

ROMATOLOJİDE KLİNİK ARAŞTIRMALAR VE NESNEL TIP III

**EDİTÖR
HASAN YAZICI**

**B. ÇAKIR
Ö. ERGÖNÜL
H. BRENNER
H. YAZICI
R. SMITH
M. İNANÇ
N. AKKOÇ**



ROMATOLOJİDE KLİNİK
ARAŞTIRMALAR VE
NESNEL TIP III

108

Romatolojide Klinik Arařtırmalar ve Nesnel Tıp III

Editör /

Prof. Dr. Hasan Yazıcı

İ. Ü. Cerrahpařa Tıp Fakóltesi

İç Hastalıkları Anabilim Dalı

Romatoloji Bilim Dalı

İstanbul

Birinci baskı © Deomed, 2010.

Yayın hakları Romatoloji Arařtırma ve Eđitim Derneđi adına Deomed Reklam ve Yayıncılık Ltd. řti.'ne aittir. 5846 Sayılı Fikir ve Sanat Eserleri Yasası uyarınca tamamı ya da bir bölümü, resim, tablo, řekil ve grafikler yazılı izin alınmaksızın elektronik ya da mekanik yöntemlerle kopya edilemez, çođaltılamaz ve yayınlanamaz. Kaynak olarak gösterilmek kořuluyla alıntı yapılabilir.

Yazıcı (Ed.)

Romatolojide Klinik Arařtırmalar ve Nesnel Tıp III

17 x 24 cm, VI + 82 Sayfa

19 Tablo, 42 řekil

ISBN 978-975-8882-24-3

Önemli Uyarı:

Bu kitabın içeriđi, temel ve klinik arařtırmalarla sürekli gelişmekte olan tıp biliminin güncel bilgi düzeyi çerçevesinde oluşturulmuřtur. Editör ve yazarlar, kitabın hazırlandığı zaman diliminde, özellikle tanı ve tedavi uygulamaları ile ilgili olanlar bařta olmak üzere tüm verilerin güncel olması ve hatasız basılması için özen göstermişlerdir. Konuyla ilgili bilgilendirme amacıyla hazırlanmış olan bu kitapta yer alan bilgi ve öneriler okuyucunun sađlık sorunları ile ilgili tıbbi tanı ve tedavi yerine geçmez. Tanı ve tedavi uygulamaları konusundaki sorumluluk kullanıcıların kendilerine aittir; yayınevi, editör, yazar, çevirmen, telif hakkı sahibi ve varsa sponsor için bađlayıcı deđildir. Kitapta adı geçen ticari markalar, özellikle ilaç adları bilgilendirme dışında bir amaçla sunulmamaktadır. Kullanıcıların ilgili üreticilerin ürün bilgi kılavuzları ya da prospektüs bilgilerine bakmaları, řüpheli durumlarda bir uzmana danışmaları önerilir. Ayrıca kuřku uyandıracak pratik durumların yayınevine bildirilmesi rica olunur.

deomed®

Deomed Medikal Yayıncılık

Acıbadem Cad. İsmail Hakkı Bey Sok. Pehlivan İş Merkezi

No: 7 Kat: 1 34718 Kadıköy / İstanbul

Tel: (0216) 414 83 43 Pbx Faks: (0216) 414 83 42

www.deomed.com / e-posta: medya@deomed.com

Baskı / Birmat Matbacılık (0212) 629 05 59

Cilt / Güven Mücellit (0212) 445 00 04

İÇİNDEKİLER

*Romatolojide Klinik Araştırmalar ve
Nesnel Tıp III*

Önsöz	V	III
Randomizasyon <i>Banu Çakır</i>	1	
Lojistik Regresyon <i>Önder Ergönül</i>	15	
<i>Propensity</i> (Eğilim) Analizi <i>Önder Ergönül</i>	23	
Uzun Dönemli Sağkalımda Periyot Analizi <i>Hermann Brenner</i>	27	
Sınıflandırma Kriterleri <i>Hasan Yazıcı</i>	37	
Bilgilendirilmiş Olur (<i>Informed Consent</i>) <i>Richard Smith</i>	43	
Eş Değerlendirme (<i>Peer Review</i>) <i>Richard Smith</i>	47	
İnteraktif Oturum I <i>Murat İnanç, Banu Çakır</i>	51	
İnteraktif Oturum II <i>Nurullah Akkoç, Önder Ergönül</i>	69	
Dizin	81	

—

—

—

—

ÖNSÖZ

Hasan Yazıcı

Oldukça gecikmeli olsa da 3-4 Mayıs 2008 “III. Bodrum Romatolojide Klinik Araştırmalar ve Nesnel Tıp Günleri”nin kitapçığını bitirebilmiş olmamızdan büyük memnurluk duyuyorum. Umarım 2009 toplantısının kitapçığı da yakında, önümüzdeki aylarda elinizde olacak. Sponsorumuz Schering-Plough artık MSD oldu. Onlara, yayımcımız Deomed’e, toplantıya ve kitaba katkıda bulunan herkese içten teşekkürler.

Düzenleme Kurulu adına
Hasan Yazıcı

RANDOMİZASYON

B. Çakır

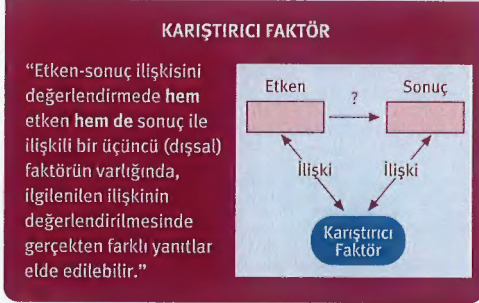
Karıştırıcı faktörler (*confounding factors*) analitik araştırmalarda verilerin yorumlanmasında sıkıntı yaratabilen, sistematik hata kaynaklarıdır (Şekil 1). Bir faktörün, verilen bir hipotezin doğruluğunu test ederken, etken ile sonuç arasındaki ilişki üzerinde karıştırıcı rol oynaması söz konusu olabilir. Bir faktörün karıştırıcı faktör olabilmesi, (1) çalışılan sonuçtan bağımsız olarak etken ile ve (2) çalışılan etkenden bağımsız olarak sonuç ile aynı anda ilişkili bulunması gerekir. Yani, sadece bir tanesiyle ilişkisi olması yeterli değildir (Şekil 2). Ek olarak, etkenin sonuca yol açtığı yolak üzerinde yer alan bir değişken de (örneğin, obezite-koroner arter hastalığı ilişkisi çalışılırken LDL-kolesterol), diğer iki özelliği taşısa bile, bir karıştırıcı faktör değildir. Bu değişkene, “ara değişken” adı verilir ve analizlerde karıştırıcı faktörden farklı olarak irdelenmelidir. Topladığımız veriler ile sonuç arasındaki ilişkiyi inceleyen, etken(ler) ve karıştırıcı faktör(ler) dışında, sonuçla ilişkili diğer risk faktörleri, ikincil ilişkiler, etkiyi değiştirici faktörler (*effect modifiers*) benzeri başka faktörler de analizlerde irdelenir, ancak, bunların hepsine birden karıştırıcı faktör muamelesi yapılması doğru değildir. Bir faktörün karıştırıcı faktör olarak tanımlanması önemlidir: Çünkü, çalışmada bir karıştırıcı faktör söz konusuysa, bunun kontrolü çok önemlidir. Mutlaka yapılmalıdır. Karıştırıcı faktör kontrolünün birkaç basamağı vardır (Şekil 3). Öncelikle böyle bir şeyin varlığından haberdarsak, bunu planlama aşamasında kontrol edebiliriz. “Randomizasyon” bu yollardan birini oluşturur. Bir grubu erkekler ve kadınlar şeklinde sıralayabiliriz. Ya da sadece erkekleri alabilir, yani “sınırlama” yapabiliriz. Olgu kontrollerde olduğu gibi eşleştirme yapabiliriz. Eğer bütün bunları planlama aşamasında yapamadıysak veya sonradan, yani analizde, bir faktörün karıştırıcı etkide olduğu ortaya çıkmış ise, çalışmada analizleri belirli bir gruba sınırlayabiliriz, tabakalandırma yapabiliriz ya da bir takım çok değişkenli yöntemlerle bu karıştırıcı faktörü kontrol edebiliriz.

Bu bölümde karıştırıcı faktörlerin kontrolü için çok değerli özellik taşıyan randomizasyon

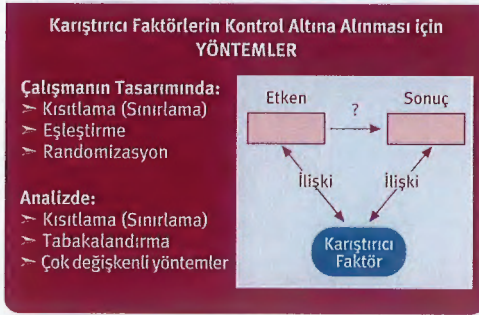
"... uzun şapkalar takip şemsiye taşımanın omuzları genişlettiği, yaşamı uzattığı, hastalıklara karşı bağışıklık ihsan ettiğini kanıtlamak kolaydır; çünkü, istatistik ilmi, bu nesneleri kullanan sınıftan insanların bunlara sahip olmayı rüyalarında bile görmemiş sınıftan olanlara kıyasla daha yapılı, daha sağlıklı olup, daha uzun yaşadıklarını gözler önüne sermektedir."

George Bernard Shaw: Doktorun Derdi (Doctor's Dilemma) için önsöz, 1906.

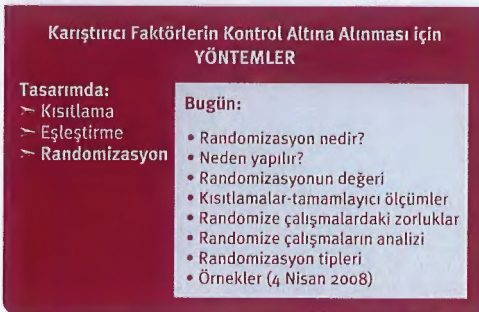
Şekil 1: İstatistik bilimi doğru kullanılmalıdır.



Şekil 2: Karıştırcı faktör nerede söz konusudur?



Şekil 3: Karıştırcı faktör nerede söz konusudur?



Şekil 4: Tasarımda randomizasyon. Konu akış planı.

konusu tartışılmaktadır. "Randomizasyon nedir, nasıl yapılır, neden değerlidir, ne tür kısıtlılıkları vardır, gruplar rastgele seçilmiş ise artık bu çalışmaya tam olarak ona güvenebilir miyiz, ne tür zorluklarla karşılaşabiliriz, bu çalışmalarını nasıl analiz etmeliyiz?" sorularının yanıtları ele alınmaktadır. Bu kapsamda, bazı özel randomizasyon yöntemleri, çok genel özellikleriyle, sunulmaktadır. Bölümün konu akış planı Şekil 4'te verilmektedir.

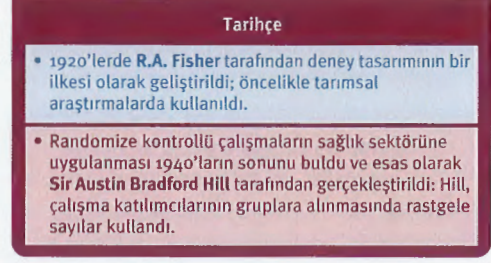
Randomizasyon, ilk olarak Fisher'in tarımda yaptığı bir takım araştırmalarda, tarımı geliştirmek için ortaya koyduğu, yeni ürün ortaya çıkarmak için kullandığı bir yöntemdir (Şekil 5). Daha sonra, ilk kez 1940'larda, insanların rastgele sayılar kullanılarak çalışmaya alındığını, ilk randomize kontrollü çalışmaların ön ayağı olarak da Wehtegill'in çalışmalarını görüyoruz.

Randomizasyon nedir sorusundan önce şu soruya açıklık getirmek gerekir: Bir randomize çalışmaya, çok rastlandığı şekilde, nasıl randomize edildiğini bilmiyorsak da güvenmeli miyiz? Burada ele alacağımız herşey, randomize kontrollü bir çalışmanın öncelikle yapılabilir olmasını, hasta açısından doğru ve etik olmasını ve uygulanabilir olmasını gerektirmektedir. Ondan sonra gerçekten randomizasyondan bahsedip, bunun yararlarından konuşabiliriz (Şekil 6). Bir grubun içindeki hastaları, iki farklı gruba incelemek istediğimizde, örneğin bir gruba bir müdahale için bir ilaç verip, diğer grubu da kontrol grubu olacak şekilde planladığımızda, bu iki insan grubunu nasıl ayırmalıyız ki, çalışmanın sonuçları doğru olsun? Burada hasta grubundaki herkesin tedavi grubuna ya da alternatif gruba girme şansının eşitliğinin biliniyor olması ve diğer katılımcı bireyden de etkilenmiyor olması gerekir. Örnek olarak, hekim yanınızdaki diğer birey sizden daha ağır hasta diye onu tedaviye alıp sizi plasebo grubunda bırakmamalıdır. Herkesin bilinen, tercihen eşit (ancak şart değil) ve diğer bireyden etkilenmeyen bir alınma şansı olmalıdır. Bu randomizasyonu farklı şekillerde oluşturabilirsiniz; yazı tura atabilirsiniz, kart çekebilirsiniz, rastgele sayılar tablosunu kullanabilirsiniz (Şekil 7). Ancak örnek büyüklüğünün çok küçük olduğu durumlarda, yazı tura attığınız za-

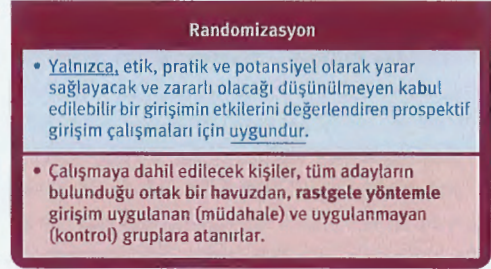
man 10 tanede beşi yazı, beşi tura olmayacaktır. Bin tane atabilseniz belki bu şansınız olabilir. O nedenle küçük örnekler şeklinde olan gruplarda iki grup birbirine eşit olmayabileceği ve bazen çok da farklı olabileceği için, kısıtlanmış (*restricted*) adı verilen randomizasyon yöntemleri kullanılabilir. Ama hepsinin mantığı aynıdır: Bir olasılık üzerinden işlem görürler ve kişilerin kendi karar ve hükümlerinden bağımsız olarak hastaların ya da çalışmaya dahil edilen bireylerin benzer olarak (ikiye) ayrılmasını sağlarlar.

Randomizasyon çok zaman alan ve emek harcanan bir konudur, ama önemli özelliği bilimsel olarak size bunun karşılığını veriyor olmasıdır. Onun için bu çaba göze almaya değer bulunmaktadır. Randomizasyon yapılmadan önce bireylerin tümüne bu araştırma için ulaşılmış, görüşülmüş, bunların kaçının çalışma için uygun olduğunun ve kimlerin bu çalışmaya gireceğinin belirlenmiş olması önemli bir noktadır. Randomizasyon daha sonra uygulanmalıdır. Bir diğer deyişle, tüm faktörler ve bireyler hazır olduktan sonra, bu bireyleri gruplara alabiliriz. Burada iki nokta çok önemlidir; birincisi seçimi random yapmak, yani rastgele olarak bu bireyleri doğru seçmek ve ikincisi, bunu bireyler gruplara atanıncaya kadar izlemek ve sabit tutmaktır (Şekil 8). Örneğin, yazı tura attığınızda fikrinizi değiştirmemek için o bireyin özelliklerinden etkilenmeden, yazı turanın gerçek sonucunu çalışmanızda göstermeniz gerekir. Çünkü orada (yani kişiyi atandığı gruba dahil etmekte) bir bozukluk olursa yazı tura atmanın (yani randomizasyonun) hiçbir anlamı kalmaz.

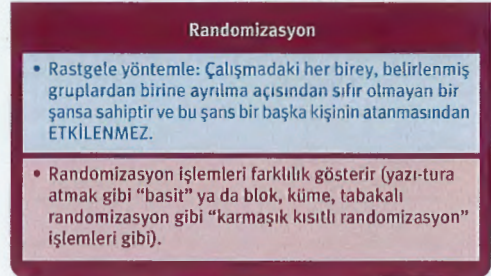
Daha önce de belirttiğim gibi, bir müdahale herhangi bir şey olabilir, bu mutlaka ilaç olmak zorunda değildir. Herhangi bir medikal yöntem, cerrahi yöntem olabilir, bir diyet, hatta bir davranış değişimi bile olabilir. Tüm bunlara dikkat etmemiz gerekiyor. İki noktayı, tekrar hatırlatmak istiyorum. Birincisi, katılımcı bireyleri rastgele gruplarken, bunun bizim dışımızda, tayin edemeyeceğimiz, tamamen rastgele, olasılık üzerine dayanan bir işlem olduğundan emin olmamız gerekir. İkincisi öyle bir seçim yapmalıyız ki, bireyler gruplara alınıncaya kadar verdiğimiz bu karar değişmemelidir.



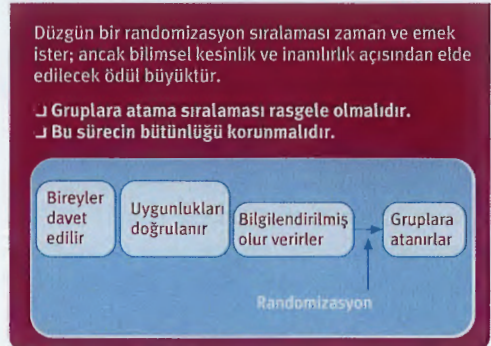
Şekil 5: Tarihte randomizasyon örnekleri.



Şekil 6: Randomizasyon nedir?



Şekil 7: Randomizasyon nasıl yapılır?



Şekil 8: Randomizasyonun aşamaları.

- Temel mantık: Randomizasyon, yaratılan iki grubun belli bir değişken açısından benzer (aynı değil) olmasını mümkün kılar.
- **Bu, olasılığa dayanan bir yöntemdir:** Ancak belli sayıda katılımcı (denek) varsa işe yarar. n küçükse, araştırmacı yanlılığını önleyebilir, ama gruplar arası farklılıkları her zaman azaltamayabilir.
- Tamamen tesadüfî bir farklılık kalabilir. Bu da analizde kontrol edilebilir!

Şekil 9: Randomizasyon ne sağlar?

“Allocation concealment” adı verilen bu durum, körleme yani “blinding” yöntemi ile çok karışır. Körleme dediğimiz şey, çalışma başladıktan sonra katılımcı bireylerin hangi ilacı almadıklarını bilmemeleridir. Ama “allocation concealment” basamağı bundan önceki bir döneme aittir. Yani, randomizasyonda üç önemli aşama vardır; önce bireyleri rastgele gruplara atarsınız; bu durumu bireyleri belirleyip çalışma başlayıncaya kadar gizli tutarsınız; katılımcı bireyler bu kurala göre belirlendikten sonra ise (gerek duyulursa) katılımcıları kör kılıp (blinded), kişinin kendisine uygulanan müdahalenin türünü bilip bilmemesinin sonucu etkilemesini engellemeye çalışırsınız.

Aslında, ancak bu üçünü bir arada yaptığımız zaman randomizasyonun yararlarından bahsetmiş oluyoruz; bunlardan bir tanesini ayırırsak, randomizasyonun getirisi önemli düzeyde azalacaktır. Örneğin bazı çalışmalarda “allocation” yani kişilerin gruplara atanması ile ilgili yeterli bilgi bulunmamaktadır. Bu durumda çalışmaya, yalnızca yazarlara inanma doğrultusunda bir güven söz konusu olmaktadır, ki bu her zaman doğru bir yaklaşım değildir.

Burada önemli bir ayrıntı unutulmamalıdır: Randomizasyon grupları benzer kılar, aynı kılmaz. Randomize kontrollü çalışmaların ilk tablosuna baktığımız zaman iki grup görürüz; placebo ve tedavi grubu. Tabii 3-4 grup da olabilir. Bunların benzerlikleri vardır, ortalama özellikleri vardır ve yanlarında da p değerleri vardır. Bunun amacı, gruplar arasındaki farklılıklarda bu farkın şans eseri mi değişkenlik gösterdiğini anlamaktır. P değerlerinin burada altı çizilmelidir. Örneğin, 5 yaşında ikiz iki erkek çocuk aldığımızda, 3-4 özelliğe bakarsak aynıdır, ama 24 özelliğe bakarsak, bir yerde mutlaka fark ortaya çıkacaktır. Onun için randomizasyon yaparken hangi özelliklerde benzerliğin bizim için önemli olduğunu bilip, karşılaştırmaları hangi özellikler açısından yapacağımıza baştan karar vermemiz gerekir. Aksi takdirde şu da mı aynı, bu da mı aynı dersek bir yerde mutlaka farklılık buluruz, ki bu yanlıştır. Ayrıca, bu özelliklerin tamamen aynı olması gerekmez, benzer olması gerekir. Eğer benzer değilse de, bunu bilmemiz gerekir (Şekil 9).

Sonuç olarak randomizasyon iki yarar sağlar: Bunlardan birincisi araştırmacının, istediği bireyi istediği gruba koymasını engellemesidir; bu seçime bağlı taraf tutmayı engellemektir ve randomizasyon bunu mutlak gerçekleştirir. Diğer ise grupları benzer kılmasıdır. Ama bunun bir olasılık olduğunu düşünürsek, bir şans eseri benzerlik olamama olasılığı bulunduğunu da unutmamamız gerekir.

