.82	3.16	-0.76	-1	0.91	0.66	0	1.41

Çok alelli bir sistem olduğu için, test istatistiklerinin mutlak değerlerinin en büyük olanı sonucu yorumlamak için kullanılır. Bonferroni düzeltmesi yapılır. Test istatistiği $3.16 > z_{0.05/10} = 2.56$ olduğu için markerla hastalık arasında ilişki tespit edilmiştir.

Ancak $b_1 + c_1$ bazı aleller için yeterince büyük değildir. Bu nedenle Binom dağılım kullanılarak hesaplanan p değerleri, standart normal dağılım kullanılarak elde edilen sonuçlardan daha doğru sonuçlar verir.

Tablo 3.4 1-TDT için binom dağılıma göre sonuçlar

	Alel 2	Alel 3	Alel 4	Alel 5	Alel 6	Alel 7	Alel 8	Alel 9	Alel 10	Alel 11
p değeri	0.5	0.99	0.99	0.60	0.99	0.99	0.99	0.99	0.99	0.56

Tablodan marker ile hastalık lokusu arasında bir ilişki bulunamadığı görülmektedir.

4.TARTIŞMA

Bilimin ilerlemesi, teknolojinin gelişmesi ile birlikte hastalıkların nedenlerinin ve çözümlerinin araştırılmasında genetik çalışmaların önemi büyüktür. Ulaşılan çok sayıda genetik bilginin değerlendirilmesi için var olan yöntemler yetersiz kalmaktadır. Bu nedenle her geçen gün yeni analiz yöntemleri geliştirilmeye çalışılmaktadır.

İnsanın, toplumların karmaşık yapısı, genetik bilginin büyüklüğüyle birleşince, yapılması gereken pek çok varsayım sağlanamamaktadır. Bu nedenle olası karmaşıklığı önlemek için alınan önlemlerden birisi aile bazlı analiz yöntemleridir. Bu analiz yöntemleri toplumdaki tabakalaşmadan etkilenmediği için tercih edilmektedir. Aynı zamanda hasta bireylerin ailelerindeki sağlıklı bireylere ulaşmak da genellikle zor değildir. Anne ve babanın genotipine ulaşılabilirdiği zaman Kalıtımda Dengesiz Aktarım Testi uygulanabilmektedir (Spielman ve ark., 1993). Ebeveynlerden sadece birine ulaşılabilirdiği durumlar için 1-TDT, sağlıklı kardeş ya da kardeşlere ulaşılabilirdiği durumlar için ise S-TDT, SDT gibi çeşitli testler geliştirilmiştir. En güçlü sonuç ebeveynlerin her ikisinin de genotipine ulaşılabilirdiği zaman alınmaktadır.

Çalışmamızda bipolar bozukluk tip 1 tanılı 93 ailenin ana-baba-hasta çocuk üçlülerinden oluşan veriye kalıtımda dengesiz aktarım testi uygulanmıştır. Veriler çok alelli kalıtımda dengesiz aktarım testine uygundur. Bu nedenle hem Genelleştirilmiş TDT hem de Genişletilmiş TDT uygulanmıştır. Aynı zamanda çok alelli sistemde her bir alel için hesaplanan TDT_i yöntemi de kullanılmıştır. Bunların yanı sıra, anne ve babanın genotiplerinin bilinmesine karşın, yalnız bir ebeveynin verisine ulaşıldığı durumlar için kullanılan 1-TDT analiz yöntemi de rasgele olarak bir ebeveynin seçilmesiyle uygulanmıştır.

Sonuçlara bakıldığında marker ile hastalık lokusu arasında beklenen ilişki istatistiksel olarak sağlanamamıştır. Bu durum verilerin çok markerli sistemde ancak az sayıda olmasından kaynaklanmış olabilir. Veriler, bireylerin 10 alelli sistemdeki marker genotiplerinden oluşmaktadır. 10 alelli sistem olduğundan dolayı, daha sağlıklı sonuç alınabilmesi için daha fazla aileye ihtiyaç duyulmaktadır. 93 aile her bir marker çeşidinin aktarılma ve aktarılmama dağılımının çıkarılması ve aktarımda söz konusu olabilecek dengesizliğin ortaya çıkarılması için yeterli değildir.