H. Yazıcı: Buraya kadar randomizasyonun adeta salt prospektif ilaç veya girişimsel çalışmalar için hayati olduğu izlenimini aldık. Ama bence ilaç çalışmalarında büyük sorun, ağır hastaların, bir etki görülebilecek hastaların ilaç çalışmasına alınması, yani gerçek hayatı yansıtmamasıdır. Halbuki, örneğin eğer üç merkezde 900 hasta varsa ve halbuki onlardan siz kendi kuvvet hesabınızla 150’lerden 450 tane almak istiyorsanız, bunun yerine ilk ana grubu randomize ederek almak, gerçek hayatı çok daha iyi yansıtır kanısındayım. Onun için bu mantıkla giderek deney hayvanlarını da randomize etmek mümkün. Yani bu randomizasyonun o kısmına inmek lazım; bu birinci nokta. İkincisi retrospektif olarak da randomizasyon çok mümkün. Cerrahpaşa’da 6 bin tane Behçet dosyası var ise, biz Behçetlilerin ortalama kaç kadın, kaç erkek derken, 6 binini saymak yerine onlardan randomizasyon yapıp da sayıyı indirirsek o zaman yine o büyük küme hakkında fikir elde ederiz. Bu da randomizasyonun işe yarayan bir başka yönü sanırım. Kontrollü bir ilaç çalışması bir yana, herhangi bir gözlem yaparken de büyük bir gözlem hakkında küçük bir gözlem yapıp büyük gözlemin temsilcisi olması söylemek istediğim. Ama senin anlattığın, yine gözlem içinde eşit olarak iki tarafa gitmesi; öyle değil mi?

B. Çakır: Hocam çok önemli bir noktaya değindi. Aslında belki bunu baştan vurgulamam gerekirdi. Randomizasyon iki farklı alanda kullanılabiliyor. Bir grubun hepsine ulaşamadığımızda seçeceğimiz grubun tümü temsil etmesi için kullandığımız “rastgele örnekleme (*random sampling*)” yöntemi var, bir de deneysel araştırmalarda karşılaştırma gruplarına seçeceğimiz kişilerin benzer olması ve seçime bağlı

taraf tutmanın engellenmesi için kullanılan “randomizasyon (*randomization*)” var. İkisi de benzer temele dayanır. Hocamın dediği gibi, ikisinde de birey faktörü önemlidir. Yani iyi niyetli olarak ya da kötü niyetli olarak yaptığımız çalışmayı değiştirebiliriz. Hiç bilmeden, farkına varmadan da değiştirebiliriz. Bu tür durumlarda birey faktörünü ortadan kaldırmak için retrospektif çalışmalarda rastgele örnekleme kullanıyoruz. Dediğiniz gibi 6 bin dosyaya ulaşp, bu dosyaların hepsinden veri almamıza gerek yoktur. Ama bunlardan örneğin, son 100 tanesini seçersek, değişen zaman diliminde sistem değişmiş, hasta özellikleri, hastaya yaklaşım, markerler değişmiş olabileceğinden başka bir grup insan profili ile karşılaşabiliriz; ilk 100’ü seçersek de bir başka profile karşılaşabiliriz. Başka bir açıdan bakarsak; örneğin Cerrahpaşa’da, tanı için bir yöntem Hacettepe’de başka bir yöntem uygulanıyor olabilir. Dolayısıyla, birey faktörünü yok etmek ve gruptaki bireyler arasından örneğe çıkacak olanları rastgele seçebilmek için random bir örnekleme yöntemi kullanıyoruz. Diyelim 6 bin kişi var ve bunlardan belli bir yüzde, örneğin %1 seçeceğiz. Bu %1’i, basit rastgele sayılar tablosundan bulup, rastgele 60 kişi numarası seçiyoruz ve bu numaraları dosyaları üzerinde çalışıyoruz. Bu rastgele örneklemedir. Bu uygulamayı, haklısınız, retrospektif de, prospektif de yapabiliriz.

Randomizasyonun bir özelliği ise, klinik çalışmaların doğasından gelen bir özellik olarak, sadece rastgele seçim yapma değil, olgu kontrol çalışmalarında ya da kohort çalışmalardan farklı olarak, katılımcı bireyin etkenle karşılaşma durumuna araştırmacının bizzat kendisinin karar vermesidir. Randomize kontrollü çalışmalarda ya da klinik çalışmalarda, bireylerin seçiminin kararı araştırmacıya aittir. Kişiye, “sen bunu yapacaksın ya da sen bu ilacı içeceksin” denir. Seçenek bırakılmaz. Bunun avantajı katılımcı bireyin seçimi değiştirebilecek tarzda taraf tutmasını ortadan kaldırmasıdır. Birey, örneğin ikisi de metabolizmayı hızlandıran ‘ya her gün egzersiz yapacaksın ya da yemeğin üstüne bir tabak daha yiyeceksin’ şeklinde iki seçeneğe sahip olursa ikincisini seçebilir. Bu da yol açacağı farklı etkileşimler-

le araştırmacıyı yanlış sonuçlara yönlendirebilir. Araştırmacı, seçme kararını hastaya bırakmadan bizzat verdiği için randomizasyon burada artık seçilmiş örneklem üzerinden, örneğin “ilacı kime vereceğiz, plaseboyu kime vereceğiz,” “yeni ilacı kime vereceğiz, eski ilacı kime vereceğiz” ya da “bu ilacı kime vereceğiz, ilaç artı diğer yöntemi kime vereceğiz” gibi sorularla devam edecektir.

Özetle, randomizasyonda mutlaka bir müdahale söz konusudur. Bu müdahale (ki bu, etken alt gruplarından biri ile karşılaşma durumudur; örneğin, iki ilaçtan hangisinin alınacağına karar verme) araştırmacı tarafından belirlenir. Bunun yanı sıra, bu müdahalenin hangi grup hastalarda, örneğin sadece ağır hastalarda mı, sadece genç hastalarda mı, belirli bir özelliği olan bir grupta mı yapılacağına da araştırmacı karar verebilir: Bu durum randomizasyon değil, çalışma grubunun belirli bir grup ile sınırlanması yöntemidir, ancak benzer şekilde karıştırıcı faktör kontrolüne yardım eder.

Burada geçerlilik kavramını da vurgulamakta yarar vardır. Epidemiyolojide iki tane geçerlilik söz konusudur: İç geçerlilik ve dış geçerlilik. Dış geçerlilik genellenebilirliktir. Seçilen hasta popülasyonu herkese benzemediğinden, randomize kontrollü çalışmaların dış geçerliliği (*external validity*), yani genellenebilirliği (*generalizability*) düşüktür. Hatta kimi zaman o kadar kısıtlı kriterler kullanılmış olabilir ki, sonuçlar ilgili hasta popülasyonuna bile tam olarak genellenemeyebilir. Bununla birlikte deneysel çalışmalar iç geçerliliği (*internal validity*) çok yüksek çalışmalardır. Örnek vermek gerekirse, iç geçerlilik açısından, bir anlamda çok ince ve titiz bir nakışa benzetilebilir. Buna karşılık, herkes bu nakışı yapabilir mi sorusu da dış geçerliliğin zayıflığını ortaya koyabilir. Farklı bir konu da olsa, randomize çalışmalarda tedaviye niyet (*intention-to-treat*) analizinin, en azından bir grup içerisindeki genellenebilirliği artırmaya yönelik bir yöntem olduğunu burada vurgulamakta yarar vardır. Genellenebilirlik, örnekten aldığımız sonuçların benzer topluluklar için çıkarım yapmak için kullandığımız bir yöntemdir.

Randomizasyon, çalışma kör olsun ya da olmasın, kimin hangi müdahaleyi alacağı ya da kime neyin uygulanacağı kararının rastgele verilmesi yoluyla, gruplara kimlerin dahil edileceğine birileri tarafından karar verilememesini ve sonuçlar hakkındaki olası öntahminlerimizin araştırma yan tutmaya neden olmasının önlenmesini sağlar.

Randomizasyonun bir olasılık olduğunu tekrar vurgulamakta yarar vardır. Olasılıkta sa-yılar çok büyük olduğunda rastgelelik beklentisi kesinleşir: Gruplar potansiyel karıştırıcı faktörler, hatta bilinmeyen/tahmin edilemeyen faktörler ve seçme yanlılığı açısından benzer kılınır. Buna karşılık küçük örnek büyüklüklerinde grupların benzerlik olasılığı düşer. Randomize klinik çalışmaların raporlanmasında kullanılan ilk dizi tanımlayıcı tablolar bu nedenle verilmektedir.

Randomizasyonun Kısıtlılıkları

Randomizasyonda çeşitli kısıtlılık ve güçlükler söz konusudur (Şekil 10-12). Araştırmacıların aslında rastgele olmayan çalışmalara kasıtlı olmaksızın rastgele süsü verdiği durumlar olabilir. Hasta başvuru numarası üzerinden, hastaneye başvurduğu tarih üzerinden, yazı tura atarak ya da her iki günde birini alarak, küçük örnekler üzerinde randomizasyon sağlanamamaktadır. Örneğin, pazartesi gelen hastalarla, çarşamba günü gelen hastalar belli farklı özelliklere sahip olabilir. Bunlara deterministik yaklaşım diyoruz. Yapılan her şey her zaman rastgelelik sağlamayabilmekte ya da rastgeleliği insanlar bozabilmektedir. Bu gibi durumlara dikkat edilmediğinde rastgeleliği bozma riski yaratılmaktadır.

Bilimsel tasarımın ideal olduğu durumlarda dahi pratik uygulamada bu bütünlüğü sağlayamamak söz konusu olabilir. Organizasyon, zaman kısıtlılığı, yüksek maliyet ve kısıtlı kaynaklar çalışmaların planlandığı şekilde uygulanmasını engelleyebilir ve randomizasyon yapılsa dahi sonuçların epidemiyolojik açıdan kalitesini azaltabilir.

Randomizasyon yapılan çalışmalarda etik onam alınması, çalışmaya katılan kişilerin çalışmanın amaçları, kullanılacak yöntemler, olası riskler vb. konularda tam olarak bilgilendirilmeleri sonrası çalışmaya alınmaları ve eğer herhangi bir aşamda çalışmadan çıkmak isterse buna imkan tanınması esas olmalıdır.

Özetle, randomizasyon sadece çalışma başlangıcı itibarı ile grupların benzer kılınmasını sağlamaktadır ama uygulamalardaki hata ve eksikler, grupların izlem süresinde benzer kalmalarını (sayıca veya özellikler açısından) engelliyor ise randomizasyonun getirisi düşüktür (Şekil 12).

Öte yandan her klinik araştırmanın mutlaka randomize olması gerekli değildir; her çalışmanın, kendine özgü değeri vardır. Bu durumda çalışma non-randomize denebilir ve neden bu şekilde düşünüldüğü, çalışmanın kısıtlılıklarının neler olduğu çalışma bildirilirken belirtilir. Her zaman kısıtlılığı bildiride açık olarak belirtmek, onu başka bir kılığa döndürmekten daha yararlıdır.

Randomize Çalışmaların Analizi

Randomize kontrollü çalışmaların nasıl yapılması, okunması ve değerlendirilmesi gerektiği konusunda bir liste vardır: *CONsolidated Standards of Reporting Trials* (CONSORT). Burada adım adım neler yapılması gerektiği belirtilir (bkz. Interaktif Oturum II). Veriler toplandıktan sonra analizde nelere dikkat edilmesi gerektiği standarttır. Temel noktalara dikkat çekmek isterim. Tedavi veya grupların izlemi sürecinde gruplardan kayıp olabilir. Bu kayıpların sayısının çok olması ya da özel bir durumun kayıba sebebiyet vermesi verilerin analizinde hataya neden olabilir. Benzer şekilde gruplardaki bireylerin yapılan müdahaleye ek aldığı ilaçlar, tedaviler, davranışlar yanıt etkileyebilecektir. Analiz aşamasında tüm bu olası taraf tutma riskleri belirlenmeli, bu özellikler açısından analizlerde kontrol sağlanmalıdır. Çok değişkenli analiz yöntemleri, bu tür olası karıştırıcı faktörler için eş zamanlı kontrol sağlayacaktır (Şekil 13)

Randomizasyonun Kısıtlılıkları -I-

- ❑ Dikkat: Araştırmalarda, kimi zaman, kasıtlı olmaksızın, randomizasyon yapılmış izlenimi olsa dahi randomizasyonun sağlanamadığı durumlar olabilir:
 - Olgu numarasına göre atama
 - Başvuru tarihine göre atama
 - Dönüşümlü atamalar
- ❑ Olguların başvurusunda ya da kayıda alınmasında bir içsel düzen varsa ve randomizasyon yeterli bir atama gizliliğini önleyorsa, atamalarda rastgelelik sağlanamamış olabilir.
- ❑ Bu durumda, en iyisi, raporda durumu "randomize değil" şeklinde bildirmek ve karıştırıcı faktörlerin analizlerde nasıl kontrol altına alındığını belirtmektir.

Şekil 10: Randomizasyonun kısıtlılıkları I.

Randomizasyonun Kısıtlılıkları -II-

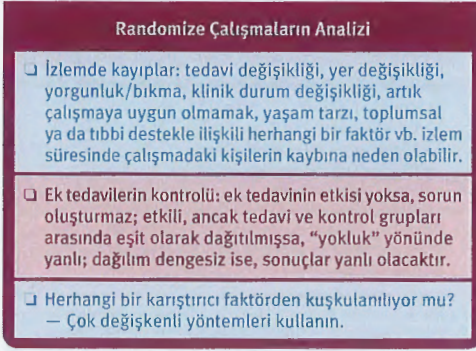
- ❑ Karıştırıcı faktörlerin rolü, etken ve sonuç ile olan ilişkisinin gücüne bağlıdır.
- ❑ İstatistiksel anlamlılığın yokluğu tek başına yeterli değildir. (CONSORT)
- ❑ Örneğin, kanserin evresi: blok randomizasyon ya da önceden tabakalı randomizasyon.

Şekil 11: Randomizasyonun kısıtlılıkları II.

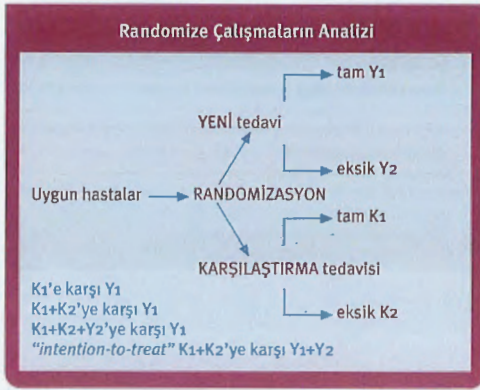
Randomizasyonda Güçlükler

- ❑ İdeal bilimsel tasarım ile pratik sorunlar arasındaki denge
- ❑ Lojistik — organizasyon, zaman, maliyet, kaynaklar
- ❑ Etik sorunlar — Bilgilendirilmiş olur
- ❑ Hem profesyonellerin hem de hastaların kurallara yüksek bağlılık göstermesi ve işbirliği; sayı yeterli mi? (girişte, randomizasyondan sonra ve izlemde kayıplar?)- randomizasyon sonrası olaylar randomizasyonun getirdiği yararı azaltabilir.

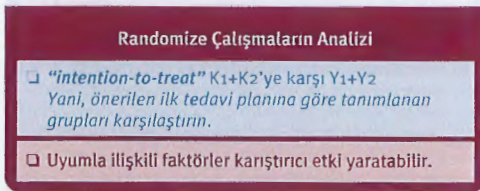
Şekil 12: Randomizasyonda güçlükler.



Şekil 13: Randomizasyon ne sağlar?



Şekil 14: Randomize çalışmaların analizi II.



Şekil 15: "Intention-to-treat" analizi.

Grupların izlem sırasında sayıca ve özellikler açısından farklılıklar göstermesi durumunda randomizasyonun etkisi azalacaktır. Öte yandan, örneğin yeni bir ilaç piyasaya sunulduğunda bu ilacın reçete edildiği kişilerin tümünün ilacı benzer şekilde kullanmayacakları, farklı beslenme, fiziksel aktivite, ilaç kullanımı vb. tarzları olabileceği de beklenen bir durumdur. Bu nedenle, randomize çalışmalarda "intention-to-treat" analiz yönteminin kullanılması giderek yaygınlaşmaktadır (Şekil 14 ve 15).

CONSORT listesi incelendiğinde iki farklı çalışma vardır (Şekil 16 ve 17). Biri açıklamak, durumu en iyi anlatmak ve iç geçerliliği en iyi şekilde göstermek için yapılan açıklayıcı çalışmalardır (*explanatory trials*). Örneğin "SLE'nin alevlenme dönemleri ile bir marker arasındaki ilişkiyi ortaya koymak istiyorum; benim görmek istediğim bu. Seçilmiş bir SLE hasta grubunda hastalığın gidişi, aktivite düzeyi için ben bu markerı kullanabilir miyim?" düşünceyle bir çalışma yapılabilir, genellenebilirliği düşük olabilir, ama markerin kullanımının etkili olarak ne kapsamda olabileceğini yansıtabilir. Bu durumda çalışma açıklayıcı tipte bir çalışmadır.

İkinci grupta ise pragmatik veya pratik çalışmalar (*pragmatic or practical trials*) yer alır. Bunlar da aslında yine klinik çalışmalardır ama burada özellikle *intention-to-treat* yaklaşımı daha önemlidir. Örneğin ilgili ilaç ruhsatlanıp reçete edilmeye başlandığı dönemde hastalara verildiğinde hasta ilacı kullanmayabilir, zamanında ve uygun dozda almayabilir, ek olarak steroid kullanabilir, ikinci bir doktora danışıp ilacını değiştirebilir. Dolayısıyla, sorumuz bu kez "ilaç ne kadar etkili?" değil, "bu ilaç hastalara reçete edildiğinde etkileri ne olacak, yani pratik uygulama sonuçları neler getirecek?" olabilir; bu durum için uygun araştırma yöntemi pragmatik çalışmalardır (Şekil 16 ve 17).

Randomizasyon Yöntemleri

Üç tane ana randomizasyon vardır. Bunlardan bir tanesi basit adı verilen yani kısıtlılığı olmayan; bir diğeri kısıtlılık yapan yani sınırlı

(restricted) randomizasyon, üçüncüsü de tabakalı (stratified) randomizasyondur.

Basit dediğimiz randomizasyonun sadece adı basittir (Şekil 18). Aslında hiç basit olarak algılanmaması gerekir, *allocation concealment* yani onu gruplara ayırma doğru yapılmalıdır. Yazı tura aslında son derece doğrudur; olasılık teorisine dayanan bir yöntemdir fakat koşul onu bozmamak olmalıdır. Bu nedenle 200'den fazla birey sayınız varsa basit randomizasyon yöntemlerini kullanmanız önerilmektedir. Bu yöntemler yazı tura atmak, zar atmak, basit rastgele sayılar tablosundan yararlanmak vb olabilir. Ancak editörler ya da okurların genellikle grupların bilgisayarda yaratılmış olmasını ya da rastgele sayılardan yararlanılmasını tercih ettikleri unutulmamalıdır. Bu tercihin bir nedeni de araştırmacının işin içine karışmadığından emin olmaktır. Ama 200'ün altındaki gruplarda bu teknikte sık olarak sorunlar yaşanabilmektedir.

H. Yazıcı: En fazla 200 mü, toplam 200 mü?

B. Çakır: Toplam 200, ama burada çalıştığınız olayın yaygınlığı da önemlidir. 200 derse-
niz, 100-100, %1'den daha sık bir şey beklediğiniz takdirde kendini gösterebilecektir. Yani örneğin bir ilaçla ilgili bir advers etki çalışıyorsanız onun sıklığı da önemlidir. Ama kabaca ne olsun dersek, biz 200 sınırını koyuyoruz.

Peki, kısıtlı randomizasyon niye var? Küçük ya da özelliği olan gruplarda iki grubun yakın olmasını istiyorsanız, bir grupta 15 diğerinde 60 kişi varsa ve bu sizin istemediğiniz bir şey ise ve bunu "olmadı, tekrar yazı turayla 15'i artıralım" şeklinde yapmak istemiyorsanız, baştan sınırlı randomizasyon tekniklerini kullanabilirsiniz. Bunlardan ilki restricted randomization adı verilen, daha eski bir yöntemdir. Şimdilerde çok sık kullanılmamaktadır. Çalışma başlamadan önce grupları objektif bir şekilde eşleştirmek için kullanılan bir tekniktir.

Random allocation da yine bu gruptandır. Büyük bir blok düşünün, sistem şöyle bir var sayım yapıyor: "Ben 200 tedavi hastası bulacağım, 200 tane de plasebo grubuna yerleştireceğim, ama sonuç itibarıyla grupların her iki-

- Açıklayıcı çalışmalar (Explanatory trials)
"Ulaşılan pratik sorunlardan bağımsız olarak, girişimin ideal koşullarda, çalışmayı kabul eden bireyler üzerindeki etki nedir?" (etkililik-efficacy)
- Yönetim çalışmaları (Management trials)
İzlemdeki kayıplar nedeniyle etkiler dilüsyona uğrayabilir.
"Pratik bir durumda, belli bir tedaviyi reçetelemenin ya da belli bir girişimi önermenin gerçek ortamdaki etkisi nedir?"