5. SONUÇ VE ÖNERİLER

Bu çalışmanın amacı, ülkemiz için yeni olan genetik analiz yöntemlerinden aile bazlı olan yöntemleri tanıtmaktır. Bu yöntemlerden çok alelli sistemde kalıtımda dengesiz aktarım testi, 93 ailenin anne, baba ve çocuk üçlülerinden oluşan verilere uygulandı.

Bu çalışmada olduğu gibi hipotezin reddedilememesi durumunda aile üçlülerinden oluşan veri sayısı artırılıp analiz tekrar yapılabilir. Ya da “İlgilenilen markerla hastalık lokusu arasında bir bağlantı yoktur.” sonucu ışığında şüpheli başka bir markerla yeni bir çalışma başlatılabilir.

İleride yapılacak olan çalışmalarda geliştirilmekte olan farklı aile bazlı analiz yöntemleri de incelenip bu yöntemlerle karşılaştırılabilir. Ayrıca az sayıdaki verilerle alel sayısı çok olduğunda yapılabilecek daha güçlü testler geliştirilebilir.

ÖZET

Kalıtımda Dengesiz Aktarım Testi

Hastalıklara çözüm getirilebilmesi için, DNA üzerindeki genetik kodların çözülmesiyle, hastalıklara neden olan genler ve bu genlerin ebeveynlerden çocuklarına aktarılması araştırılmaktadır. Hastalıkla genler arasındaki ilişkiyi göstermek amacıyla yapılan ilişkilendirme analiz yöntemleri, toplumda oluşan tabakalaşma nedeniyle yetersiz kaldığı için, Spielman ve ark. tarafından bu tabakalaşmadan etkilenmeyen Kalıtımda Dengesiz Aktarım Testi (TDT) geliştirilmiştir.

Kalıtımda Dengesiz Aktarım Testi; ilişki olduğu durumda gösterge ve hastalık geni arasındaki bağlantıyı test eden güçlü bir yöntemdir. Bu testte anne baba çocuk üçlüsü üzerinden çalışılır. Ebeveynlerin belirli bir lokustaki çocuklarına aktardıkları gösterge aleli ile aktarmadıkları alel birbiri ile karşılaştırılır. Böylelikle ilgilenilen lokusta belirli bir alelin Mendel Kalıtım Kuralları'ndan beklenen orandan daha yüksek bir oranda kalıtılıp kalıtılmadığı araştırılır.

Bu çalışmada aile bazlı ilişki analizi yöntemleri tartışılmış, anne-baba-hasta çocuk üçlülerinden oluşan verilere kullanılabilir yöntemler uygulanmıştır.

Anahtar sözcükler: TDT, ilişki çalışmaları, istatistiksel genetik, popülasyonda tabakalaşma, bağlantı dengesizliği.

SUMMARY

Transmission Disequilibrium Test

To find treatments or solutions to diseases, genes that cause diseases are investigated and researchers try to find how these genes are transmitted from parents to affected offsprings. However, because of population stratification, the methods used for detecting the association between diseases and genes are not valid. Spielman et al. developed Transmission Disequilibrium Test which is not affected by the population stratification. (1993)

Transmission disequilibrium test is a powerful test which detects linkage between the marker and disease locus in cases where association between the marker and disease is found. Family trios are used for this test. In a specific marker locus, the alleles that are transmitted from parents to the affected offspring and the alleles that are not transmitted are compared. Therefore, in that locus, it is investigated that whether the alleles are transmitted according to the Mendel's Inheritance Laws or not.

In this thesis study, family based association methods are discussed and the suitable analysis methods are applied to the data consisting of family trios.

Key Words: TDT, association studies, statistical genetics, population stratification, linkage disequilibrium.