Şekil 16: Müdahale tipi araştırmaların amaca göre sınıflandırılması I.

- Açıklayıcı çalışmalar
Pratik değerden çok potansiyel değeri ortaya koyar; Belirlenen etkiler sadece çok özelliği olan (toplumu yansıtmayan) bir grup şeklinde birey için geçerli olabilir.
- Yönetim çalışmaları
PRAGMATİK ya da PRATİK klinik çalışmalarda elde edilen sonuçlar, klinik karar vermede ya da sağlık politikaları belirlemede kullanılabilir.

Şekil 17: Müdahale tipi araştırmaların amaca göre sınıflandırılması II.

BASİT: temel nitelikte, ancak, her bir atamanın önceden tamamen tahmin edilemez olması kuralını koruyan

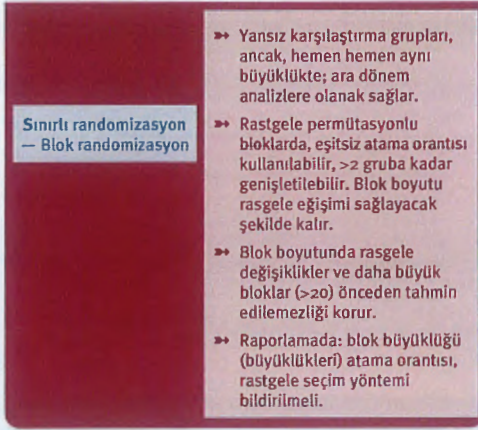
- Basit (kısıtlı olmayan) randomizasyon
- Sınırlı randomizasyon
 - Yer değiştirmeli randomizasyon
 - Blok randomizasyon
 - Rastgele atama kuralı
 - Yanlı yazı-tura ve torbadan çekmeye dayalı
- Tabakalı randomizasyon

► N. ise, tesadüfi olarak gruplardaki kişi sayısı farklı olabilir. N. durumunda daha iyidir (iki kollu çalışma için >200 yeterlidir)

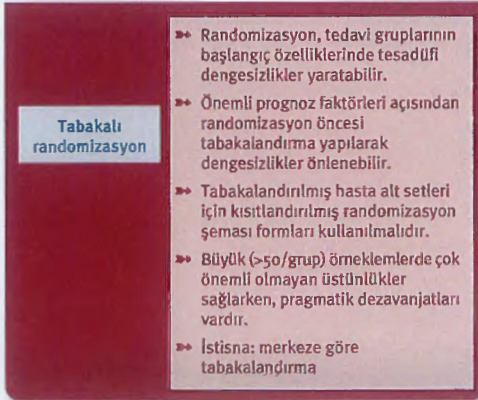
► Yazı-tura atmak, zar atmak, karıştırılmış kart destesi;

► Rastgele sayılar tablosu, rastgele sayı üretici bilgisayar programı

Şekil 18: Randomizasyon ne sağlar?



Şekil 19: Blok randomizasyon.



Şekil 20: Tabakalı randomizasyon.

sinde de yüzer kişi (eşit) olmasını istiyorum". Bu durumda, kohortun 5. yılında insan almaya devam ederken ara araştırma yapamayabilirsiniz çünkü burada sonuç itibarıyla gruplar eşitlenecek denilmişti; bu önemlidir.

Ancak, izlem süresi devam ederken arada da çalışma yapmak isteyebilirsiniz. Örneğin, bir artrit vakasında ilaç deniyorsunuz ve 6. ayını görmek istiyorsunuz. Ayrıca "birinci, ikinci, üçüncü yıllarını görmek istiyorum" diyorsanız, o zaman bu ara dönemlerde de plasebo ve tedavi gruplarının yakın olmasını istediğiniz için, bu durumda blok randomization dediğimiz gruplamayı yapabilirsiniz (Şekil 19). Küçük bloklar seçebilirsiniz; 10'luk, 20'lik, 50'lik. Her blokun yarısı tedavi, yarısı plasebo olabilir. Bu grup içi oranlar bire bir olabilir, bire iki olabilir. Bunu kendiniz belirleyebilirsiniz. Önemli bir nokta, blok kullandığınız zaman kişilerin bir aşamadan sonra, bloklamanın örneğin her 10'da bir değiştiğini fark edebilmesidir. Eşitlemenin farkedilmesi nedeniyle araştırmacı yine bazı şeylerle, örneğin verdiği ilaçlarla oynayabilir. Bu nedenle blokları değiştirebiliriz, yani önce 10'luk blok, sonra 20'lik blok, sonra yine 10'luk blok gibi.

Sonucu ana randomizasyon yöntemlerinden bir diğeri ise tabakalı randomizasyon (*stratified randomization*) adı verilen yöntemdir (Şekil 20). Tabakalı randomizasyon, prognoz faktörlerin farklı olduğu düşünülen rahatsızlıklar söz konusu olduğunda, yani başlangıç itibarıyla tedavi ve plasebo gruplarının birbirlerinden farklı prognozistik özellikleri varsa uygulanır.

Örneğin, Interaktif Oturum II bölümünde verilen çalışma örneğinde olduğu gibi, istatistiksel olarak plasebo grubuyla tedavi grubunda başlangıç SLE aktivitelerine bakıldığında ortalamaları yakın görünse de olasılıkla bir grupta daha çok tekrarlayan alevlenmelerin olabileceği varsayılan durumda tabakalı analiz yapılabilir. Yani, alevlenme riski yüksek olan grupta tedavi ve plasebo, alçak olan grupta tedavi ile plasebo. Buna baştan karar verilebilir, önce alt grup analizi yapmak mümkündür.

Baştan sadece tedavi ile plaseboyu karşılaştırmak üzere yola çıkılıp, ona uygun örnek büyüklüğü seçildikten sonra, alt grup analizi yapmak istenirse sayılar alt gruplarda kimse kalmayacak şekilde eriyebilir. Bu durumda da “anlamlı sonuç bulunamadı” sonucuna varılabilir (tip II hata) ancak bu yorum yanlış olacaktır: Çalışmaya başlarken asıl amaç biliniyorsa tabakalı randomizasyon tekniği kullanılması yarar sağlayacaktır.

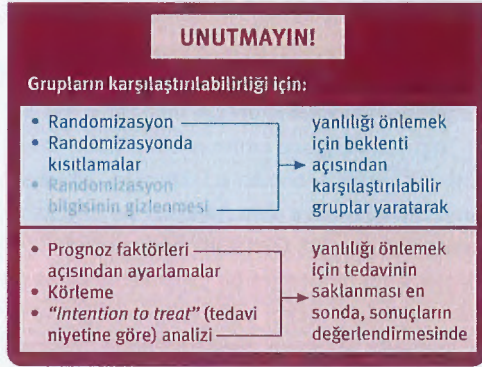
Grupların Karşılaştırılabilirliği

Grupların benzer kılınması istendiğinde, karıştırıcı faktörlerin önemli olduğu unutulmamalıdır. Çok değişkenli analizlerde, sayı yeterli olmak koşuluyla istendiği kadar karıştırıcı faktör için analiz yapılabilir. Fakat bir karıştırıcı faktörün olası karıştırıcı faktör olduğu bilinmiyorsa analiz yapılamaz. Bununla birlikte randomizasyonda grup ikiye bölündüğünde aslında her iki grup bilinmeyen özellikleri açısından da benzer kılınmış olur. Örneğin, bugüne kadar hiç tanımlanmamış bir romatizmal hastalık geni, tanımlanmamış bir davranış sistemi vb açısından da gruplar ikiye bölündüğü için, randomizasyon tüm diğer şeylerden önemlidir ve mutlaka yapılması gereklidir. Tabii burada grupların sonuç itibarıyla da bu randomizasyonu aldığına dikkat etmelidir.

Sonraki aşamada bu işin yine kör gitmesi gerekir (Şekil 21). Bu yöntemin adı körlemedir (*blinding*).

Sonuçlar değerlendirilirken yapılan *intention-to-treat* analizinde, birey baştan hangi gruba kondu ise aynı grupta devam etmiş gibi değerlendirmeye alınır. Örneğin bir katılımcı birey başlangıçta tedavi grubuna konmuş olmasına karşın bu bireyin tedaviden bir ay sonra bu gruptan çıkıp geri kalan 11 ayını plasebo grubunda geçirdiği bilindiği halde, tedavi grubunda imiş gibi değerlendirilmektedir. Çünkü bu durum normal hayatta ve doğal gidişte de benzerdir.

Şekil 21’de özetlenen altı yöntem birlikte ele alınmadığı zaman randomizasyon tek başına etkili olmamaktadır. O nedenle tek bir biri-



Şekil 21: Randomizasyon ne sağlar?

mi almak yerine, hepsini birlikte değerlendirmek suretiyle yapılan uygulamanın doğru olduğunu bilmek zorundayız.

Randomizasyon klinik çalışmaların iç geçerliliği olabildiğince artırmaktadır. Bununla birlikte çalışma sonuçları açıklanırken çalışmanın kısıtlılıklarını ve olası hataları okuyucu ile paylaşmak gerekir. CONSORT bildirgesinde (Başlık 8-10) randomize bir çalışmanın sunulmasında nelere dikkat edilmesi gerektiği sıralanmaktadır ve bunlar araştırmacıların çalışmalarını karşılaştırılabilir ve standart hale getirmesi için değerlidir (bkz. Interaktif Oturum II).

Her araştırmada raporun bulgular ve tartışma kısmında çalışma grubunun detaylı olarak tanımlanması sonuçların kimler için genellenebileceği açısından önemlidir. Son olarak her ne kadar randomizasyon grupların belirlenmesi aşamasında sistematik taraf tutmayı ortadan kaldırır ise de, prognoz faktörler açısından mükemmel şekilde dengelenmiş gruplar üretmesi şart değildir.

H. Yazıcı: Ben yine bir şey sorabilir miyim? Tabakalı örneklerdeki alt grup analizi olayını bir kere daha açıklar mısınız?

B. Çakır: Örnek üzerinden konuşursak; iki tane grup var, bunlar SLE hastası. Bir grup o olayın gelişme olasılığı daha fazla olan yani prognozu daha kötü olan A grubu. Diğer grup da, prognozu daha iyi olan B grubu. Bu iki grubu birlikte aldığımız zaman belirtilen prognoz özellikleri yüzünden, tedavi ve plasebo grubu arasındaki ilişki yanlış görülebilir. Onun için grubu ikiye bölüp bakmamız gerekir. Prognozu kötü olan grupta bu yöntem iyi ama prognozu iyi olan grupta değil diyebilirsiniz. Bu nedenle bir tabakalama yapmamız gerekir ancak bunun ötesinde, tabakalanmayı yapabilmek istiyorsanız, örneği belirlerken de ilgili prognostik faktörler açısından her iki grupta kötü ve iyi olan gruptan yeter sayıda insan seçmek gerekliliği vardır.

H. Yazıcı: Benim söylemek istediğim şu; tabii çok doğru, ama bu hiçbir zaman daha sonra kendini çürütmek yönünde subgrup analizi kapısını bize kapatmasın. Eğer dürüstçe

kendimizi çürütme yönünde yapıyorsak o subgrup analizinde sorun yok. Bana, kanıtlamak yönünde yanlış ama çürütmek yönünde olabilir gibi geliyor. Öyle değil mi?

B. Çakır: Kesinlikle, hocam çok doğru söylüyor. Diğer bölümde belirtilecek zaten, çok değişkenli analize geçmeden önce mutlaka tabakalandırma yapmamız gerekir. Orada karıştırıcı faktör var mı, varsa bunun ilişki üzerindeki etkisi nedir diye bakıyoruz. Yani, ilişki gizlenecek, az görünecek diye alt grup analizleri yapmaktan vazgeçmiyoruz. Ancak, örnek büyüklüğü yeterli olmadığında var olan ilişkiler de gizlenebileceği için, alt grup analizleri yapma planımız varsa, çalışma başlangıcında seçeceğimiz örnek hacmini o kadar çok tutmak gereklidir. Bu kapsamda örnek büyüklüğü nasıl hesaplanır ve yeterli sayı nedir konularını hiç tartışmadık.

Her iki (ya da daha fazla) müdahale kolundan birinde en az kaç kişi olmalıdır? NNT (*number needed to treat*) denilen kavramın hesaplamasında daha önceki çalışmalardan ya da tahminlerden elde edilen sayılar ile yola çıkıyoruz. Gruplar arasında en az x kadar fark saptamak için her kolda en az kaç kişi olmalı bilgisine varıyoruz. Ama yine grupları karşılaştırmak için, alt grup analizi gerektiğinde bu sayı da az olacaktır.

Soru: 90 kişilik lupus hastamız var, ikiye böldük. Yaklaşık 10 ayrı parametreden tabakalama yapıldığı söylendi. Bu kadar küçük grupta, diyelim yaş, cinsiyet, kreatinin vs ile ilgili randomizasyon yapılırken grupların eşit olması nasıl sağlanıyor? Yani sayısal olarak sonuçta 45-45 bölünmüyor.

B. Çakır: Şu nedenle yapmış olabilirler. Belki gruba özel değer hesaplayabilmek için değil de grubu heterojen (ana grubu olabildiğince yansıtır) kılmak için yapmış olabilirler. Yani her cinsten insan bu grupta yer alsın diye. Ama size ortalama bir değer verilir. Şunu düşünün, Türkiye de bir sigara araştırması yapacaksınız. Türkiye’de sigara araştırması yapmakla, bölgeye özel sıklık hesaplanması araştırması farklı örnek hacimleri gerektirir ama yine de her kesimden insan almaya çalışıyoruz:

Türkiye'nin her bölgesinden insan seçiyoruz ki Türkiye insanını yansıtabilsin. Ama sonuçta ülkeye özel bir değer verebiliyoruz. Bölgeye özel hesap için ise, gerekli minimum örnek sayısını bölge sayısı kadar artırmak gerekecektir. Verdiğiniz örnekte de, aynı örneklemede farklı gruplardan yeterli sayıda insanın eşit şekilde plaseboyla tedavi almasını garantilemek için böyle bir şey yapmış olabilirler. Öte yandan, "40-49 yaşında erkek olup son 2 ayda stabil giden SLE hastasında şöyledir" demeyip, ortalama bir sayı vereceklerdir, çünkü örnek hacimleri aksi durum için çok küçüktür.

M. İnanç: Randomizasyon deyince ben şöyle bir şey anlıyorum: Hocamın dediği doğru; bir kere baştan çok iyi planlama yapılması gerekiyor, ilk anladığım bu. Şöyle bir şey söyleyeyim size, çok önemli sıkıntılarla araştırmacıların bulunduğu bir ortamda bir protokol yapılıyor, hastalar toplanıyor, sonra geriye dönüp düşünüyorlar, "yeni bir başlangıç semptomunu alsa mıydık, almasa mıydık?". İnsan faktörü çok önemli ve sonunda yapılan şeylere dönüp baktığımız zaman, insanların bir şeyleri yapıp bitirdiklerini sonra başka bir sonuç almak istediklerini görüyoruz. Mesela SERENA'da SLE ile çalışan birçok grup var. Şöyle bir örnek gördüm: "Proteinüri nasıl azalmış? Hipertansiyon olanlarda ACE inhibitörü alınca azalmış". Bu dergilerde kutular var ya, oraya yazıyorlar ki, eğer SLE hastaları ACE inhibitörü alırlarsa, proteinürinin ortaya çıkma olasılığı azalır. Böylece bütün SLE hastalarına ACE inhibitörü önerilebilir. Halbuki yazıya dönüp bakıyorsunuz, hipertansiyonu olan grupta proteinüri çıkmasını azaltmış. Yani arada kulaktan kulağa gibi birçok olay geliyor. Burada benim anladığım, kuralları çok iyi koyup sonra ne zaman artık o kuralların değişmeyeceği gerektiğini iyi belirlemek lazım, yani işte hastanın olur verildiği andan itibaren belki bunlara dikkat edilirse daha az sonuç çıkar ama daha kolay anlaşılır diye düşünüyorum.

Soru: Körleme uygulanan araştırmalarda yine de randomizasyon gerekli mi? Günaşırı uygulama yanlış mı?

B. Çakır: Çalışmanın kör (*blinded*) olmasının etkisi randomizasyonun etkisinden farklı

bir etki. Kişinin plasebo etkisinin kör olması "ben iyi bir ilaç alıyorum, atlamayayım, bu gruptan çıkmayayım, alternatif bir tedaviye gerek yok gibi" diye düşünmesini sağlıyor. Hastanın ya da araştırmacının kör olması yanıtın değerlendirilmesi açısından iyi bir şey. Randomizasyon ise grupları sonucu etkileyebilecek faktörler açısından başlangıçta benzer kılıyor.

Yanlış yönlendirmek istemem ama çoğu durumda günaşırı uygulaması (Pazartesi, Çarşamba, Cuma seçmek de) yanlış değil. Sadece şu yanlış: O seçtiğiniz yerde bir içsel düzen dahilinde bu insanlar birbirlerinden farklıysa sonuçlarda fark ortaya çıkabilir. Gün aşırı gelen hastalar birbirine çalışma sonuçları açısından benzer ise problem değildir.

Soru: Randomizasyonun doğru yapılıp yapılmadığını ve etkilerini inceleyen çalışmalar var mı?

B. Çakır: Tabii, bu konuda çok çalışma var. Randomizasyon yapılmasına karşın aslında randomize olmayan çalışmalar ya da bozuk olduğunu fark ettiğimiz, durdurduğumuz, tekrar başlattığımız çalışmalar var. Ancak randomizasyonu tek başına düşünmemek lazım. Yazı tura atmak zararlı değil ama tura çıktığında, araştırmacının bir tane daha atayım dememesi lazım. Kurallara baştan karar verilmeli ve kurallar asla değişmemeli. Müdahalelerden biri diğerinden anlamlı olarak başarılı sonuç veriyorsa, çalışmayı süresi dolmadan durdurabilirsiniz, ama kuralları bozduğunuz zaman hastaya da çalışma sonuçlarına da daha çok zarar vermiş olursunuz.

LOJİSTİK REGRESYON

Ö. ERGÖNÜL

Marmara Üniversitesi Tıp Fakültesi
Enfeksiyon Hastalıkları ve Klinik
Mikrobiyoloji Anabilim Dalı, İstanbul

Verilerin Analize Hazırlanması

Analize verileri gözden geçirmekle başlamak gerekir. Çalışmanın bulguları sunulurken genellikle birinci tabloda çalışma grubunun cinsiyet, yaş, meslek dağılımı gibi özellikleri sunulur. İkinci tabloda ise tek değişkenli analiz yer alır. Bazen bu iki tablo birlikte verilir. Son olarak da çok değişkenli analiz bulguları tek bir tabloda sunulur.

Veriler genel olarak ikiye ayrılır. İlk olarak ham verilerden söz edebiliriz. Sürekli veya sayısal veri de denilir. Kolesterol, yaş, ALT, AST gibi hiçbir sınıflama işlemi yapmadan kullandığımız verilerdir. Bazen verileri kategorize etmek durumunda kalabiliriz ve buna kategorik veri denilir. Örneğin cinsiyet, 1 veya 0 olarak kategorize edilir; dikotom veri adını alır. Meslek grupları kategorik veridir ve nominal olarak tanımlanır. Nominal veri gruplarının aralarında, örneğin kanser evrelemesi gibi veya bazı hastalık gruplarında olduğu gibi hiyerarşi varsa buna ordinal veri denilir. Bilgisayar programı, kendiliğinden bu hiyerarşiyi bilemez. Kimin çiftçi, kimin hekim olduğunu bilemeyeceğinden, kategori numaralarını sürekli sayısal değişken yani kendi aralarında bir sırası olan veriler gibi ele alır, örneğin yaş değişkeni için nasıl bir işlem yapıyorsa burada da aynısını yapar.

Bazen ayrı bir sütun (kolon) yaratarak türev (*dummy*) değişken oluşturmayı tercih ederiz. Meslekler içinde çiftçi için ayrı bir sütun açılmış, hemşire için ayrı bir sütun açılmış, çiftçi olanlar “1”, olmayanlar “0” gibi (Tablo 1).

Daha sonra önem verdiğimiz değişkenleri seçmek gerekir. Bu seçim kritik önemdedir, aynı zamanda en çok zorlanılan işlerden de bir tanesidir.

Pek çok veri topluyoruz, her birini sütuna koyuyoruz ve bunlar çoğu zaman birbiriyle lineerlik (paralellik) gösterebilir. Buna multikollinearite denir. Multikollinearite regresyon için önemli bir kavramdır. Birbirine çok yakın şeyler ifade eden iki tane bağımsız (*independent*) veriyi ilişkilendiren en klasik örneklerden bir tanesinde, sigara içenler aynı zamanda kibrit tashışlar (bunun doğru bir örnek olduğunu düşü-

Tablo 1: Nominal verilerden yeni sütunlar oluşturulması.

Meslek	Nominal	Türev (<i>dumy</i>) Çiftçi	Türev (<i>dumy</i>) Hemşire
Çiftçi	1	1	0
Ev hanımı	2	0	0
Hekim	3	0	0
Hemşire	4	0	1
Polis memuru	5	0	0

nürsek ya da yanlışlığını bir yana bırakırsak), kibrit taşıyanlar da kanser olur diyebiliriz. Bilgisayar size hem sigara içmenin hem de kibrit taşımanın eşit ölçüde kanser etkeni olduğunu gösterebilir. İşte bu tür verileri ayıklamak gerekir, en baştan, bu ayıklama işlemini bilgisayara da devredebilirsiniz. Ama en güvenilir yöntem, çalışmayı yöneten insanların, yani, işin mantığına sahip insanların biyolojik ilişkileri ayıklayıp ortaya koymasıdır. Bazı örnekler:

1. Sigara ve alkol.
2. Hepatik kanserlerleri izlerken siroz olmayı, alkolün daha ön planda olduğu bir şeyi düşünüyorsanız, alkolü tercih edersiniz.
3. Romatoid artritte sedimentasyon ve CRP her ikisi de ayrı ayrı etkiliyor olabilir.

CRP ve sedimantasyon arasında öyle bir etkileşim söz konusu olabilir. O zaman birini tercih etmek gerekir. Etkileşim (*interaction*) testi ne bakılabilir, iki değişkeni karşılaştırdığınızda, korelasyon yaptığınızda, X ve Y eksenlerinde, eğer her iki bağımsız değişken %90'ın üzerinde korelasyon gösteriyorsa o zaman birinden fedakarlık etmek zorunda kalabilirsiniz. Kolinearite ve karıştırıcı (*confounder*) değişken farklı şeylerdir. Herhangi bir bağımsız değişken, sonuca gitmeden önce bir başka bağımsız değişkenden geçiyorsa, ara değişken (*intervening*) olarak adlandırılır. LDL, obezite veya her ikisi aynı anda paralel olarak sonlanımı (*outcome*) belirleyebilir.

Eksik Veriler Sorunu

Genel bir kural olarak, toplanan verilerin tümünün kullanılmasına özen gösterilir. İsten-

meyen bir yaklaşım olsa da, bazen kayıp verileri olan bir değişken tamamen analizden çıkarılabilir. Bazı durumlarda kayıp veriler ortalama değerler ile kapatılabilir. Kayıp verili ve doldurulmalı sonuçlar karşılaştırılır, ikisi arasında bir fark yoksa o zaman kayıp veri ile devam edilebilir. Ama şu veya bu şekilde kayıp veriler ortalama ile doldurulduğunda, farklılık çıkıyorsa daha temkinli yaklaşmak gerekir.

Tek Değişkenli Analiz

Bağımsız değişkenlerle sonuç (*outcome*) değişkeni arasında “tek değişkenli analiz” yapılır. Bir bakıma, birer birer değişkenler gözden geçirilir. Makalenin birinci veya ikinci tablolarında genelde tek değişkenli analizler sunulur. Parametrik ve parametrik olmayan testlerin tercih nedenleri belirtilmelidir. Korelasyonun tek değişkenli bir analiz olduğu unutulmamalıdır. Ancak korelasyon eğrisinin oluşturulması ve analizi, regresyon analizini daha kolay anlamamızı sağlar.

Çok Değişkenli Analiz

Lineer regresyon en kolay anlaşılandır (Tablo 2). Burada elbette ki her istatistiksel analizde olduğu gibi bir bağımsız değişkenler ve bir de bağımlı değişken vardır. Etkili olabilecek bağımsız değişkenler, parametreler veya faktörlerle birlikte sonuç olarak *outcome* veya *dependent variable* denilen değişkenler vardır ve bunların yapısına göre çok değişkenli analiz tarzımız değişecektir. Örneğin sonuç veya bağımlı değişken sürekli değişken ise o zaman lineer regresyonu kullanılacaktır. Şu örneği dü-

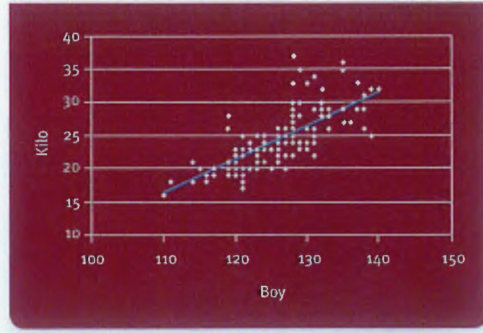
şünebiliriz: X eksenini sürekli değişken, Y eksenini yani outcome grubu kolesterol olsun. Yüksek kolesterolü etkileyen faktörleri de X eksenine koyabilirsiniz. Bunlar yaş, obezite gibi parametreler olabilir (Şekil 1). Tıpta, lojistik regresyona göre göreceli olarak daha az kullanılır.

Tıpta doğayla uğraşmaktan kaynaklanan bakış açısıyla “var” veya “yok” diye bakılır; ya vardır ya yoktur, ya ölmüştür ya kalmıştır, ya hastadır ya değildir, ya ilaç almıştır ya almamıştır. Yani 1 veya 0 ile ifade edilen dikotom değişkene bakılarak lojistik regresyon kullanılır. Eğer sürecin içinde bir de zaman faktörü varsa sağkalım (*survival*) analiz yapılır.

Cox regresyonda, uygulama lojistik regresyona benzer; bağımlı değişken dikotomdur.

Bir de çok nadiren kullandığımız, örneğin 30.000’de bir, 100.000’de 1 gibi nadiren ortaya çıkan olaylarda kullanılan *poisson* regresyon vardır. İlaç yan etkilerinde sıkça kullanılan bir regresyon türüdür.

Tıpta en fazla kullanılan çok değişkenli analiz yöntemi, lojistik regresyondur (Tablo 2). Lojistik regresyonu anlamak için lineer regresyonu gözden geçirmek gerekir. Lineer regresyonu anlamak için de korelasyonu bilmek gerekir. Bütün mesele soyut bir doğrunun çizilmesidir. Bağımsız ve bağımlı değişkenler arasındaki bir ilişkiyi doğrusal bir ifade ile gösterebilir miyiz? Matematikte doğru, bir lineer denklem ile ifade edilir. Lise matematiğinden hatırlarsak, $Y=aX$ bir doğrudur, sıfırdan başlar. İşte lineer yani doğrusal regresyonda, lineer bir doğru çizmiş oluyoruz ve bütün parametrelerimizi bir doğru üzerinden anlatmaya çalışıyoruz. Burada noktasal dağılım (*scatter plot*) yapılırsa, X’e karşılık gelen noktaları buradan alıp, yine Y’ye denk gelen noktalarla bağladığımızda bir noktalar kümesi oluşur. Grafikte X-Y eksenini koordinatlar düzleminde. Daha sonra bunları bir araya getirdiğimizde (korelasyon da aynı mantıktır) hayali bir doğru çizilir. Bunu bizim yerimize bilgisayar programı çizer. Bu aslında her bir noktanın bu doğruya olan mesafesinin toplamının sıfırlanmış bir korelasyon-



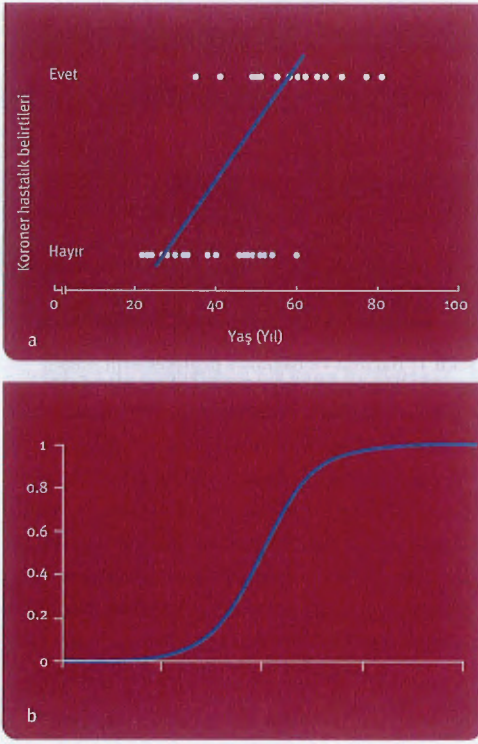
Şekil 1: Bağımlı ve bağımsız değişkenlerin sürekli (sayısal) değişkenler olması.

Tablo 2: Eksik verilerle başatmenin yolları.

1. Eksik veri olan olgular atılır.
2. Eksik verisi olan olgular için dikotom değişkenler yaratılır.
3. Eksik verilerin olduğu değişkenler çıkarılır
4. Eksik verileri sağlamak için ek bir çaba sarfedilir.
5. Eksik verilerin değeri kestirilmeye çalışılır.

Tablo 3: Regresyon analizinde bağımlı değişkenler.

Regresyon türü	Bağımlı değişkenin türü
Lineer	Sürekli (sayısal)
Lojistik	Dikotom
Cox	Dikotom
Poisson	Nadir



Şekil 2a ve b: Lojistik transformasyon.

dur; yani bilgisayarın çizdiği sanal bir doğrudur.

Bağımlı değişken (Y) ekseninde kesişim (*intercept*) noktasından başlayabilir. Sanal bir doğru, matematiksel bir denklemi ifade eder. Böyle bir denklem ile ifade edilen bir modelde boyunuz, yaşıınız, genetik mirasınız, kalp riskiniz vb ilgili parametreler hep birlikte girildiğinde, MI geçirme riskiniz nedir gibi bir kestirim değeri sunulabilir.

Lineer regresyonda, bağımlı değişkenin tipi sürekli (sayısal) değişkendir. Lineer regresyonda bir doğru çizmek kolaydır. Oysa lojistik regresyonda bağımlı değişken dikotomdur. Bu durumda bağımsız değişkenlerle bağımlı değişken arasındaki ilişkiyi ifade eden doğru çizilebilir mi? Acaba, lineer ilişki yerine lojistik bir ifadeye geçerken nasıl bir transformasyon yapılabilir? Şekil 2a'da görülen "1" ve "0" üzerindeki doğruyu biraz aşağıdan biraz yukarıdan bükerseniz, S şeklinde bir doğru elde edersiniz (Şekil 2b). Burada logaritma fonksiyonunu kullanarak, logaritmik bir transformasyonla dönüşüm yapılıyor.

Odds Oranı

Odds oranı Türkçe'ye "olma olmama" ya da "kaçakaç" oranı olarak çevrilebilir. İki odds'un oranlanmasıyla elde edilir. Bu kavramın matematiksel ifadesi yerine sezgisel anlamını ve günlük hayattaki kullanımını kavramak bizler için yeterli olacaktır. Odds oranının matematik ifadesi, e üzeri regresyon katsayısı beta şeklindedir. Lineer regresyonda *least square* denilen yani doğruya olan mesafe ile ilgili bir kavramdır.

Lojistik regresyon bize bazı olanaklar sunuyor; parametrelerin normal dağılıp dağılmadığına bakmaksızın, lojistik regresyon ile işlem yapabiliyoruz. Fakat lojistik regresyonun yaygın kullanımı, bir dizi yanlış kullanımı da beraberinde getiriyor. Diğer taraftan lojistik regresyon bize odds oranı sunan yani yorumlaması kolay bir rakam sunan bir sistemdir. Odds oranının yaygın olarak kullanımının nedeni yorumlamasının kolay ve anlaşılır olmasıdır. Bir

ilacı alanlar almayanlara oranla kaç kat daha fazla etkileniyor ifadesi anlaşılır bir şey. Ancak, odds oranı kavramı her zaman kullanılabileceğimiz ve her durumda güvenilir bir kavram değildir. Bu uyarıyı yapmak durumundayız, çünkü odds oranının relatif risk yerine rahatça kullanıldığına tanık olmaktayız. Kohort çalışmalarda kullanılan relatif risk, kolay anlaşılabilen bir kavramken, odds oranı relatif risk yerine kestirim değeri taşır. Odds oranı yükseldikçe, örneğin 6'yı, 8'i aşarsa yanıltıcı olabilir. Bazı yazılarda odds oranını örneğin 30 görüyoruz. 30 kat fazla gibi ne kadar güçlü bir ilişki yakaladık diye sevinmek için acele etmemek gerekir. Odds oranı çok yüksek saptanıyorsa, örneklem sayısının düşüklüğüne işaret edebilir. Odds oranı 6'yı ya da 8'i geçtiği zaman, relatif risk yerine kullanılma özelliğini kaybeder.

Örneğin bir çocuk doğunca, kız ya da erkek doğma olasılığı %50'dir. Kız olma ihtimalinin erkek olma ihtimaline oranı %50 bölü %50; eşittir 1. Odds oranı 1'dir. Eğer p değeri (0 ile 1 arasında bir sayı) bir olasılıksa, başarının (olma), başarısızlığa (olmama) oranı 1'dir. Elbette çocuğun kız veya erkek olması bir başarı (*success*) veya başarısızlık (*failure*) değildir ama ikisini oranladığımızda odds oranına (o'dan başlayıp sonsuza kadar gidebilecek bir sayı) ulaşabiliriz. Diyelim ki bir şeyin gerçekleşme olasılığı 0.75 ise (p), gerçekleşmeme olasılığı (1-p) 0.25 olacaktır (1-0.75). Bu durumda "kaçakça" oranı $p / (1-p)$ 3 olarak hesaplanacaktır.

Odds oranı, iki odd'un oranıdır. Bu oran olgu kontrol çalışmalarında kullanılır. Kohort çalışmalarında ise relatif risk verilir. Kohort çalışmalarını yorumlamak daha kolaydır. Örneğin Çernobil kazasından sonra radyasyona maruz kalanlar ve kalmayanları izlediğimizi farz edip 20 yıl öncesinden kohort çalışmasını başlattık diyelim, 20 yıl boyunca bu insanları takip ettiğimizi ve acaba 20 yıl boyunca kaç kişi kanser oldu, kaç kişi olmadı şeklinde izlediğimizi varsayalım. Burada riske maruz kalanların kalmayanlara göre kanser oranı ile relatif riski tanımlamış oluruz. Adı üzerinde, relatif olarak riske maruz kalanlarla kalmayanları oranlıyoruz. Örneğin Fatsa'da oturanlar, Bodrum'da oturanlara oranla, yani radyasyon bölgesinde maruz kalanlar, kalmayanlara göre 3 kat daha fazla kanser oluyorlar. Ancak basit gibi görünse de, olgu kontrol çalışmalarının matematiksel kavramlaştırılmaları zordur; yine eğip bükme yapmak zorunda kalırız.

Tablo 4'de oral kontraseptif kullanımı ile miyokard enfarktüsü arasındaki ilişkiyi görüyoruz. Olgu kontrol çalışması sistemi gibi, burada odds oranı 4.8, yani oral kontraseptif kullananlarda, kontrollere göre olay oranı 4.8 kat daha fazladır.

Tablo 5'de sigara ve miyokard enfarktüsü arasındaki ilişki gösterilmiştir. Sigara ilişkisine bakarsak, sigara içenlerde içmeyenlere göre olay gelişme oranı 2.3 kat daha fazladır.

Tablo 4: Oral kontraseptifler ve miyokard enfarktüsü arasındaki ilişki.

Oral Kontraseptif	Miyokard Enfarktüsü	Kontrol	Odds Oranı
Evet	693	320	4.8
Hayır	307	680	referans
	1000	1000	

Tablo 5: Sigara ve miyokard enfarktüsü arasındaki ilişki.

Oral Kontraseptif	Miyokard Enfarktüsü	Kontrol	Odds Oranı
Evet	700	500	2.3
Hayır	300	500	Referans
	1000	1000	

Oral kontraseptif kullananların bir kısmı sigara içiyor. Bu faktörü nasıl ayırtacağız? Eğer sigaraya göre tabakalama yaparsanız sigarayı bir değişken olmaktan çıkarırsınız. Sigara içenler arasında odds oranını 6 kat fazla (Tablo 6), içmeyenlerde 3 kat fazla (Tablo 7) bulursunuz; yani her ikisinde de oral kontraseptif daha etkilidir.

Ancak tabakalanacak çok sayıda değişken olduğunda bu işlemi yapmak kolay olmaz. Bu durumda regresyon analizi, yani çok değişkenli analizler yaparız. İşte regresyon modelleri oluşturmanın nedeni budur; değişkenleri kendi içinde parçalara ayırmak ve analiz yapmak. Bilgisayar programları bizim yerimize bu işi yapıyor. Regresyon analizi 1970 ve 1980'lerden sonra yoğun olarak kullanılmaya başlanmıştır, daha önce klasik olarak stratifikasyon yani tabakalama daha sık yapılırdı.

Model Oluşturma

Lojistik regresyon basit anlamda bir model oluşturma girişimidir ama matematiksel modelleme ile karıştırılmamalıdır. Bir sonucumuz var ve bu sonucu belirleyecek 3, 5, 10, 70 veya "n" sayıda bağımsız değişkeni modele dahil ediyoruz. Soru, bağımsız değişkenlerin seçiminde dükümlemektir:

1. Değişkenlerin tümünü bir arada modele dahil ederek okurlara p değerlerini, odds

oranlarını ve güven aralıklarını toplu halde sunabiliriz. Güven veren bir yöntemdir.

2. Tek değişkenli tabloda bulduğumuz anlamlı değişkenleri modele dahil edebiliriz. Sıklıkla tercih edilen yöntemlerden biridir. Ancak burada, biyolojik sürecin özelliklerini dikkate alarak, yapılan seçimin doğru olup olmadığı gözden geçirilmelidir.
3. Bilgisayar programlarının otomatik yaptığı ileri (*forward selection*) veya geri (*backward selection*) seçimler yapabiliriz. Bu yöntemde, değişken seçimini bilgisayara bırakmış oluruz. İleri seçimde, p değeri en küçük olandan başlayarak tek tek değişkenler modele dahil edilir ve istatistik değeri anlamlı olanlar modelde korunur. Geriye yönelik olarak da, tüm değişkenlerden başlanır, p değeri en yüksek olanlar ardışık olarak atılır. Her iki yöntemde de istatistik açıdan anlamlı kalan değişkenler sunulur.

Ancak bu yöntemlerin her birinin kısıtları vardır. Biyolojik ilişkinin altını bir kez daha çizmeliyiz. Seçim yaparken düşünüp gerekli ve anlamlıysa örneklerin seçilmesi ve bunun okuyucuya dürüst bir şekilde gösterilmesi gerekmektedir. Sonuç anlamlı çıkmayabilir; tablodaki bütün p değerleri anlamlı çıkacak diye bir şart yoktur. Anlamsız olduğunu görmek de güzeldir. Bu güven aralığıyla (*confidence interval*) telafi edilerek ifade edilebilir.

Tablo 6: Sigara içenlerde oral kontraseptif ve miyokard enfarktüsü arasındaki ilişki.

Oral Kontraseptif	Miyokard Enfarktüsü	Kontrol	Odds Oranı
Evet	517	160	6
Hayır	183	340	referans
	700	500	

Tablo 7: Sigara içmeyenlerde oral kontraseptif ve miyokard enfarktüsü arasındaki ilişki.

Oral Kontraseptif	Miyokard Enfarktüsü	Kontrol	Odds Oranı
Evet	176	160	3
Hayır	124	340	referans
	300	500	

Modelin Geçerliliği

Modelin uyumluluğu (*goodness of fit*), uygulanan regresyon modelinin uyumluluğunun bilinmesi de önemlidir. Bir anlamda bütün modelin anlamlı olup olmadığı saptanmaya çalışılır. Bir bakıma, metodun geçerli olup olmadığı nı tek bir p değeri ile gösterilmeye çalışılır. Bunun için birçok sistemde uygulanan olabilirlik oranı (*likelihood ratio*), *R square* veya Hosmer ve Lemeshow'un adlarıyla anılan bir test uygulanabilir.

Tablo 8'de yaş ve cins değişkenlerinin metisilin dirençli *Staphylococcus aureus* (MRSA) gelişmesi üzerindeki etkisinin lojistik regresyon sonuçlarını görüyoruz.

Tablo 8'de, olabilirlik oranı (*likelihood ratio*) 6, *R* 0.04 iken, Tablo 9'da insülin kullanımını modele eklemekle *R* 0.017 olmuş yani azalmış, buna karşılık olabilirlik oranı (*likelihood ratio*) artmış görünüyor. Uygun bir parametre (insülin kullanımının MRSA oranını arttıran bir parametre) eklemekle testin gücünü artırmış olduk. Modelin *R*2 değeri 0 ve 1 arasında yer alır. Bu değer 1 olması mükemmel uyumu, 0 olması da zayıf uyumu gösterir. An-

cak bu değerler analizin sonucunda genellikle sunulmaz.

Regresyon Sonuçlarının Sunumu

Eğer odds oranı ile beraber güven aralığı veriliyorsa o zaman p değeri verilmeyebilir, çünkü güven aralığı ile aynı zamanda p bilgisi de verilmiş oluyor. Eğer güven aralığı 1'i kapsıyorsa (odds oranı olduğu için, herhangi bir şey bir şeyin bir katıysa o zaman kendisidir ve dolayısıyla kendisi de olabilen bir şeyin anlamı yoktur düşüncesiyle), arada "1" varsa zaten anlamlı değildir. Örneğin, Tablo 8 ve 9'da cins değişkeni odds oranı 1.29, güven aralıkları ise 0.38 ve 4.3 olup, 1'i kapsamaktadır. Görüldüğü gibi p değeri de 0.678 olarak belirlenmiştir ve istatistiksel açıdan anlamlı değildir.