KAYNAKLAR

- ACKERMAN, H., USEN, S., JALLOW, M., SISAY-JOOF, F., PINDER, M. (2005). A comparison case-control and family based association methods: The example of sickle-cell and malaria. *Annals of Human Genetics*. **69**: 559-565.
- AKARSU, N.A., ÇAKIR, B. (2004). Psikiyatrik Genetik Araştırmalarında Kullanılabilecek Genetik Yöntemler:IV.A.Hastalık Geni Haritalanması. *3P DERGİSİ* **12**:31-49
- ALMGREN, P., BENDAHL, P., BENGTSSON, H., HÖSSJER, O., PERFEKT, R., (2003). Statistics in Genetics Lecture Notes. *Lund University*.
- ALMGREN, P., BENDAHL, P., BENGTSSON, H., HÖSSJER, O., PERFEKT, R. (2002). Statistics in Genetics Lecture Notes. Erişim: [<http://www.maths.lth.se/matstat/kurser/statgen/book/StatisticsInGenetics-20031125.pdf>] Erişim Tarihi: 10.03.2007
Erişim: [<http://www.sizes.com/units/centimorgan.htm>] Erişim Tarihi:9.04.2007
- ARMITAGE, P., COLTON, T. (1998). Encyclopedia of Biostatistics. John Wiley & Sons, West Sussex, England.
- BALDING, D.J., BISHOP, M., CANNINGS, C. (2003). Handbook of Statistical Genetics. *John Wiley & Sons, West Sussex*.
- BOEHNKE, M., LANGEFELD, C.D. (1998). Genetic Association Mapping Based on Discordant Sib Pairs: The Discordant-Alleles Test. *Am. J. Hum. Genet.* **62**: 950-961.
- CHEN, W., DENG, H. (2001). A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. *Genetic Epidemiology* **21**: 53-67.
- CLAYTON, D., JONES H., (1999). Transmission/Disequilibrium Tests for Extended Marker Haplotypes *Am. J. Hum. Genet.* **65**:1161–1169.
- CURTIS, D. (1997). Use of siblings as controls in case-control association studies. *Ann. Hum. Genet.* **61**: 319-333.
- CURTIS, D. SHAM, P.(1995) A note on the application of the transmission disequilibrium test when a parent is missing. *Am. J. Hum. Genet.*, **56**, 811-812.
- EWENS W.J., SPIELMAN R.S., (1995). The Transmission/Disequilibrium Test: History, Subdivision, and Admixture. *Am. J. Hum. Genet.* **57**:455-464.
- FALK, C.T., RUBINSTAIN, P. (1987). Haplotype relative risks: An easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51**: 227-233.

- GENE THERAPY SOLUTIONS. (2004). Eriřim:[<http://www.genesolutions.com/>].
Eriřim tarihi: 4.12.2006
- GORDON, D., HAYNES, J., JOHNNIDIS, C., PATEL, S.B. (2004). A transmission disequilibrium test for general pedigrees taht is robust to the presence of random genotyping errors and any number of untyped parents. *Eur. J. Hum. Genet.* **12(9)**: 752-761.
- GRIFFITHS AJF, MILLER JH, SUZUKI DT, LEWONTIN RC, and GELBART WM (2000). *An Introduction to Genetic Analysis*. New York: W.H. Freeman and Company. [ISBN 0-7167-3520-2](#). genetik tanım kaynagi
- GUO, W. FUNG, W.K. (2005) Combining the case-control methodology with the small size transmission/disequilibrium test for multiallelic markers. *European Journal of Human Genetics*, **13**, 1007-1012.
- HARDY, G.H. 1908. 'Mendelian proportions in a mixed population.' *Science*, vol.28, 49-50.
- HORVATH, S., LAIRD, N.M. (1998). A discordant-sibship test for disequilibrium and linkage: No need for parental data. *Am. J. Hum. Genet.* **63**: 1886-1897.
- KANEKO, N., MURATAKE, T. AMAGANE, H., SAKURAI, M., TANAKA, T., TSUJI, S., SOMEYA, T., (2004). Transmission disequilibrium test and haplotype analysis of the *NOTCH4* gene in Japanese patients with schizophrenia. *Psychiatry and Clinical Neurosciences.* **58**: 199–205.
- KAPLAN, N. L., MARTİN, E. R., MORRIS, R. W., WEİR, B. S., (1998). Marker Selection for the Transmission/Disequilibrium Test, in Recently Admixed Populations. *Am. J. Hum. Genet.* **62**: 703-712.
- LAZZERONI, L.C., LANGE, K. (1998). A conditional inference frame work for extending the transmission/disequilibrium test. *Hum. Heret.* **48**: 67-81.
- MARTIN, E.R., BASS, M.P., KAPLAN, N.L. (2001). Correcting for a potential bias in the pedigree disequilibrium test. *Am. J. Hum. Genet.* **68**: 1065-1067.
- MARTIN, E.R., KAPLAN N.L., WEIR, B.S. (1997). Test for linkage and association in nuclear families. *Am. J. Hum. Genet.* **61**: 439-448.
- MARTIN, E.R., MONKS, S.A., WARREN L.L., KAPLAN N.L., (2000). A Test for Linkage and Association in General Pedigrees: The Pedigree Disequilibrium Test *Am. J. Hum. Genet.* **67**:146–154.
- McGINNIS, R. E.,(1998). Hidden linkage: a comparison of the affected sib pair (ASP) test and transmission/disequilibrium test (TDT). *Ann. Hum. Genet.* **62**: 159-179.
- MORRIS, A.P., CURNOW, R.N., WHITTAKER, J.C. (1997). Randomization tests of disease-marker associations. *Ann. Hum. Genet.* **61**: 49-60.
- MONKS, S.A., KAPLAN, N.L., WEIR, B.S. (1998). A comparative study of sibship test of linkage and/or association. *Am. J. Hum. Genet.* **63**: 1507-1516.

- ROCKEFELLER UNIVERSITY, Lab of Statistical Genetics. (2005). An Alphabetic List of Genetic Analysis Software.
Eriřim:[<http://linkage.rockefeller.edu/soft/list4.html>] Eriřim tarihi: 6.8.2006.
- RUBINSTEIN, P. WALKER, M. CARPENTER, C. CARRIER, C. KRASSNER, J. FALK, C. GINSBERG, F. (1981). Genetics of HLA disease associations: the use of the haplotype relative risk (HRR) and the 'haplo-delta' (Dh) estimates in juvenile diabetes from three racial groups. *Human Immunology*, **3**, 384.
- SHAM, P. (1998) Statistics in Human Genetics. *Arnold Publisher, NY*.
- SHAM, P.C., CURTIS, D. (1995). An extended transmission disequilibrium test (TDT) for multi-allele marker loci. *Ann. Hum. Genet.* **59**: 323-336.
- SHEPHARD, N. (2005). Guide to Genetic Analysis. Eriřim: [http://slack.ser.man.ac.uk/progs/genetic_analysis.html] Eriřim tarihi: 10.11.2006.
- SPIELMAN, R.S., EWENS, W.J. (1998). A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test. *Am. J. Hum. Genet.* **62**: 450-458.
- SPIELMAN, R.S., MCGINNIS, R.E. and EWENS, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *Am. J. Hum. Genet.*, **52**, 506-516.
- SUN, F., FLANDERS, W.V., YANG, Q., KHOURY, M.J. (1999). The transmission disequilibrium test (TDT) when only one parent is available: The 1-TDT. *American Journal of Epidemiology* **150(1)**: 97-104.
- WALDMAN, I. D., ROBINSON, B. F., ROWE D. C. (1999). A logistic regression based extension of the TDT for continuous and categorical traits. *Ann. Hum. Genet.* **63**: 329-340.
- WANG, D., SUN, F., (2000). Sample sizes for the transmission Disequilibrium tests: TDT, S-TDT and 1-TDT. *Commun.stat., Theory methods* **29**: 1129-1142.
- WEILING F (1991). "Historical study: Johann Gregor Mendel 1822-1884". *American Journal of Medical Genetics* **40** (1): 1-25; discussion 26. [PMID 1887835](#)
- WEINBERG, C.R. (1999). Allowing for missing parents in genetic studies of case-parent triads. *Am. J. Hum. Genet.* **64**: 1186-1193.
- WHITTAKER, J. C., MORRIS, A. P. (2001). Family-based tests of association and/or linkage. *Ann. Hum. Genet.* **65**: 407-419.
- WICKS, J. (2000). Exploiting excess sharing: A more powerful test of linkage for affected sibpairs than the transmission/disequilibrium test. *Am. J. Hum. Genet.* **66**: 2005-2008.