Güven aralığı ne kadar genişse çalışmanın gücü o kadar az demektir. Örneklem büyüklüğünüz yeterliyse, güven aralığı dardır. Güven aralığı genişlemişse o zaman örneklem büyüklüğü olasılıkla küçüktür diye düşünmek ya da çalışmanın gücü küçüktür diye düşünmek mümkündür.

Tablo 8: Lojistik regresyon ile yaş ve cinsin MRSA üzerine etkisinin STATA versiyon 11 (USA) programında gösterilmesi.

Logistic regression		Number of obs	=	900
		LR chi2(2)	=	6.08
		Prob > chi2	=	0.0479
Log likelihood = -56.343561		Pseudo R2	=	0.0512

mrsa	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
yas	1.040297	.017497	2.35	0.019	1.006563 1.075162
gender	1.288707	.7886503	0.41	0.679	.3883686 4.276261

Tablo 9: Lojistik regresyon ile yaş, cins ve insülin kullanımının MRSA üzerine etkisi STATA versiyon 11 (USA) programında gösterilmesi.

Logistic regression		Number of obs	=	900	
		LR chi2(3)	=	10.10	
		Prob > chi2	=	0.0177	
Log likelihood = -54.332075		Pseudo R2	=	0.0850	

mrsa	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]

vas	1.037847	.0179971	2.14	0.032	1.003166 1.073727
gender	1.291911	.7965018	0.42	0.678	.3858045 4.325445
insulin	7.362583	6.092891	2.41	0.016	1.454186 37.27695



Şekil 3: Karıştırıcı değişken olarak hastalığın ciddiyeti.

Bağımsız Değişkenlerin Etkileşmesi (Interaction)

İki ya da daha fazla değişken birbiriyle etkileşebilir ve yanlış sonuçlar elde etmemize neden olabilir. Bu sorunu gidermek için istatistiksel yöntemler kullanılabilir. Bu amaçla, ilgili iki değişkenin çarpımıyla yeni bir değişken oluşturulur ve bu değişkenin modele dahil edilmesiyle sonucun değişip değişmediğine bakılır.

Bu yönteme başvurmadan daha pratik yöntemlerle etkileşimi kestirmek mümkündür. Bir değişkeni modele katmakla, bir başka değişkenin odds oranında yüzde 20'nin üzerinde bir değişme varsa veya anlamlılığı değişiyorsa etkileşim (interaction) vardır. Ayrıca, p değeri anlamlı iken anlamsız hale geliyorsa veya tersi söz konusuysa etkileşim olduğunu anlayabiliriz.

Karıştırıcı Değişken (Confounder)

Sözcük anlamı olarak “*confounder*”, karıştırıcı anlamında İngilizce’de kullanılmaktadır.

Karıştırıcı değişken, hem bağımlı değişkeni, hem de başka bir bağımsız değişkeni etkileyen değişkendir (Şekil 3). Karıştırıcı değişkenlerle baş etmenin yollarından ilki, biyolojik süreçleri gözden geçirmektedir. Ayrıca istatistiksel yöntemlerle de baş etmek mümkündür. *Propensity* (eğilim) skoru uygulaması bu önlemlerden biridir.

KAYNAKLAR

1. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott William&Wilkins, 2008.
2. Katz MH. Multivariable Analysis. A Practical Guide for Clinicians. 2nd ed. New York; Cambridge University Press, 2006.
3. Hosmer DW, Lemeshow SL. Applied Logistic Regression. 2nd ed. New York; Wiley, 2000.

PROPENSITY (EĞİLİM) ANALİZİ

Ö. ERGÖNÜL

Çeşitli nedenlerden dolayı randomizasyon yapma şansı yoksa, eldeki verilerle gözlemsel çalışma yapmak durumunda kalındığında, bu verilerin en doğru biçimde analiz edilmesi istatistikçilerin ve epidemiyologların çok önemsendiği konulardan biridir. *Propensity* (eğilim) analizi bu yaklaşımlardan bir tanesidir. Gözlemsel çalışmalarda şans faktörünü en aza indirmek için bir dizi işlem yapılır. Uygun eşleme, tabakalandırma ve nihayet regresyon sıkça başvurulan yöntemlerdir. Ek yöntemlerden birisi de *propensity* skoru ile yapılan *propensity* analizidir.

Propensity Skoru

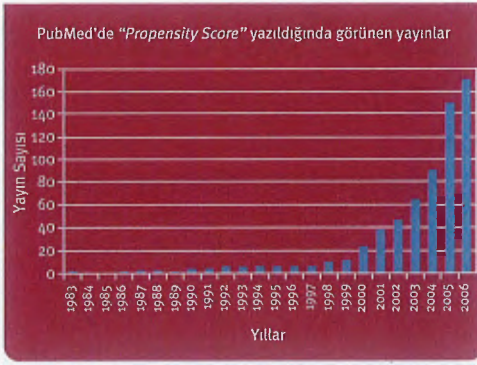
Propensity skoru, lojistik regresyon yardımıyla elde edilen ve regresyona bağımsız değişken bir parametre gibi girilen bir skordur. Böylelikle, *propensity* skoru, karıştırıcı (*confounder*) değişkenlerin kontrol edilmesine ek olarak uygulanan bir yöntemdir. Özellikle bağımlı değişken olarak kullanılan sonuç sayısı az ise, *propensity* skoru daha önemli hale gelebilir. Anonim olarak, lojistik regresyonda 10 bağımsız değişken verisi için en az bir sonuç verisi olması istenir, bu da %10 demek olur ki, yüksek orandır. Rosenbaum, 1983'ten bu yana *propensity* skorlarının anlamlı olması gerektiğine dair yayınlar yapmış, Cochrane *propensity* skorlarının öneminden söz etmiştir. Yöntemsel açıdan, 2002 ve 2004 yıllarında romatoloji dergilerinde sıkça karşılaşılan ve 2004'te *Journal of Rheumatology*'de ele alınan bir yöntemdir. Son yıllarda daha sık karşımıza çıkmaktadır (Şekil 1).

Sözcük olarak *propensity*, eğilim, temayül anlamındadır; örneğin '*a propensity to drink too much*', yani içmeye eğilimli gibi. *Propensity*, hastanın geleceğine dair sağkalım açısından fikir vermesi veya eğilim belirtmesi gibi anlaşılabilir.

Swan-Ganz Kateteri Örneği ve Propensity Skoru

Bugüne kadar yapılmış en geniş çalışmalardan biri, çok önemli tarihsel bir örneği Boston merkezli, kardiyologların yaptığı, Harvard

Marmara Üniversitesi Tıp Fakültesi
Enfeksiyon Hastalıkları ve Klinik
Mikrobiyoloji Anabilim Dalı, İstanbul



Şekil 1: Yıllar içinde *propensity* skoru kullanımının artışı.



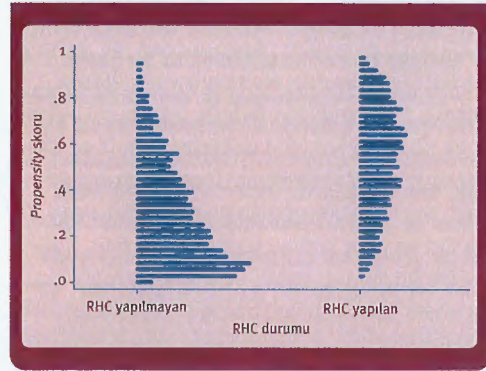
Şekil 2: Karıştırıcı değişken olarak hastalığın ciddiyeti.

dahil başka merkezlerin de katılımıyla yapılmış olan bir çalışmadır. Swan-Ganz kateteri, sağ kalp kateterizasyonu ve sağ kalp fonksiyonunu takip etmek için kullanılır. İlk kez, 1970'lerde bulunuyor ve hekimlere bir güven hissi veriyor. Hekimler, bu kateterleri taktıklarında hastalarını daha iyi takip edebiliyorlar. Ama bir süre sonra 1980 ve 1990'lara gelindiğinde, Swan-Ganz kateterleri takıldığında ölüm oranlarının arttığı iddia ediliyor. Bu durumda, Swan-Ganz kateteri kullananlar acaba hata mı yapıyorlardı şeklinde sorular sormaya başlıyorlar. Diğer yandan şu haklı soru beliriyor: Acaba hekimler daha hasta olanlara mı Swan-Ganz kateteri takma eğilimindeydi, böyle bir eğilim (*propensity*) mi gösteriyorlardı? Bu durum, "*confounding by indication*" olarak bilinir. Soru şudur, "*confounding by indication*"dan nasıl kurtuluruz? Lojistik regresyon yapmak kendi başına her zaman yeterli olmaz. Çünkü lojistik regresyon yaparken model oluşturulmasında ciddi sorunlar vardır. Yanlış modelleme ile yanlış sonuçlara gidilebilir. Tasarım aşamasında dikkatli olmak gerekir. Bilgisayar programlarında en ileri sistem, Hosmer-Lemeshow veya *R square* değerleri olsa da, en başta misklasifikasyon yapılmışsa, hastalar yanlış gruplanmışsa, epidemiyolojide temel bir yanlışlık (*bias*) oluşur ve dolayısıyla yanlış sonuç kaçınılmaz hale gelebilir. İşte bu yanlışlık çoğu kez "*confounding by indication*" nedeniyle ortaya çıkar. Aynı tablo, sağ kalp kateterizasyonu mu yoksa hastanın ciddiyeti mi daha etkili sorusunda ortaya çıkar. Kateterizasyon ve sağ kalp yetmezliğinin ciddiyeti ölümle ayrı ayrı ilişkilidir. Ayrıca kendi aralarında da etkilenirler. İşte bu tam bir karıştırıcı (*confounder*) değişken sorunudur (Şekil 2). Benzer şekildeki olgulara yoğun bakımlarda karşılaşıyoruz. Bir enfeksiyon çıkar, ölüme neden olabilir ama olmaya da bilir. Bu durumda sosyal bilimlerden ödünç alınan "*counter-factual*" kavramı konuyu anlamamızda yardımcı olabilir. Zaman tüneli örneğinden biliriz; zaman tüneline girsek, 100 sene öncesine gitme şansımız olsa ve gözlemlesek, acaba tarih farklı gelişebilir miydi? Her şeyin başına dönsek, acaba süreç nasıl gelişirdi? Bu sorularla başlayan yaklaşım "*counter-factual*" olarak anılır.

Bu kavramsal yaklaşımlardan sonra kateter sorununa gelelim; 1996'da 5 merkezli bir çalışmada araştırma sorusu, Swan-Ganz kateterleri etkili mi, değil mi, sonucu nasıl etkiliyor? Prospektif kohort bir çalışma tasarlanıyor [1]. Yedi uzman seçiliyor. İlk başta, her bir hasta için bir *propensity* (eğilim) belirleniyor. Hastalara ölüm-kalım için bir skor konuyor (bu skora, 1 den 5'e kadar veya sıfırdan 100'e doğru percentil verilerek yapılabilir). Elbette bu skora yine hastanın belli parametrelerine bakılarak yapılıyor. Bu parametrelerin bir kısmı zamana bağlı olarak değişebilir ve bu durumda işe yaramaz; klinisyen bunu bilir ama istatistikçi bilemeyebilir. *Propensity* skorunun temel sorunlarından biri, yapılan skorlamanın zaman içinde değişebilmesidir. Bu skorlamayı yapmak ve önemli parametreleri belirlemek için de lojistik regresyon kullanabilirsiniz. Ancak, bu skora daha sonra daha geniş bir lojistik regresyon modelinde parametrelerden biri olarak kullanılacaktır. Ya da hastalar *propensity* skorlarına göre eşleneceklerdir. İşte bu çalışmada 1008 çiftte ulaşılmış ve sonuçta *quasi* (yarı) randomizasyon yapılmış. Mademki burada, etik nedenlerle (örneğin doktorun, "biz aslında Swan-Ganz kateterizasyonunu öneriyoruz ama siz itiraz ederseniz size yapmayacağız, kontrol grubunda olacaksınız" demek durumunda kaldığı hastaların varlığında) randomizasyon yapamıyoruz, o zaman yarı randomizasyon anlamına gelen *quasi* randomizasyon yapıyoruz.

Hastaların eşleştirilmesinde değişik teknikler kullanılabilir. Sonuçta *propensity* yapmaksızın elde edilen odds oranlarında, çalışmanın sonunda *propensity* skorun eklenmesiyle bir düzeltme (*adjustment*) yapılıyor.

Sağ kalp kateterizasyonu (RHC) yapılmıyorsa, hastaların durumlarının daha iyi olması halinde *propensity* skoru düşük olacaktır. Şekil 3'ün sol tarafında görüldüğü gibi, *propensity* skoru büyüdükçe, hasta sayısı giderek (yukarıya doğru) azalmıştır. Oysa diğer tarafta, sağ kalp kateterizasyonu yapılan olgularda *propensity* skorunun hemen hemen normal dağıldığını görüyoruz. Sorun asimetrik (normal dışı) dağılan grafiğin sonuçlarımızı etkilemesinden korunmaktır.



Şekil 3: Sağ kalp kateterizasyonu (RHC) yapılmayan ve yapılan hasta gruplarında propensity skorlaması.

Kırım Kongo Kanamalı Ateşi Örneği ve Propensity Skoru

Başka bir örnek Kırım-Kongo kanamalı ateşi (KKKA) ile ilgili. KKKA'da kullanılacak tek ilaç ribavirin. Ülkemizden bazı meslektaşlarımıza göre ribavirinin faydası yok, hatta öyle ki ribavirin alan grupta ölüm oranı daha yüksek. Sağlık Bakanlığının ilk verilerine göre ribavirin alan hastalarda ölüm oranı %20, almayanlarda %5 olarak sunulmuştu. Bu durumda ribavirin çok zararlı denilmesi gerekir. Ama burada, aynen Swan-Ganz kateterlerinde olduğu gibi temel bir hata yapılıyor. Hastalar, düzgün bir şekilde eşleştirilmeden yani bir *propensity* skorlaması yapılmadan karşılaştırma yapmak hataya neden oluyor. *Propensity* skoru değil ama ciddiyet skalası ile KKKA hastalarının eşleştirilerek bakılması gerekir. Ayrıca, ribavirin bu hastalığın ilk evresinde, vireminin olduğu dönemlerde daha fazla etkin, daha sonra başka bir mekanizma devreye girmekte ve bir tür sepsis gibi, immünolojik bir mekanizma söz konusu olmakta. O evrede ribavirin artık çok etkili değil gerçekten. Anadolu'dan geç gelmiş bir hastaya ribavirin verirsiniz ve bu hasta üçüncü merkezlerden birinde ölürse, bu hastayı da ribavirin kolunda değerlendirmek ne kadar doğrudur? *Propensity* skorlaması kullanmak, randomizasyon yapamayacaksa veya zaman ve para sorunu varsa yararlı olur. Kaliteli bir gözlemsel çalışmada randomizasyona yakın bir sonuç sunar; o açıdan faydalıdır. Ancak randomizasyonda farkında olmadığımız bağımsız değişkenleri kontrol etme şansımız daha yüksektir. Örneğin, henüz bilinmeyen genetik faktörler önemli olabilir ama istatistik analize başlanıldığında ancak toplanılan verilerden sorumlu olunabilir. Onun dışında bilmediğimiz değişkenler sisteme dahil edilmez ve sonuçta yine yanlılık (*bias*) kaçınılmaz olabilir. Bir yazıda *propensity* skoru kullanılmışsa, bunun nasıl yapıldığı ve nasıl tanımlandığı dikkatle incelenmelidir. Hiç bir şeyde olmadığı gibi bu da mutlak bir çözüm değildir.

KAYNAK

1. Connor's AF Jr, Speroff T, Dawson NV, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. JAMA 1996; 276: 889-97.

UZUN DÖNEMLİ SAĞKALIMDA PERİYOT ANALİZİ*

H. Brenner

Almanya Kanser Araştırma Merkezi,
Heidelberg, Almanya

*Hazırlayan: Konuşma özeti Hasan Yazıcı
tarafından, ses kayıtlarından hazırlanmıştır.

Bu bölümde, kanser sağkalım istatistikleri kapsamında geliştirilmiş bir tür sağkalım analizi olan uzun dönemli sağkalımda periyot analizinden söz edeceğim. Günümüzde, yalnız uzmanlar değil hekimler, hastalar ve hatta herkesin kolayca erişilebildiği 5 veya 10 yıllık uzun dönemli sağkalım oranları sadece kanser için değil birçok kronik hastalık için majör sonuç ölçümleridir. Böyle uzun dönemli sağkalım oranlarının sayısının günümüzde birçok kronik hastalık için geliştirilmiş olması sevindiricidir. Bu ölçü özellikle kanser için önemlidir. Hastanızın sağkalım hesaplamasını irdelemek istediğinizde, eldeki veriler çoğu örnekte tatminkar olmaktan uzaktır. Bu sadece akademik bir sorun olmayıp güncel hekimlik uygulaması için de önemlidir. Hastanıza, geçerliliğini yitirmiş sağkalım tahminleri vermek istemezsiniz. İşte bu durum bizim, periyot analizi olarak adlandırdığımız, bu yeni sağkalım analizi yöntemini geliştirmemiz için bir motivasyon oldu. İlk olarak 1996'da kanser hastalarında güncel uzun dönem sağkalım oranlarını hesaplama amaçlı geliştirilen bu yöntem bugün benzer şekilde diğer hastalıklara da uygulanabilmektedir.

Sağkalım Analizinde Yeni Bir Yöntem Olarak Periyot Analizi

Önerdiğimiz yöntemin çok yararlı olabileceğini gösterebilmek için olabildiğince kapsamlı ve titiz bir çalışma yaptık. İlk tanıtımında yöntemimiz, oldukça kuşkuyla karşılanmıştı. Ancak günümüzde artık kanser verilerinde giderek artan sayıda araştırmacı tarafından tercih edilmektedir. Özellikle kanser kayıt sistemleri (*cancer data registries*) tarafından, kanser sağkalımında popülasyon bazlı, güncel kanser sağkalım tahminlerini hesaplamak için kullanılmaktadır. Bugüne kadar romatoloji gibi diğer alanlarda uygulandığını pek görmedim ancak yakın bir gelecekte bu durumun değişeceğini tahmin ediyorum. Ayrıca eğer bir yöntem tanıtıyorsanız, bu insanlar tarafından ulaşılabilir ve kullanımı kolay bir yöntem olmalıdır. Biz bunu sağladık; periyot analizi, uygulaması oldukça kolay, kullanıcı dostu bir yöntemdir.

Yöntemin temel prensibi çok basittir. Hastaların sağkalım deneyimi yakın bir zaman dilimine indirgenir. Periyot adı verilmesinin nedeni budur. Yakın bir zaman diliminin başlangıcındaki gözlemlerin sola budanması (*left truncation: istatistikte kullanılan, eşik değer üstüne çıkıldığında uygulanan yuvarlama durumu*) söz konusudur. Sağkalım analizi, çoğunuzun aşına olduğu bir şeydir. Yeni olan ise, bu sola budama işlemidir. Tamamen yeni bir şey keşfettiğimizi iddia edemeyiz. Periyot analizinin prensibi, bir süredir diğer disiplinlerde kullanılmaktadır ve bunların en bilineni demografidir. Yaşam beklentisini demografide nasıl hesapladığımızı düşünürsek bunun aslında önerdiğimiz yöntemden farklı olmadığını görürüz. Yapılan, zaten on yıllardır kullanılmakta olan bir periyot prensibidir.

Periyot analizinin nasıl işlediği, bir kanser kayıt sistemi üzerine yapılandırılmış haliyle Tablo 1’de basitçe görülebilmektedir. Tanı yılları 1992’den 2004’e kadar sıralanmıştır. Elbette başka hastalık kayıtlarında da örneklendirilebilir. Bir uzun dönem sağkalım oranı elde etmek için, öncelikle insanların uzun süreler boyunca gözlemlenmesi gerekir. Burada hastalar sağkalım yönünden zaman içerisinde izlenerek bir veritabanının ve bu veritabanı üzerinden de sağkalım fonksiyonlarının ögesi oluyorlar. Eli-

mizde 2004'e kadar olan izlem bilgisi var. Tablo 1'den anlaşılacağı üzere, 1992'de tanı konmuş bir hasta ilk yıl kontrollerini, kısmen 1992'de, kısmen de 1993'te olmuştur. İkinci yıl kontrolleri ise kısmen 1993'te ve kısmen de 1994'te gerçekleştirilmiştir ve bu durum benzer şekilde devam etmektedir. En son 2004'te tanı konulan hastalar ise sadece tanı sonrası ilk yıl olan 2004'te ilk sağkalım deneyimlerini tamamlamışlardır.

Sağkalım analizindeki geleneksel yöntemlerle yani kanser sağkalım istatistiklerinde sık kullanılan kohort analizi dediğimiz yöntemle 10 yıllık sağkalım sonuçları için, gerçekten 10 yıl boyunca izlenmiş hastaların eğrisine bakılır. Bunu yapabilmek için, 1990'ların başında tanı konan ve o zamandan sonra 10 yıl boyunca izlenen hastaların saptanması gerekir. Bunun yanında söz konusu yöntemle ancak 1990'ların başındaki bakım kalitesi hakkında fikir edinebiliriz. Analizimizin son yılı olan 2004 yılına yaklaştığımızda tıbbın ilerlemesiyle kanser tedavisinde önemli gelişmeler olmuştur. Bu gerçek doğal olarak geleneksel kohort analiziyle ortaya konamaz.

Geçmişte kullanılmış olan ve benim ‘komplet analiz’ adını verdiğim ya da yaygın bilindiği şekliyle Kaplan-Meier analizi adlı bir başka

Tablo 1: Son 10 yıldaki sağkalım.

[illegible]

yöntemde ise daha yeni tanı konmuş ve 10 yıllık izlemelerini tamamlamamış hastaların prognozu da incelenir. Ancak bu yöntem erken sağkalım oranlarından fazlaca etkilendiğinden öngörüler geçerliliğini yitirebilir.

H. Yazıcı: 1993'te 1/2 derken, aslında 2 hastadan biri mi izlendi demek istiyorsunuz?

H. Brenner: Bu sadece izlem yılları anlamına geliyor, yani 1992'deki hastalar 1992'deki izlemin birinci yılına katkıda bulunmaktadır ve birinci yıllarının bir kısmı aynı zamanda 1993'te yer almaktadır. Yani buradaki sayılar, tanı sonrası yıllardır.

H. Yazıcı: 92'de bir tanı ve bir izlem var. 92'nin altındaki "1" neyi simgeliyor?

H. Brenner: Tanı sonrası birinci yıl.

H. Yazıcı: Tamam, peki bir buçuk (1/2) nedir?

H. Brenner: Bir buçuk (1/2) değil; bu hastalar kısmen 1993'te izlemelerinin birinci yılindalar ve kısmen de ikinci yılindalar. Bu soruları sormanız çok iyi, bunu açığa kavuşturmanın oldukça önemli olduğunu düşünüyorum.

Burada, Tablo 1'deki kırmızı veri kutucuğuna bakalım. Bu son dönem verilerin hepsi 2002-2004 yıllarına aittir. Görünüşe göre, 2001'den 2004'ün son zamanlarına kadar olan dönemde tanı konmuş hastaların, tanı sonrası ilk yıldaki sağkalımları hakkında çok fazla bilgi vardır. Tanı sonrası ikinci yıldaki sağkalımı, 2000-2003 arasında tanı konmuş hastalardan elde ediyoruz ve sistem bu şekilde devam ediyor. 9 ila 10 yıllık izlem için, çok uzun yıllar önce tanı konmuş yaşlı hastalara dönmeniz gerekiyor. Önerilen yöntemle, çok daha güncel uzun dönem sağkalım tahmini elde ediliyor. Bu tür bir sağkalım varsayımı, bize 2002'den 2004'e kadar tanı konulan hastaların beklenen veya öngörülen sağkalımını verecektir şeklinde yorumlanabilir. Bu yorumdaki sağkalım varsayımı, yani tanı sonrası her izlem yılı için varsayılan sağkalım ve araştırma dönemi olan 2002-2004 periyodu için uzlaşmış varsayım yorumu, genellikle periyot yaşam tablolarından çıkarılan varsayılan yaşam beklentisine ol-

dukça analog bir yorumdur. Periyot yaşam tablolarında belli bir dönem için (bizim örneğimizde 2002-2004 yılları arası söz konusudur) mortalite oranları vardır. Bu bize, 2002-2004 yıllarındaki yaşa özgü mortalite oranlarının her yaş için sabit kaldığı varsayılarak, 2002-2004 yılları arasında doğmuş kişilerin yaşam beklentisini vermektedir. Yaşam beklentisinin özüne baktığımızda, her zaman bu periyot prensibine uymamız gerekmektedir. 100 yıl geriye dönüp, doğduktan sonra tüm yaşamları boyunca izlenmiş kişilere bakarsak günümüzdeki yaşam beklentisi ile kıyaslanamayacak düşüklükte beklentiler görürüz. Yöntemimizin temel prensibi özellikle son yılların verilerine verilen ağırlıktır. Söz konusu süre örnekteki gibi 3 yıl, bazı durumlarda 4-5 yıl da olabilir.

Yöntemimize, ilk önerdiğimizde insanlar oldukça kuşkuyla yaklaşmıştı. Onları inandırmak için doğal olarak çok emek vermemiz gerekti. İnsanları inandırmak için kullandığımız yöntem eski kanser kayıtlarındaki tarihi verileri incelemek oldu. Eğer elimizde olsaydı da bu yöntemi o döneme uyarlasaydık, analizimiz o dönemlerde tanısı konmuş hastalarda uzun dönem sağkalımı öngörür müydü sorusunu araştırdık. Bu konuda Finlandiya Kanser Veritabanından çok yararlandık (Tablo 2). Yasal zorunlulukla tutulan, bütünlüğü ve kalitesi çok iyi olan bu veritabanı 1953'ten beri işlevseldir. Biz ilk değerlendirmemizi yaptığımızda veriler 1997'ye dek tamamlanmıştı.

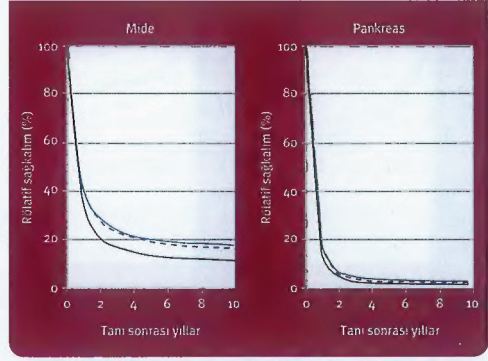
Biz çok basit bir şey yaptık; geleneksel kohort analizini ve önerdiğimiz yöntem olan periyot analizini kullanarak, 1983-1987 arasında tanı konulan hastaların 10 yıllık sağkalım eğrilerini karşılaştırdık. İncelenen grup 1983 ile 1987 yılları arasında tanı almış ve 1997'de 10 yıllık izlemelerini tamamlamış olan en genç ve en yeni kohortu içeriyordu (Tablo 2). Bu hastalar için yaptığımız öngörülerin gerçekleşip gerçekleşmeyeceğini görmek amacıyla varsayım larımızı hastaların sağkalım deneyimleriyle karşılaştırdığımızda yöntemimizin işlediğini anladık.

[illegible]

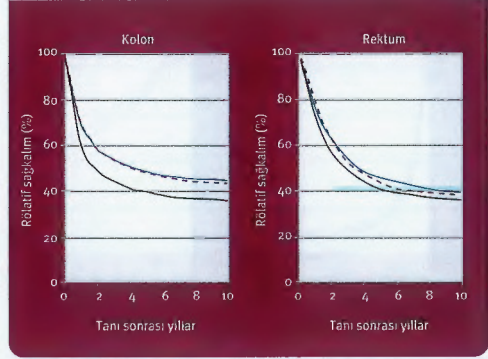
ise, diğer kanser türlerinden görece farklıdır. Birçok hasta tanıdan sonraki ilk iki yılda ölmekte ve daha sonra eğriler birbirlerinden ayrılmaktadır. Meme kanseri için periyot analizi, hala diğer geleneksel yöntemlerden daha üstündür. Fakat periyot analizi bile burada oldukça karamsar kalmaktadır. Çünkü sonuçta kısmen de olsa, daha önce tanı konmuş hastaların geç sağkalım deneyimlerine bağlıdır. Ancak over kanserinde periyot analizi gözlemlenen gerçeği mükemmel bir biçimde öngörmektedir.

Endometriyum kanserinde hangi yöntemi kullanırsanız kullanın, aslında değişen bir şey olmamaktadır. Serviks kanseri için ise durum oldukça gariptir (Şekil 4). İlk yılda, mavi eğrinin bu kanser türünde siyah eğriden daha alçakta yer aldığı görülmektedir. Yani, prognozu zaman içerisinde kötüleşmektedir, ki bu da çok şaşırtıcıdır. İngiltere'deki hekimler bu hastaları eski yıllara kıyasla kesinlikle son yıllarda daha kötü tedavi etmemişlerdir. Ancak bu, aslında seleksiyonla kolay bir biçimde açıklanabilir. Diğer birçok ülkede olduğu gibi İngiltere'de de serviks kanseri için çok sıkı taramalar yapılmıştır. İnsidans oldukça düşmüştür ve popülasyonun oldukça küçük bir kısmı hala kansere yakalanmaktadır. Bunlar, oldukça yaşlı hastalar, diğer kişilere göre hastanelere daha az başvuranlar ve benzeri örneklerdir. Burada ilginç olan, periyot analizinin diğer geleneksel eğri analizinden daha yüksek tahminler vermemiş olmasıdır. Yani eğer Türkiye veya Almanya'da, herhangi bir zamanda tıbbi bakım sistemi çökecek olsa ve her şey daha da kötüleşse bile, bu durumu diğer metodolojilere kıyasla periyot analizi ile daha erken görme şansına sahipiz.

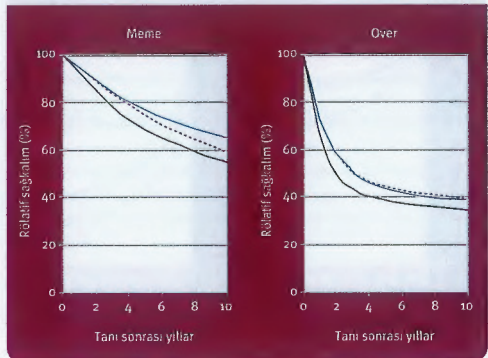
İnsanlar elbette şüpheli olabilir ve tartıştığımız şeyin sadece bir Finlandiya fenomeni olduğunu ileri sürebilir. Sevindirici olan, eşzamanlı olarak örneğin İsveç'ten, Kanada'dan veya Hindistan'dan diğer birçok araştırmacının da bağımsız olarak verilerimizi almış ve birçok değerlendirmede bulunarak çok benzer sonuçlara varmış olmasıdır. Bu nedenle yöntem sadece bir Finlandiya fenomeni değildir. En azından kanser için oldukça evrensel bir fenomen gibi görünmektedir.



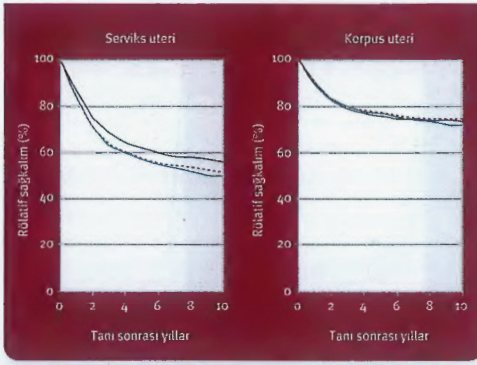
Şekil 1: Mide ve pankreas kanserlerinde periyot analizi ve sağkalım.



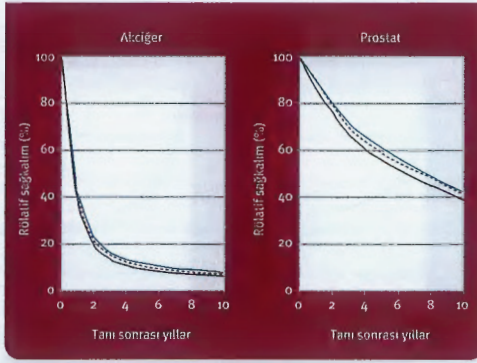
Şekil 2: Kolon ve rektum kanserlerinde periyot analizi ve sağkalım.



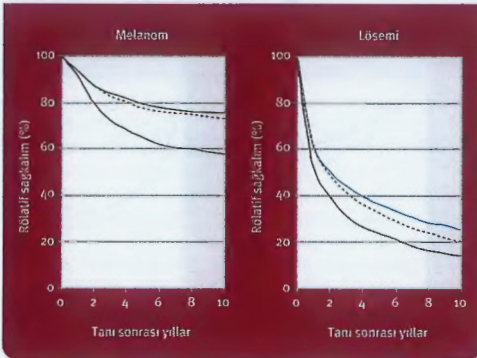
Şekil 3: Meme ve over kanserlerinde periyot analizi ve sağkalım.



Şekil 4: Serviks ve endometriyum kanserlerinde periyot analizi ve sağkalım.



Şekil 5: Akciğer ve prostat kanserlerinde periyot analizi ve sağkalım.



Şekil 6: Melanoma ve lösemide periyot analizi ve sağkalım.

Diğer bir uygulamamız da çocukluk dönemi kanserleri ile ilgiliydi. Özellikle, aynı yöntemin çocukluk dönemi kanserine de uyup uymayacağını görmek istiyorduk. Çocukluk dönemi kanseri oldukça özeldir çünkü sağkalım üzerine çok fazla odaklanılırken bir yandan da tedavinin son dönem etkilerine yönelik bazı endişeler bulunmaktadır. Bu yüzden, Birleşik Devletler Ulusal Kanser Enstitüsü'nde bulunan veriler üzerinde, erişkin kanserinde olduğu gibi, çocukluk dönemi kanseri için de benzer türde bir değerlendirme gerçekleştirdik (Tablo 3). 1999'a kadar tam 10 yıl boyunca izlenmiş hastaların sağkalım verilerine sahiptik. 1980'lerde, 1985'te ve 1989'da tanı konulan hastaların sağkalımlarına baktık ve bu hastaları 10 yıl boyunca izledik. Ardından sağkalımlarını, erişkinlerde yaptığımız gibi geleneksel kohort analizi ve periyot analizi ile kıyasladık. Periyot analizinin (noktalı kırmızı çizgi) aslında gerçekleşeni yakaladığı (mavi çizgi) ve her ikisinin, geleneksel kohort analiziyle (siyah çizgi) kıyaslandığında oldukça iyi bir sonuç sergilediği görülmektedir (Şekil 7). Daha sonra bundan yeterince emin olduk ve 1990'ların sonlarında (kırmızı kareler) tanı konulan hastaların sağkalımını öngörmek için de en son aşamaları periyot analizini kullandık. Periyot analizi ile elde ettiğimiz eğri, çocukluk çağıının tüm kanserlerine uygulandığında daha iyi bir prognoz sergilemektedir (Şekil 7). Çocuklar için güncel sağkalım tahminleri varken daha eski verileri kullanmanın, tamamen etik dışı olduğunu düşünüyorum. Aynı şey, en yaygın çocukluk dönemi kanserleri olan lösemi ve lenfomada da görülmektedir (Şekil 8). Tedavideki gerçek ilerleme nedeniyle son on yıldaki sağkalımda etkileyici bir artış vardır. Biz bunu, geleneksel sağkalım analizine kıyasla periyot analiziyle daha erken bildirebilirdik.

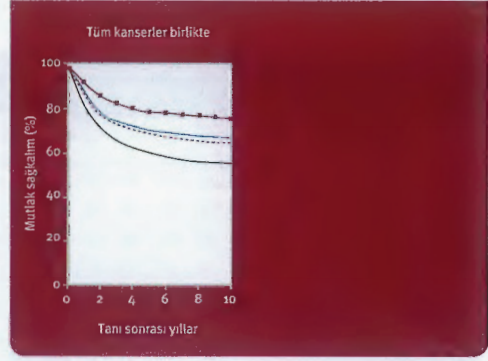
Kendi ülkem olan Almanya'da, ulusal çapta faaliyet gösteren bir kanser kayıt sistemine sahip olacak kadar şanslı değiliz. Fakat Saarland eyaletinin güney batısındaki küçük bir kasabada iyi bir kanser kayıt merkezi bulunmaktadır. Bir süre önce, kanser hastalarının en güncel uzun dönemli sağkalım oranı tahminlerinin orada bulunmasını istedik. Buradaki kanser kayıtlarından 2002'ye kadar olan verilerle o ta-

rihte aynı analiz türlerini gerçekleştirdik. Son 3 yılı kapsayan bir periyot analizi yaptık ve bunu, geçmiş son 10 yılda geleneksel sağkalım analizi yapıldığında ortaya çıkacak durumla karşılaştırdık (Tablo 4). Kötü prognozlu iki kanser olan özofagus kanseri ve mide kanserini ele aldık. Şekil 9, güncel periyot tahminlerimizin oldukça düşük olduğunu göstermekle birlikte 10 yıllık sağkalımın kohort tahminlerine kıyasla %20 daha fazla arttığını da işaret etmektedir.

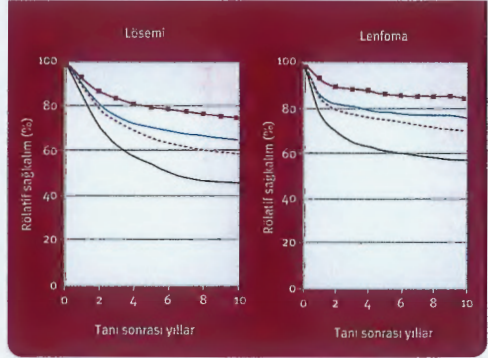
Biraz da kanser sağkalım istatistiklerinde çok kullanışlı olan rölatif sağkalım kavramına kısaca değinmek istiyorum. Kanser kayıt verilerinde çok sık yapılan bir şey, bu hastaların beklenen sağkalımının, genel popülasyonun yaşam tablolarından alınmış beklenen sağkalımına bölümüyle, kanser hastalarında gerçekten gözlemlenen sağkalımın oranı olarak rölatif sağkalımı hesaplamaktır. Yani bu, hepinizin bildiği gibi her zaman güvenilir olmayan ölüm kayıtlarında belirtilen ölüm sebebinin önüne sürmeksizin, kanser nedeniyle beklenen aşırı mortaliteyi elde etme yöntemidir.

Periyot Analizinde Modelleme

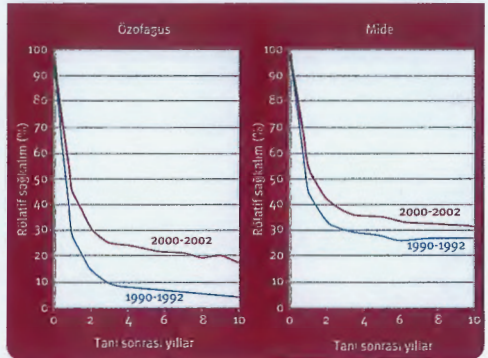
Konuyu buradan, periyot analizi yaklaşımımızı model tabanlı periyot analizine genişlettiğimiz son yıllardaki bazı yeni metodolojik gelişmelere getirmek istiyorum. Bu yöntemle, analizlerimizde daha duyarlı periyot tahminleri elde etmek istedik. Sağkalımda son eğilimleri hesaplamak ve ayrıca kanser hastalarında gelecekteki sağkalım için bir sütun projeksiyonu yapmayı hedefledik. Kısaca özetlemek gerekirse, tekrar Finlandiya Kanser Veritabanından bir örnek aldık. O tarihte Finlandiya'da 5 sağkalım periyodunu hesapladık. Verilerimiz 2002'ye kadardı. Tablo 5'te tekrar aynı veri grubunu görüyorsunuz. Yıllara göre tanılar solda, izlem yılları ve çeşitli hastaların sağkalım analizine katkıları da tablonun daha geniş alanında görülmektedir. Örneğin burada, 1998'den 2002'ye kadarki en son 5 yıl için, en güncel 5 yıllık sağkalım tahminlerini elde etmek amacıyla standart bir periyot analizi kullanabiliriz. Bu örnekte yaptığımız şey, bir adım ilerisini araştırmaktır. Verilerin tümünü bu 5 yıldan öy-



Şekil 7: Tüm çocukluk dönemi kanserlerinin izlemi.



Şekil 8: Çocukluk döneminde lösemi ve lenfomada periyot analizi ve sağkalım.



Şekil 9: Özofagus ve mide kanserlerinde periyot analizi ve sağkalım.

Tablo 4: 2000-2002 arasında Almanya'da periyot analizi.

[illegible]

Tablo 5: Standart ve modellenmiş periyot analizinin karşılaştırılması.

Tanı yılları	İzlem yılları									
	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
1993	1	1/2	2/3	3/4	4/5	5				
1994		1	1/2	2/3	3/4	4/5	5			
1995			1	1/2	2/3	3/4	4/5	5		
1996				1	1/2	2/3	3/4	4/5	5	
1997					1	1/2	2/3	3/4	4/5	5
1998						1	1/2	2/3	3/4	4/5
1999							1	1/2	2/3	3/4
2000								1	1/2	2/3
2001									1	1/2
2002										1

lece almadık. Bu 5 yılı Tablo 5'te görülen 5 dikey kutuya katmanlandırdık ve bunu yaparken de tüm bu kutulardan bilgi aldık. Ardından, ne olduğunu görmek için modeli bu son 5 yıllık periyoda da uyguladık. Örneğin, eğer sağkalım bir kutuda bile yükselirse, bu bilgiyi daha duyarlı tahmin elde etmek için kullanabilirsiniz.

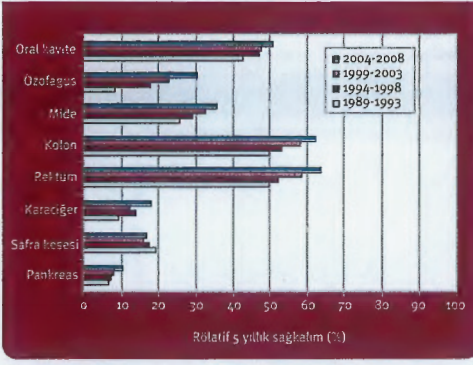
Burada yaptığımız şey, bu modellenmiş yaklaşımı, modellenmiş ve modellenmemiş basit kohort ve periyot analizleriyle kıyaslamaktı. Gördüğümüz şey ise, periyot modellemesiyle elde edilen sağkalım tahminlerinin, modellenmemiş periyot analiziyle hesaplananlara kıyasla daha düşük standart hatalara sahip olduğuydu (Tablo 6). Dahası, modellenmiş periyot analizi, çoğu kanserde, kohort modellemesine nazaran daha iyi bir 5 yıllık sağkalım öngörüsüne sahipti.