- WICKS, J., WILSON, S.R. (2000). Evaluating linkage and linkage disequilibrium: use of excess sharing and transmission disequilibrium methods in affected sib pairs. *Ann. Hum. Genet.* **64**: 419-432.
- ZHAO, H. (2000). Family-based association studies. *Stat. Methods Med. Res.* **9**: 563.

EK 1

```

using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Text;
using System.Windows.Forms;
using Microsoft.Office.Interop.Excel;
namespace GenAnaliz
{
    public partial class Form1 : Form
    {
        private Microsoft.Office.Interop.Excel.Application ExcelNesnesi = null;
        int selectedIndex;
        int alel_sayisi;
        int max_alel;
        int min_alel;
        public Form1()
        {
            InitializeComponent();
            ExcelNesnesi = new Microsoft.Office.Interop.Excel.Application();

            if (ExcelNesnesi == null)
            {
                MessageBox.Show("Problem! Dosya Açılmadı.");
                System.Windows.Forms.Application.Exit();
            }
        }
        private void btnHesapla_Click(object sender, EventArgs e)
        {
            lbFrekans.Items.Clear();
            max_alel = Convert.ToInt32(txtMaxAlel.Text);
            min_alel = Convert.ToInt32(txtMinAlel.Text);
            string[] str_aleller = txtAleller.Text.Split(';');
            alel_sayisi = str_aleller.Length;
            int[] aleller = new Int32[alel_sayisi];
            for (int i = 0; i < alel_sayisi; i++)
            {
                aleller[i] = Convert.ToInt32(str_aleller[i]);
            }
            int elemanSayisi;
            elemanSayisi = lbA1.Items.Count;
            bool kontrol = true;
            for (int i = 0; i < elemanSayisi; i++)
            {
                kontrol = true;
                if ((lbC1.Items[i].ToString() == lbA1.Items[i].ToString()))
                {
                    if ((lbC2.Items[i].ToString() == lbB1.Items[i].ToString()) && kontrol)
                    {
                        lbG.Items.Add(lbA1.Items[i].ToString());
                        lbK.Items.Add(lbA2.Items[i].ToString());
                        lbG.Items.Add(lbB1.Items[i].ToString());
                        lbK.Items.Add(lbB2.Items[i].ToString());
                        kontrol = false;
                    }
                }
            }
        }
    }
}

```

```

    }
    if ((lbC2.Items[i] == lbB2.Items[i]) && kontrol)
    {
        lbG.Items.Add(lbA1.Items[i].ToString());
        lbK.Items.Add(lbA2.Items[i].ToString());
        lbG.Items.Add(lbB2.Items[i].ToString());
        lbK.Items.Add(lbB1.Items[i].ToString());
        kontrol = false;
    }
}
if ((lbC1.Items[i].ToString() == lbA2.Items[i].ToString()))
{
    if ((lbC2.Items[i].ToString() == lbB1.Items[i].ToString()) && kontrol)
    {
        lbG.Items.Add(lbA2.Items[i].ToString());
        lbK.Items.Add(lbA1.Items[i].ToString());
        lbG.Items.Add(lbB1.Items[i].ToString());
        lbK.Items.Add(lbB2.Items[i].ToString());
        kontrol = false;
    }
    if ((lbC2.Items[i].ToString() == lbB2.Items[i].ToString()) && kontrol)
    {
        lbG.Items.Add(lbA2.Items[i].ToString());
        lbK.Items.Add(lbA1.Items[i].ToString());
        lbG.Items.Add(lbB2.Items[i].ToString());
        lbK.Items.Add(lbB1.Items[i].ToString());
        kontrol = false;
    }
}
//*****
if ((lbC1.Items[i].ToString() == lbB1.Items[i].ToString()))
{
    if ((lbC2.Items[i].ToString() == lbA1.Items[i].ToString()) && kontrol)
    {
        lbG.Items.Add(lbA1.Items[i].ToString());
        lbK.Items.Add(lbA2.Items[i].ToString());
        lbG.Items.Add(lbB1.Items[i].ToString());
        lbK.Items.Add(lbB2.Items[i].ToString());
        kontrol = false;
    }
    if ((lbC2.Items[i].ToString() == lbA2.Items[i].ToString()) && kontrol)
    {
        lbG.Items.Add(lbA2.Items[i].ToString());
        lbK.Items.Add(lbA1.Items[i].ToString());
        lbG.Items.Add(lbB1.Items[i].ToString());
        lbK.Items.Add(lbB2.Items[i].ToString());
        kontrol = false;
    }
}
if ((lbC1.Items[i].ToString() == lbB2.Items[i].ToString()))
{
    if ((lbC2.Items[i].ToString() == lbA1.Items[i].ToString()) && kontrol)
    {
        lbG.Items.Add(lbA1.Items[i].ToString());
        lbK.Items.Add(lbA2.Items[i].ToString());
        lbG.Items.Add(lbB2.Items[i].ToString());
        lbK.Items.Add(lbB1.Items[i].ToString());
    }
}

```

```

        kontrol = false;
    }
    if ((lbC2.Items[i].ToString() == lbA2.Items[i].ToString()) && kontrol)
    {
        lbG.Items.Add(lbA2.Items[i].ToString());
        lbK.Items.Add(lbA1.Items[i].ToString());
        lbG.Items.Add(lbB2.Items[i].ToString());
        lbK.Items.Add(lbB1.Items[i].ToString());
        kontrol = false;
    }
}
}
double[,] GK = new double[50,50];

for (int i = 0; i <= max_alel; i++)
{
    for (int j = 0; j <= max_alel; j++)
    {
        GK[i, j] = 0;
    }
}
int x, y;
//MessageBox.Show(lbG.Items[0].ToString());
//MessageBox.Show(lbG.Items.Count.ToString());
for (int i = 0; i < (elemanSayisi * 2)-2; i++)
{
    x = Convert.ToInt32(lbG.Items[i].ToString());
    y = Convert.ToInt32(lbK.Items[i].ToString());
    GK[x, y] = GK[x, y] + 1;
}
lbFrekans.Items.Add(".....Tekrar Sayısı");
double toplam = 0;
bool hata=false;
for (int i = min_alel; i <= max_alel; i++)
{
    foreach (int u in aleller)
    {
        if(i==u)
        {
            for (int j = min_alel; j <= max_alel; j++)
            {
                foreach (int v in aleller)
                {
                    if ((j == v))
                    {
                        lbFrekans.Items.Add(i.ToString() + "-" + j.ToString() + "-----" +
GK[i, j].ToString());
                        if (GK[i, j] + GK[j, i] == 0)
                        {
                            hata = true;
                        }
                        else
                        {
                            if ((i > j))
                            {
                                toplam = toplam + ((GK[i, j] - GK[j, i]) * (GK[i, j] - GK[j, i]) / (GK[i,
j] + GK[j, i]));
                            }
                        }
                    }
                }
            }
        }
    }
}

```

```

    }
    }
    }
    }
    }
    }
    if (hata)
    {
        MessageBox.Show("Hatalı veri var, Sonuç hatalı olabilir.");
        txtSonuc.Text = toplam.ToString();
    }
    else
    {
        txtSonuc.Text = toplam.ToString();
    }
}
private void btnOku_Click(object sender, EventArgs e)
{
    lbA1.Items.Clear();
    lbB1.Items.Clear();
    lbC1.Items.Clear();
    lbA2.Items.Clear();
    lbB2.Items.Clear();
    lbC2.Items.Clear();
    lbG.Items.Clear();
    lbK.Items.Clear();
    lbFrekans.Items.Clear();
    this.openFileDialog1.FileName = "*.xls";
    if (this.openFileDialog1.ShowDialog() == DialogResult.OK)
    {
        Microsoft.Office.Interop.Excel.Workbook theWorkbook =
            ExcelNesnesi.Workbooks.Open(
                openFileDialog1.FileName,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing,
                Type.Missing);
        Microsoft.Office.Interop.Excel.Sheets sheets = theWorkbook.Worksheets;
        Microsoft.Office.Interop.Excel.Worksheet worksheet =
            (Microsoft.Office.Interop.Excel.Worksheet)sheets.get_Item(1); //sadece ilk çalışma sayfası
        için
        int satirSayisi = Convert.ToInt32(txtVeriMiktari.Text);
        for (int i = 2; i <= satirSayisi + 1; i++)
        {