Sözünü edeceğim son şey, bu model olgu yaklaşımından bilginin nasıl alınacağı ile ilgilidir. Klinik tıptaki diğer birçok veritabanından bildiğim gibi, verilerin yayınlanması aslında genellikle biraz zaman alır. Model yaklaşımını kullanarak o dönem elimizde olan, örneğin 2003'e kadarki Saarland Kanser Kayıtları verilerinden elde ettiğimiz, daha önceki gözlemleri 1989-1993, 1994-1998, 1999-2003 şeklinde 5 yıllık dönemlere ayırdık. Ve ardından, bu periyotlar için benzer modelleme yaklaşımını gerçekleştirdik ve bu modelleme yaklaşımını, 2004-2008 dönemindeki sağkalımı projekte

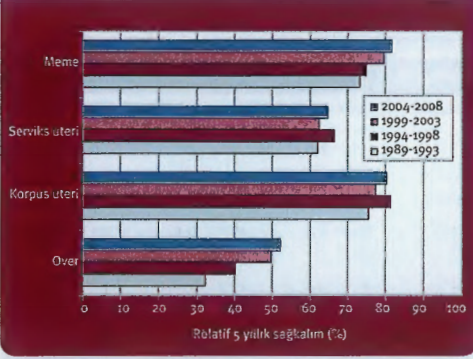
Tablo 6: Standart ve modellenmiş periyot analizinin istatistiksel olarak karşılaştırılması.

Kanserin yerleşimi	Standart periyot analizi				Modellenmiş periyot analizi	
	1998-2002		2002		2002	
	PE	SE	PE	SE	PE	SE
Kolon	59.1	0.8	59.4	1.8	60.2	1.3
Akciğer	10.5	0.8	9.8	0.8	10.2	0.5
Meme	86.4	0.4	89.7	0.8	88.8	0.6
Over	47.0	1.1	49.7	2.6	50.7	1.9
Prostat	83.3	0.6	88.6	1.2	87.8	0.8
Tiroid	89.7	0.9	90.2	2.1	92.4	1.3

	Standart periyot analizi		Modellenmiş periyot analizi			
Kanserin yerleşimi	1998-2002	1998	2002	Değişim	P değeri	
Kolon	59.1	58.0	60.2	+2.2	0.29	
Akciğer	10.5	10.8	10.2	-0.6	0.40	
Meme	86.4	83.5	88.8	+5.3	>0.0001	
Over	47.0	43.3	50.7	+7.4	0.01	
Prostat	83.3	77.1	87.8	10.7	<0.0001	
Tiroid	89.7	86.7	92.4	+5.7	0.03	



Şekil 10: Almanya'da gastrointestinal kanser için sağlık projeksiyonları.



Şekil 11: Almanya'da jinekolojik kanserler için sağlık projeksiyonları.

etmek için kullandık. Çünkü gerçekten de bugün tanı konmuş hastalar için bir tahmin yürütmek, ne beklediğimizi bilmek istedik. "Eğer geçmişte yapsaydık aynı sonucu alır mıydık?" sorusuna yanıt aramak için, daha önce yaptığımız gibi bu yaklaşımı, Finlandiya Kanser Veritabanındaki 50 yıllık verileri kullanarak tekrar değerlendirdik. Sorunun yanıtı evet idi. Şekil 10 ve 11'de, belirli kanser türleri yer almaktadır. Özofagus kanseri için en son projeksiyon %30, mide kanseri için %35'tir. Jinekolojik kanserlerle ilgili analizlerimizin bazıları en etkili gelişmeyi gerçekleştirdi. Over kanserinde son tahminler, 5 yıllık sağkalım için yaklaşık %50'dir ve testis kanseri için %100'e yakındır. Akciğer kanseri için bile bu projeksiyonlar, Almanya'da en az %20'ye ulaşmış durumdadır.

Sonuç

Özetle, daha doğru uzun dönem sağkalım oranı elde etmenin sadece epidemik bir uygulama olmadığını düşünüyorum. Bunun gerçekten bir anlamı vardır çünkü hastalar, sağlık hizmeti sağlayıcıları ve komiteler bu istatistikleri görmektedir. Kanımca bu yöntemin uygulanması, sadece kanserde değil tıbbın tüm diğer alanlarında da klinik prognozun değerlendirilmesini destekleyecek ve periyot analizinin rolü, bu yönde büyük bir katkı sağlayabilecektir. Uzun dönem sağkalımın tahmininde periyot analizi bir standart haline gelmelidir.

KAYNAKLAR

1. Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. *Cancer* 1996; 78: 2004-10.
2. Brenner H, Hakulinen T. Up-to-date long term survival estimates of patients with cancer by period analysis. *J Clin Oncol* 2002; 20: 826-32.
3. Brenner H. Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. *Lancet* 2002; 360: 1131-5.
4. Brenner H, Gefeller O, Hakulinen T. Period analysis for up-to-date cancer survival data: theory, empirical evaluation, computational realization and applications. *Eur J Cancer* 2004; 40: 326-35.
5. Brenner H, Hakulinen T. Up-to-date and precise estimates of cancer patient survival: model based period analysis. *Am J Epidemiol* 2006; 164: 689-96.
6. Brenner H, Hakulinen T. Up-to-date estimates of cancer patient survival even with common latency in cancer registration. *Cancer Epidemiol Biomarkers Prev* 2006; 15: 1727-32.
7. Brenner H, Gondos A, Arndt V. Recent major progress in long-term cancer survival disclosed by modeled period analysis. *J Clin Oncol* 2007; 25: 3274-80.

SINIFLANDIRMA VE TANI KRİTERLERİ

H. Yazıcı

*İstanbul Üniversitesi Cerrahpaşa Tıp
Fakültesi İç Hastalıkları Anabilim Dalı
Romatoloji Bilim Dalı, İstanbul*

Biliyorsunuz, tıp uygulamasında femur başı kırığı, Addison hastalığı veya kolon kanseri örneği elimizde radyoloji, laboratuvar veya histolojik kesin tanı yöntemleri bulunmayan hastalıklarda birçok kez tanı veya sınıflama kriterlerine başvuruyoruz. Bu bağlamda biz romatologların uğraştığı birçok hastalıkta böyle tanı kriterleri var. Kanıta dayalı tıpta güncel “sofistikasyon” düzeyimiz ise böyle kriterlerin hemen hiçbir zaman tanı koymak amacıyla hazırlanmadığı. Deniyor ki söz konusu kriterler gerçekte sınıflama (klasifikasyon) kriterleri ve ancak araştırmalarda hasta gruplarını kıyaslamakta işe yarıyor. Hastalara tek tek tanı koymak ise başka bir süreç ve bunun için de gerçek tanı kriterleri gerek.

Bu bölümün ana amacı, kriterleri tanı ve sınıflama diye ikiye ayırmanın işin esasına indigimizde oldukça yapay olduğu. Aritmetik yöntem bakımından her iki tür kriter de tümüyle aynı mantıkla hazırlanıyor. Sizlere tüm tanı kriterlerinin esasında klasifikasyon kriterleri olarak başlangıçta hazırlandıklarını ancak “test öncesi olasılık” arttıkça kriterlerin tanı koyucu nitelik aldığından söz edeceğim. Yazının sonunda da söz konusu sınıflama/tanı kriterlerinin tıp uygulamasında daha etkili kullanılabilmesi için bir öneride bulunacağım.

Beni bu konuda düşünmeye zorlayan vaskülit sınıflandırmaları oldu. Vaskülit tanısı koymak için iki ana kriter grubu var: 1. *American College of Rheumatology* (ACR) kriterleri, 2. Chapel Hill kriterleri. Ancak vaskülitle ilgilenenler çok iyi biliyor, gerçek hayatta söz konusu kriterler pek de işe yaramıyor. Esasında bu, kabaca 10 yıl evvel yapılmış iki çalışmayla da gösterilmiş. Önce Rao ve arkadaşları, aralarında 51 tane vaskülit tanısı almış 198 hasta arasında ACR kriterlerinin pozitif prediktif değerinin (kriterler pozitif bulunduğunda gerçekten aranan hastalığın bulunması) %17- %29 arasında değiştiğini bildirdi [1]. Bundan iki yıl sonra yapılan diğer bir çalışmayla da Chapel Hill kriterlerinin 27 Wegener’li hastanın ancak 8, mikroskopik poliarteritli 12 hastanın ise 4’üne doğru tanı koyabildiği gösterildi [2].

Tablo 1: Behçet çalışması [3].

- 7 ülkede, 12 merkezden 914 tanı konmuş Behçet hastası alındı.
- Oral ülser anamnezi olmayan 28 hasta gruptan çıkarıldı.
- Behçet hastası gönderen merkezlerden 308 kişilik bir kontrol grubu istendi. Kontrol grubu 211 bağ dokusu ve 97 izole oral ülseri olan hastalardan oluştu.
- Çalışmaya katılan merkezler her Behçet hastası ve kontrol için standart bir form doldurdular. Bu formlar Behçet hastalığının göz tutulması, genital ülser, paterji testi gibi çeşitli belirti veya bulgularının bir dökümünü içeriyordu.
- Her belirti/bulgunun Behçet hastalığında bulunma sıklığının kontrol grubuna kıyasla duyarlılık ve özgüllüğü bulundu. Bu duyarlılık ve özgüllüklerden önce pozitif ve negatif "log olasılık oranları" ve bunların, hastalık ve kontrol gruplarının bütünlüğü de göz önüne alınarak "ağırlıklı değerleri" hesaplandı.
- Çalışma grubunun %60'ı kriterlerin hazırlanması geri kalan %40'ı da kriterlerin geçerliliğinin (*validity*) sınanması için kullanıldı.

Tablo 2: Bayes kuramı [4].

- Test sonrası odds = test öncesi odds x olasılık oranı
- Bu formülde "olasılık oranı" doğrudan kriterlerin duyarlılık ve özgüllüğünden hesaplanır [4].
- Herhangi bir hastanın tanımlanan hastalık kriterlerini doldurması "pozitif olasılık oranını" (=duyarlılık/1-özgüllük), doldurmaması ise "negatif olasılık oranı"nı (=1- duyarlılık/özgüllük) oluşturur.

Tablo 3: Duyarlılığı %90 ve özgüllüğü %95 olan Behçet kriterlerinin işe yararlılığı.

- Kriterler + (+ olasılık oranı):
 $0.90/1 - 0.95 = 18 \times \text{test öncesi odds}$
- Kriterler - (- olasılık oranı):
 $1 - 0.90 / 0.95 = 0.10 \times \text{test öncesi odds}$

Avrupa Romatizmayla Savaş Ligi'nin (EULAR), benim de üyesi olduğum bir vaskülit çalışma grubu var. İşte bu grup geçen yıl vaskülit kriterlerini geliştirmek, daha işe yarar hale getirmek için bir çalışma başlattı. İlk toplantıya ben katılamadım. Ancak tutanaklardan anladım ki toplantıda önemli bir sorun ortaya çıkmış. Şöyle ki: Komite hazırlanmasında öncül olduğum ISBD Behçet Hastalığı tanı kriterlerini [3] kendilerine örnek alıyor, bu kriterlerin şimdiye kadar geçerliliği (*validity*) gösterilmiş yegane vaskülit tanı kriteri olduğunu söylüyor ve nihayet söz konusu kriterlerin hem tanı hem de sınıflamada işe yaradığından söz ediyor. Bütün bunlar oldukça gurur verici ama gerçeği de tam yansıtmıyor.

Behçet kriterleri Lancet'te tanı kriterleri diye yayımlanmıştı [3]. Ancak kısa sürede anlaşıldı ki kriterler Behçet hastalığının seyrek olarak görüldüğü, yani test öncesi olasılığın düşük bulunduğu yörelerde pek işe yaramıyor. O kadar ki yıllar boyu tüm konuşmalarım da ve yayınlarda bu kriterlerden sınıflama kriterleri diye söz ettim. Durumu açıklığa kavuşturmak için, Behçet çalışmasının yöntemini anımsatmak yerinde olacak (Tablo 1).

Yöntemi Tablo 1'de özetlenen Behçet kriterlerinin Behçet hastalığı için duyarlılığı %90, özgüllüğü ise %95 olarak bulundu.

Buradan giderek ünlü Bayes kuramı ışığında [4] tanımladığımız kriterlerin gerçek hayatta ne işe yaradığına bakalım. Tablo 2 tanı koymakta Bayes kuramını özetlemekte.

Tablo 1 ve 2'de açıklananlar ışığında herhangi bir hastanın Behçet kriterlerini doldurup doldurmamasının ne anlama geldiği Tablo 3'te özetlenmekte:

Tablo 3'deki bilgiyi bir örnekle açıklayayım. Diyelim bir hekim arka arkaya 100 hasta görüyor. Bu hastaların sadece 1 tanesinde Behçet var diyelim. Bu örnekte test öncesi olasılık 0.01, test öncesi odds ise 1:99'dur. Şimdi bu gruba Behçet kriterlerini uygulayalım. Bu gruptan herhangi bir hasta kriterleri doldurdu-

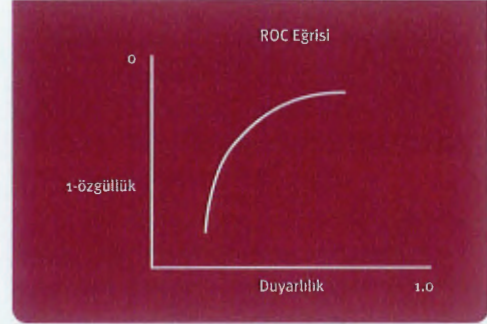
ğunda Behçet’li olma olasılığı $1:99 \times 18 = 18:82$ olur. Diğer bir deyişle olasılık 0.01’den 0.22’ye çıkar. Demek ki tüm gruba Behçet kriterlerini uyguladığımızda 5 kişi pozitif bulunacak, bunlardan 4’ünün Behçet hastalığı olmayacaktır.

Şimdi bir de kriterlerin negatif bulunduğu duruma gelelim. Yine test öncesi (örneğimizde kriter öncesi) grupta Behçet hastalığı bulunma odds’unu $1:99$ alırsak kriterleri doldurmayan bir hasta karşısında bu odds $1:99 \times 0.10 = 1:999$ olur. Olayı olasılık üzerinden ifade edersek kriterler öncesi 0.01 olan olasılık şimdi 0.001 olmuştur. Diğer bir deyişle verdiğimiz örnekte kriterler hastalığa tanı koymaktan çok uzaklaştırmakta işe yaramaktadırlar.

Burada duyarlılık ile özgüllük arasındaki çok önemli bir ilişkiyi vurgulamak gerek. Kural olarak duyarlılık ve özgüllük birbirleriyle ters orantılıdır. Bu ters olan çalışılan örnekte örneğe göre değişir ancak genellikle $y = \log X$ gibi bir eğri gösterir.

Şekil 1’e dikkat edildiğinde x ekseninde duyarlılık, y ekseninde 1-özgüllük olduğu görülür. Böyle bir aritmetik “oyun” aslında, yukarıda vurguladığımız üzere, duyarlılıkla özgüllük arasındaki ters ilişkide, genellikle düz ilişki eğrilerine bakmaya alışık olan biz hekimlerin işini kolaylaştırmak için yapılmış bir düzenlemedir. ROC eğrileri bize önümüzdeki soruna göre hangi duyarlılık ve özgüllük düzeyinde çalışmak istediğimiz yönünde yardımcı olur. Örneğin elimizde kişiyi sıtmadan başarıyla tedavi eden ve de yan etkisi de çok az olan bir ilaç olsun. Bu durumda sıtmayla mücadelede artık duyarlılığı çok yüksek ve dolayısıyla da özgüllüğü görece azalmış, sıtma tanı kriterleri kullanabiliriz. Verilen örnekte, kullanılması düşünülen ilacın toksisitesinin çok az olması önemlidir. Çünkü artmış olan duyarlılık, görece düşük özgüllük birçok kişiye gereksiz yere sıtma tedavisi yapmayı gerektirecektir.

Duyarlılık ve özgüllük kavramlarını tanı kriterleri açısından da gözden geçirmek uygun olacaktır.



Şekil 1: ROC (Receiver operating curve).

a	b
Duyarlılık=0.95 (0.90) Özgüllük = 0.95	Duyarlılık=0.95 Özgüllük = 0.97 (0.95)
■ + olasılık oranı Duyarlılık/1-Özgüllük = 19 (18)	■ + olasılık oranı Duyarlılık/1-Özgüllük = 30 (18)
■ - olasılık oranı 1-Duyarlılık/1-Özgüllük = 0.10 (10)	■ - olasılık oranı 1-Duyarlılık/1-Özgüllük = 0.05 (10)

Şekil 2: Özgüllük (a) ve duyarlılığın (b) artmasının kriterlerin işe yararlılığına etkisi.

Daha önce %90 duyarlı ve %95 özgül Behçet sendromu tanı kriterlerinin, pozitif veya negatif bulunmalarının hastalık sıklığı %1 olan bir toplumda ne anlama geldiğinden söz ettik. Şimdi, hastalık sıklığını sabit tutup söz konusu kriterlerin duyarlılık ve özgüllüklerinin artmasının kriterlerin işe yararlılığına nasıl etki edeceğine bakalım.

Şekil 2a'da özgüllüğü sabit tutup duyarlılığı %90'dan %95'e çıkartıyoruz. Görüldüğü üzere bu değişiklik pozitif olasılık oranını ancak 18:1'den 19:1'e arttırıyor. Bunun yanında negatif olasılık oranını ise 20 kat azaltıyor. Yani duyarlılığın %90'dan %95'e çıkması eldeki kriterlerin bir hastada pozitif bulunmasının değil, bulunmamasının daha çok işe yaradığını gösteriyor. Şimdi bir de Şekil 2b'ye bakalım. Burada ise duyarlılık %90 kalmış, özgüllük ise sadece %95'ten %97'ye çıkıyor. Şekilde görüldüğü üzere bu değişiklik kriterlerin bir hastada pozitif bulunmalarının Behçet tanısı koymakta işe yararlılığını arttırıyor. Öyle ki %95 özgüllükte kriterler test öncesi hastalık odd'unu 18 kat arttırırken bu kez 30 kat arttırıyorlar. Dikkat edilirse özgüllüğü arttırmak negatif olasılık oranına, yani kriterlerin bir hastada negatif bulunması halinde tanıyı uzaklaştırma yeteneğine hiç etki etmiyor. Bütün bunlardan şu sonuçlar çıkar: Yaygın olarak yanlış algılandığının aksine bir tanı kriterleri grubunda özgüllüğün artması hastalık tanısı koymaya, duyarlılığın artması ise söz konusu hastalık tanısını dışlamakta işe yarar.

Özgüllük ve duyarlılıkla ilgili bu vurgulamalardan sonra tanı kriterlerinin işe yararlılığını saptamakta çok önemli ancak önemli olduğu kadar da sık unutulmuş test öncesi olasılığa geelim.

Tablo 4: Japonya ve Kuzey Amerika'daki iki üveit kliniğinde Behçet sendromu sıklığı [6,7].

Japonya ve ABD'de üveit nedenleri (%)		
	Japonya (n=3060)	ABD (n=1237)
İdiyopatik	38.9	34.9
B27	1.5	10.4
Sarkoidoz	13.3	9.6
JRA	0.5	5.6
SLE	1.0	4.8
Behçet	6.2	2.5
HIV	0	2.4

Tanı kriterleri açısından "test öncesi olasılığı" irdelerken söz konusu olasılık tanı kriterlerini kullanmakta ayrı, kullanılan kriterleri hazırlamakta ise ayrı düşünülmesi gerekir. Kriterleri kullanım söz konusu olduğunda test öncesi olasılık, kriterlerin uygulandığı ortamda aranan hastalığın gerçek sıklığıdır. Kriterleri hazırlamakta ise test öncesi olasılık birinci aşamada söz konusu hastalığın önde gelen belirti ve

bulgularının o hastalık, onunla ayırıcı tanıya girebilen hastalıklar ve nihayet sağlıklı kontrollerdeki sıklığıdır. Buradan, biraz önce Behçet hastalığı örneğinde açıkladığımız üzere “ağırlıklı değerler” saptanır. Saptanan değerler listesinden ise, değerlerin büyüklüğüne, yani ayırıcı tanıya işe yararlılığına göre, tanı kriterleri formüle edilir. Ancak böyle oluşturulmuş kriterlerin geçerliliğini saptamakta bu kriterlerin hastalık ayırıcı tanısını yapabilmek yeteneğini sınamak gerekir. İşte bu 2. aşamada, yani geçerlilik saptama (*validity*) sürecinde, test öncesi olasılık yine gerçek hastalık sıklığı olur.

Y. Karaaslan: Aftı olan ve başka hiçbir şikayeti olmayan hastalar var. Aftı olmayan, şikayeti olan hastalarda tanı kriterleri hiç mi işe yaramıyor?