```

```

        Microsoft.Office.Interop.Excel.Range range = worksheet.get_Range("A" +
i.ToString(), "F" + i.ToString());
        System.Array myvalues = (System.Array)range.Cells.Value2;
        string[] strArray = ConvertToStringArray(myvalues);
        lbA1.Items.Add(strArray[0]);
        lbA2.Items.Add(strArray[1]);
        lbB1.Items.Add(strArray[2]);
        lbB2.Items.Add(strArray[3]);
        lbC1.Items.Add(strArray[4]);
        lbC2.Items.Add(strArray[5]);
    }
    ExcelNesnesi.Workbooks.Close();
    btnHesapla.Enabled = true;
}
}
string[] ConvertToStringArray(System.Array values)
{
    string[] theArray = new string[values.Length];
    for (int i = 1; i <= values.Length; i++)
    {
        if (values.GetValue(1, i) == null)

            theArray[i - 1] = "";
        else
            theArray[i - 1] = (string)values.GetValue(1, i).ToString();
    }
    return theArray;
}
private void lbA1_Click(object sender, EventArgs e)
{
    selectedIndex = lbA1.SelectedIndex;
    lbA2.SelectedIndex = selectedIndex;
    lbB1.SelectedIndex = selectedIndex;
    lbB2.SelectedIndex = selectedIndex;
    lbC1.SelectedIndex = selectedIndex;
    lbC2.SelectedIndex = selectedIndex;
    lbG.SelectedIndex = (selectedIndex * 2);
    lbK.SelectedIndex = (selectedIndex * 2);
}

private void tsmCikis_Click(object sender, EventArgs e)
{
    this.Close();
}

```

ÖZGEÇMİŞ

Bireysel Bilgiler:

Adı Soyadı : Umut Özbek
 Doğum Yeri ve Tarihi : Ankara 12/06/1982
 Uyuşu : T.C.
 Medeni Durumu : Bekar
 İletişim Adresi : Ankara Üniversitesi Tıp Fakültesi Dekanlık Binası
 Biyoistatistik A.D. 06100 Sıhhiye, Ankara
 Telefonu : 312 310 30 10/375

Eğitim Durumu:

2004-	Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D. Biyoistatistik Yüksek Lisansı,	Ankara
1999-2004	Orta Doğu Teknik Üniversitesi Fen-Edebiyat Fakültesi, İstatistik Bölümü İstatistik Lisansı	Ankara
1996-1999	Özel Bilim Fen Lisesi	Ankara

Sözlü Bildiriler

9. Ulusal Biyoistatistik Kongresi, bildiri sunumu: '*Kalıtımda Dengesiz Aktarım Testi*' (2006 Eylül)
8. Ulusal Biyoistatistik Kongresi, bildiri sunumu: '*Tekrarlı Ölçümlerin Eğri Altında Kalan Alan Yardımıyla Analizi*' (2005 Eylül)
4. İstatistik Kongresi, bildiri sunumu: '*Klinik Denemelerde Koşullu Güce Bağlı Olarak Örneklem Büyüklüğünün Hesaplanması*' (2005 Mayıs)

Kurslar

2006 Haziran İstatistiksel Genetik Yaz Kursu, (burslu olarak)

Washington Üniversitesi, Seattle WA, A.B.D.

2006 Mayıs Kompleks Fenotiplerin İstatistiksel Genetik Analizi, Hibrid kurs

Avrupa Genetik Kuruluşu

Yabancı Dil

İngilizce (iyi derecede)

Almanca (başlangıç düzeyinde)

Bilgisayar Bilgisi

C, SQL

SPSS, MINITAB, MATLAB, SAS, Microsoft Office ve İnternet Uygulamaları

Üyelikler

İstatistik Mezunları Derneği

Uluslararası Biyometri Topluluğu

ODTÜ İstatistik Topluluğu

ODTÜ Orienteering Topluluğu

ODTÜ Tiyatro Topluluğu