H. Yazıcı: Dediğiniz aynen doğru, bu tanı kriterlerinin yapılmasında aftı olan bağ dokusu hastalıklarından ayırıcı tanı önemliydi. Siz şimdiki Ankara Numune Hastanesinde çeşitli hastalıklarda aftöz stomatit sıklığı nedir diye araştırmanız çok faydalı bir yayın yapmış olursunuz.

Buraya kadar anlattıklarımın tanı kriteri hazırlamakla, tanı kriteri uygulamanın birbirine çok bağlı iki olay olduğu açık. Tanı kriteri koymak bir yerde bir fasit daire olgusu. Yukarıda açıklamaya çalıştığım üzere bir yandan çeşitli hastalık belirti ve bulgularının olasılık oranlarını hesaplarken test öncesi olasılıklardan yararlanıyoruz. Ancak bir kez aranan kriterler hazırlandığında bunların işe yararlılığı da aranan hastalığın o toplumdaki sıklığına bağlı.

Bütün bu nedenler bize belki de çeşitli hastalıklar için artık evrensel tanı kriterlerinden vazgeçmek gerektiğini düşündürmeli. Önerimiz, Behçet sendromu örneği, multisistem, nedeni bilinmeyen ve laboratuvar, histolojik veya radyolojik patognomonik özellikleri olmayan hastalıklarda artık evrensel tanı kriterleri hazırlanmamalı [5]. Böyle bir yaklaşımın iki önemli getirisi olacaktır:

1. Ayırıcı tanıya girecek hastalıkların çeşidi azalacaktır. Örneğin bir gastroenterolog açısından Behçet hastalığının ayırıcı tanısı

hemen tümüyle Behçet ve Crohn arasında, bir nörolog açısından da Behçet ve multipl skleroz arasında olur. Buradan giderek söz konusu yan dallara özgü hazırlanacak kriterlerin özgüllük ve duyarlıklarında %100’e yakın değerlere ulaşmak olasıdır.

2. Yine Behçet hastalığı örneği toplum veya genel hekimlik bazında sıklıkları, yani test öncesi olasılıkları düşük olan hastalıklara yan dal bazında bakıldığında test öncesi olasılıklar çok daha yüksek olabilir. Bunu bir örnekle açıklayayım. Tablo 4’de Kuzey Amerika ve Japonya’daki iki üveit merkezinde Behçet hastalığının diğer üveit çeşitlerine kıyasla görece görülme sıklıkları verilmektedir. Bu tabloya bakınca ilginç olarak sıklığın Kuzey Amerika’daki merkezde %2.5 [6], Japonya’daki merkezde ise %6.5 olduğu [7] anlaşılmaktadır. Aynı sıklığın genel toplum bazında da Japonya’da, Amerika’ya kıyasla en az bin defa daha az olduğu düşünülürse, test öncesi olasılık açısından özetlediğimiz verilerin önemi iyi anlaşılır.

Soru: Toplum bazında düşünürsek, dahiliyecinin, pratisyenin Behçet tanısını koymasını istiyoruz, bu anlamda toplum bazında bizim hastanede tanı koymamıza göre özgünlüğün daha yüksek olmasını isteyeceğiz

H. Yazıcı: Hayır tam tersi. Behçet olabileceğinden kuşkulansın yeter ve uzmanına göndersin. Oraya gelmedim, ama bunun bir de eğitim tarafı var. Bu hastanın aftı var, bir de ishali var, tenesmusu var vs.; göndersin ve Crohn’la Behçet arasındaki ayırımı gastroenterolog yapsın.

Soru: O zaman duyarlılığı arttırınca özgünlükten kaybımız olur,

H. Yazıcı: Hayır, önce duyarlılıkla yakalaya-cağız. Genel kural; duyarlılıkla tara, özgülle tanı koy. Bu arada tanı koymamızdaki amacı da unutmamak gerek. Şimdi siz eğer Behçet veya Crohn dersiniz ciddi ilaç vereceksiniz. Ama siz sadece yüksek duyarlılıkla tanı koyup ben bunlara zezem suyu vereceğim diyorsanız hiç özgül olmayıp, son derece duyarlı davranabilirsiniz.

Topluma artık bunu söylemek gerek. Hekimler olasılıklarla çalışırlar. Hekimler bir yerde mide kanseri tanısını koymayı çok istemez ama mide lenfoması tanısını koymak isterler. Çünkü lenfoma daha kolay tedavi edilebilir. Akciğerdeki leke karşısında önce tüberküloz düşünmek gerek. Sağlıkta reformdan çok söz eder olduk. Ciddi bir reform mutlaka toplumun bilinçlendirilmesiyle başlamalıdır.

KAYNAKLAR

1. Rao JK, Allen NB, Pincus. Limitations of the 1990 American College of Rheumatology classification criteria in the diagnosis of vasculitis. *Ann Intern Med* 1998; 129: 345-52.
2. Sorensen SF, Slot O, Tvede N, et al. A prospective study of vasculitis patients collected in a five year period: evaluation of the Chapel Hill nomenclature. *Ann Rheum Dis* 2000; 59: 478-82.
3. International Study Group for Behçet's Disease. Criteria for diagnosis of Behçet's disease. *Lancet* 1990; 335: 1078-80.
4. Max MB, Lynn J, editors. Symptom Research: Methods and opportunities. URL: <http://symptomresearch.nih.gov/tablecontents.htm>.
5. Yazici H, Seyahi E, Yurdakul S. Behçet's syndrome is not so rare: why do we need to know? *Arthritis Rheum* 2008; 58: 3640-3.
6. Goto H, Mochizuki M, Yamaki K, et al. Epidemiological survey of intraocular inflammation in Japan. *Jpn J Ophthalmol* 2007; 51: 41-44.
7. Rodriguez A, Calonge M, Pedroza-Serez M, et al. Referral patterns of uveitis in a tertiary care center. *Arch Ophthalmol* 1996; 114: 593-9.

BİLGİLENDİRİLMİŞ OLUR (INFORMED CONSENT)*

R. Smith

*British Medical Journal, Eski Editörü,
İngiltere*

***Hazırlayan:** Konuşma özeti Hasan Yazıcı tarafından, ses kayıtlarından hazırlanmıştır.

Dr. Smith konuşmasında bilgilendirilmiş oluru, gelişen ve halen de gelişmekte olan bir süreç olması dolayısıyla tarihsel boyutta ve öncelikle araştırma bağlamında ele alarak özellikle Helsinki deklarasyonuna değinmiş, kendisinin bir dönem editör olarak görev yaptığı *British Medical Journal*'daki (BMJ) uygulama örneklerini dile getirmiştir. Bunun yanı sıra klinik pratikte ve eğitimde bilgilendirilmiş olur üzerine çeşitli soru ve düşünceleri değerlendirmiştir.

Tarihsel Bakış

Dr. Smith konuşmasına, 20. yüzyıldan itibaren gelişen süreci, İkinci Dünya Savaşı öncesinde, hastalardan izin almaksızın hastalar üzerinde çalışmalar yapmanın rutin bir uygulama olduğunu anımsatarak başlamıştır. O dönemde doktorların ve araştırmacıların en iyisini bildikleri, hastaların ise sadece hasta kabul edildikleri görüşünün yaygınlığını dile getirmiştir. II. Dünya Savaşı sırasında Nazi doktorlar tarafından gerçekleştirilen, bilinen korkunç olayların meydana geldiğini anımsatarak, Nürnberg duruşmasına yanıt olarak, içeriğinde insan araştırmaları konusunda 10 ana prensibin bulunduğu Nürnberg Kodu'nun ortaya çıktığını ve bu prensiplerinden ikisinin bilgilendirilmiş olur olduğunu belirtmiştir. Burada ilk prensip bilgilendirilmiş olurun çok önemli bir koşul olduğudur. Gönüllülük temeldir. Onay çok yoğun bir şekilde bilgi içermeli, araştırma yapmadan önce alınmalı, şahsın onay verebilecek yasal kapasitede olması gerekmektedir. Onay veremeyecek derecede hasta olanlarla araştırma yapılmaması ve ayrıca onayın vekaleten alınması söz konusudur. Konuşmacıya göre bu metin özetle çok güçlü bir onay özelliği taşımaktadır.

Helsinki Deklarasyonu

Nürnberg Kodu sonrasında durumun, Dünya Tıp Birliği'nden gelen ve halen de devam eden Helsinki Deklarasyonu ile sürdüğünü belirten Dr. Smith bu süreci ayrıntılı bir biçimde

aktararak gelişimi tüm boyutlarıyla gözler önüne sermiştir. İlk Helsinki Deklarasyonu Haziran 1964 tarihli'dir. Sonrasında 5 revizyon ve 2 açıklama yapılmış ve içerik zenginleştirilmiştir. Helsinki Deklarasyonu öncesinde yine Dünya Tıp Birliği tarafından, 1954'te yayınlanan İnsan Araştırmaları Ortakgörüş Bildirgesinde insan deneylerinde vekaleten onay alınabilir denerek Nürnberg Kodu'nda değişikliğe gidilmiştir. Buna göre araştırmalar, onay veremeyecek kadar hasta veya mental açıdan yetersiz kişiler, örneğin çocuklar üzerinde de kurgulanabilir hale gelmiştir. Sonrasında gelen Helsinki Deklarasyonu'nda da 'hasta psikolojisinin uygun olduğu her durumda, hastadan öncesinde onay alınmalıdır' ifadesi yer almaktadır. Burada vekalet yoluyla onay da söz konusu olmuştur.

Sonrasında gelen 1975 deklarasyonunda bilgilendirilmiş olurun temel olduğu, vekalet onayının ise yaşça küçükler, çocuklar ve fiziksel veya mental olarak kapasitesi yetmeyen hastalar gibi durumlarla sınırlı olabileceği belirtilmiştir. Burada ilk kez değinilen önemli olan bir nokta, onay almadan ilerlenmek istendiğinde, çalışma için bir etik komiteden onay alınması ve her araştırma protokolünün, dikkate alınması, tartışılması ve rehberlik edilmesi için özel olarak atanmış bir tarafsız komiteye sunulması gereklilidir. Böylece ilk etik komiteler oluşmuştur.

Sonrasında, 80'lerde, gerçekte tedavisi olan durumlarda plasebo kullanımı, özellikle de gelişmekte olan ülkelerde HIV'in anne-badan çocuğa geçişini önlemek için yapılan plasebo kollu çalışmalar üzerine tartışmalar olmuştur. Etik olarak kabul edilemez bulunan bu konuda tartışmalar hala sürmektedir.

1994'te Helsinki Deklarasyonu'nun dördüncü revizyonu, tedavisi olan durumlarda, plasebo kontrollü çalışmaların kabul edilemez olduğunu ilan etmiştir. Tartışmalar sonrasında, 2000'de tüm belge yeniden düzenlenmiş ve araştırmanın değeri üzerinde çok duran bambaşka bir belge ortaya çıkmıştır. Burada tedavi edici ve etmeyici araştırma arasında ciddi ayırım yapılmıştır. Plasebo üzerinde de durularak

kanıta dayalı bir tedavinin varlığında plasebo vermenin kabul edilemez olduğu belirtilmiştir. Sonra 2002'de, orta bir yol bulunmuş ve plaseboya ciddi veya geri dönüşümü olmayan tehlikeli durum yokluğunda, minör durumlarda, metodolojik gereklilik olduğunda izin verilmiştir.

Konuşmacı, 1950'lere tekrar dönerek, o yıllarda tıbbi araştırmaların sponsorluğunun temel karar organı olan İngiliz Tıbbi Araştırma Komisyonu Sekreterliği'nin şu saptamasını anımsatmıştır: "Bir tıp dalına kendilerini adanmış çok az sayıda tecrübeli araştırmacı, belirli bir araştırmanın yapılması konusunda öğüt verebilecek ve fikir yürütebilecek kapasitededir". Konuşmacıya göre bu saptama, hastaların bu konularda karar vermek için yeterli olmadığı ve gerçekte sadece biz uzmanların neyin kabul edilebilir olduğuna karar verebileceği şeklinde bir ifadedir. Burada esas söylenmek istenen, bir yerde, bilgilendirilmiş olurun sadece Naziler için söz konusu olduğu, tıp bilimcileri için geçerli olmadığıdır. Konuşmacı böylece Nürnberg Kodu'nun ironik bir şekilde Nazilerin garipliklerinden doğduğunu belirterek, insanların bunun çok garip bir uç nokta, korkunç bir durum olduğunu "Nazilere uygulanan bize uygulanmaz, çünkü biz çok mantıklı insanlarız ve hastalar için iyi olanı yaparız" sözleriyle ifade ettiklerini vurgulamıştır. Dr. Smith'e göre insanlar hala böyle düşünmekte ama bu düşüncelerini artık yazıya dökmemektedirler.

Konuşmacı Dr. Smith, Moise Pappert tarafından, 1962'de, "İnsan göçmenleri" başlığıyla yayınlanan, günümüz standartlarında yapılmış ve onay alınmamasına karşın BMJ gibi büyük dergilerde yayınlanmış, çocuklar, mahkumlar ve zihinsel özürliülerde yapılmış deneyleri içeren ve birçoğunda kardiyak kateterizasyon gibi hastaya hiçbir yararı dokunmayacak olan manipülasyon yöntemleri olan zoo çalışmayı irdelleyen bir makaleyi örnek vermiştir. Makalede sözü edilen çalışmalar sadece bilginin gelişimi için yapılmışlardı. Dr. Smith, yazarının bundan altı yıl sonra kitap haline getirebildiğini ifade ettiği bu yapının hala ilgi çekici olduğunu belirtmiştir.

Konuşmacının vermiş olduğu diğer bir örnek, hemen hemen aynı zamanlarda Harvard'da anestezi profesörü olan Henry Nose Beacha tarafından *New England Journal of Medicine*'de (NEJM) yayınlanan "Etik ve insan deneyleri" başlıklı makaledir. 6 tanesi NEJM'de yayınlanmış, etik olmayan 22 deney olgusunun sunulduğu bu makale Amerika'da çok kısa bir zamanda büyük tepkiye yol açmış ve ardından hızla kurumlar bazında araştırma eleştirileri kurulları oluşmaya başlamıştır. Ancak güncel anlamda etik kurulların oluşması bir 20 yıl daha almıştır.

Konuşmacı, 90'lı yılların ikinci yarısında bilgilendirilmiş olurun gelişmesini de çeşitli örneklerle açıklamıştır. 102 önde gelen İngilizce derginin tarandığı *Journal of American Medical Association*'da (JAMA) yayınlanmış bir çalışmada, bunların sadece yarısının yayınlar için etik komite onayının gerekliliğini koşul olarak istediklerini belirtmiştir. Yayınlanmış 53 yayınının incelendiği bir başka çalışmada, yayınların yarısında bilgilendirilmiş olur alınıp alınmadığının belirtilmemiş, 6 tanesinde, ilgili dergiler gerektirdiği halde bilgilendirilmiş olur veya etik komite onayından hiç söz edilmemiş olduğunun altını çizmiştir. Konuşmacının sunduğu son örnekte ise dört geriyatri dergisinde yayınlanmış 586 girişimsel çalışma incelemiştir. Bunların sadece yarısından biraz fazlasında bilgilendirilmiş olur saptanmış ve sadece %40'ında etik komite veya kurumsal onay komitesi (*International Review Board*) onayı alınmıştır.

BMJ Deneyimleri

Konuşmacı BMJ yıllarına ait bilgilendirilmiş olur deneyimini de, o günlerdeki kavram karışmasının da altını çizerek şöyle özetlemiştir:

"1997'de bilgilendirilmiş oluru olmayan 2 çalışma yayınladık. Her iki çalışma da Helsinki deklarasyonu'na uymaktaydı. Helsinki Deklarasyonu'nun izin verdiği ayrıcalıklı durumlara uyduğu düşünüldüğünden etik komite tarafından kabul edilmişlerdi.

İlk çalışma, inmeli ailelere yardımcı olan sağlık çalışanlarının hastalarda ve ailelerde bir

fark yaratıp yaratmadığının araştırıldığı rando-mize bir çalışma idi. Yazarlar inmeli ailelerden iki nedenle bilgilendirilmiş olur almayı uygun bulmamışlardı: Birinci neden insanların bu durumu bilmeleri halinde, 'evet biz bu çalışmayı çok beğeniyoruz çok muhteşem' diyerek sonuçları tehlikeye atacağıydı, çünkü sonuç değerlendirmeleri oldukça sübjektifti. Diğer neden olarak inme hastaları ile çalışan sağlık personelinin bu durumu daha iyi veya daha kötü yapma olasılığı pek akla yakın görünmediğinden, çok risk yoktu. Ama bu çalışmayı yayınlarken hepimiz çok yoğun bir tartışmanın içine girdik ve sonunda bu durumu dünya ile paylaşmaya karar verdik. Glasgow'dan bir etik profesörü bu durumu kabul edilmez buldu. Nedenlerimizin evrensel doğrulardan sapmamız için yetersiz olduğu görüşündeydi. Yerleşik kurallara saygılı olmalıydık.

BMJ'de yayınlanan bir başka çalışma ise şöyleydi: HIV'li hastalar yoğun bakım ünitelerine kabul edilmiyordu. Bu durum daha kötü sonuçlar doğurabilirdi. Hastalar bu çalışmaya ve HIV için kan alınmasına onay vermediler. Yazar onayın hastaların çok hasta olmaları nedeniyle alınmadığını, çalışmanın ise çok önemli olduğunu, bu nedenle hastaların bilgilendirilmiş olur hakkının çiğnenebileceğini savundu. Etik komite bu çalışmayı reddedecekti fakat tüm girişimlerin yoğun bakımda uygulanacağı ve hastalara çok az zarar verecek olmaları nedeniyle sonunda kabul etti. İlk yorum bir nörologdan geldi. Etik komitenin hatalı olduğunu ve bizlerin de BMJ olarak yanlış yaptığımızı söyledi. Çünkü çalışmanın standartları çok düşüktü ve gelişmiş bir ülkede asla kabul edilmezdi. O halde neden gelişmekte olan bir ülkede kabul edilmişti?..

Karşı durumu savunanlar da vardı: İnsan olarak yayın yapma hakkımız var ve eğer yayın yapmak doğruysa, bazı durumlarda yayın bilgilendirilmiş olur olmaksızın da yapılır."

Dr. Smith sonuçta tartışmayı BMJ okurlarına götürdüklerini anlatmış ve okurlardan gelen mektuplarda belirtilen görüşlerin böyle bir durumda bilgilendirilmiş olur gereği konusunda tam bir ikiye bölünme gösterdiğini söylemiştir.

Hatta bununla ilgili bir de konferans düzenlediklerini belirtmiştir. Dr. Smith'e göre tartışma devam etmektedir ve olasılıkla sonsuza dek devam edecektir.

Klinik Pratikte ve Eğitimde Bilgilendirilmiş Olur

Dr. Smith'e göre, araştırmalara oranla, tıbbi uygulamada etik standartlar çok daha düşüktür. Örneğin cerrahlar hangi sıklıkta değişik tedavi yöntemlerinin artı ve eksi yönlerini hastalara açıkça anlatmaktadırlar? Hekimlerin tıbbi pratiğin belirsizliklerine ne kadar hakim oldukları ve bunun ne kadarını hastalarıyla paylaştıkları önemli bir sorudur. Hakem değerlendirmesinde olduğu gibi burada da çifte standart çok yaygındır.

Dr. Smith'e göre, hastaneye yatış sırasında bir standart form olması gerekir. Ayrıca tıp öğrencilerinin de hasta kayıtlarına erişebilmesi için hastanın onayı olmalıdır. Bu görüşlerin yanı sıra Dr. Smith, bilgilendirilmiş olur konusundaki ciddiyetin varabileceği noktalara dikkat çekerek, bunun tıp eğitimi de olumsuz etkileme potansiyelinden söz etmiştir.

Konuşmacı son olarak, konu ile ilgili sorunları, yanıtlarını kendisinin de bilmediğini belirttiği bazı sorular yönelterek dile getirmiştir: "Araştırmalardaki bilgilendirilmiş olur orijinal Nürnberg Kodu'nda olduğu gibi, hiç çıkışı olmayan mutlak bir şey mi olmalıdır? Pratik ve eğitimdeki standart bilgilendirilmiş olur, araştırma düzeyine çıkarılmalı mıdır? Bir hekim olarak belirsizlikleri hasta ile hangi oranlarda paylaşmalısınız? Olasıdır ki bu durum kültür farkına göre, yaşa göre ve disipline göre değişecektir; güven, belirsizliği paylaşmaktan mı geçiyor?" Ve konuşmasının sonunu esprili bir şekilde, tüm bu soruları her yönüyle tartışabilmek için dinleyicileri, kurduğu ve "Hayat Zordur ve Elimizde Çözüm Yoktur" adını verdiği politik partiye katılmaya davet ederek bağlamıştır: "Manifestomuz şu: Ölüm kaçınılmazdır. Hazırlıklı olun. Bilim ile rekabeti görmezlikten gelmemeliyiz; nadiren anlayabiliriz, her zaman yanlış anlıyoruz. Problem olarak tanımladıkla-

rımızın büyük bir kısmını toplum kendisi yaratmaktadır. Siz onlar için bir şeyler yaparken tüm çözümlerin aynı zamanda geçici bir süre sorunun bir parçası olduğunu ve başka sorunlara yol açacağını bilerek yaparsınız. Hayat çelişkilerle doludur, insanoğlu sadece kusurların toplamıdır, olaylar ters gittiğinde birilerini suçlamakla bir yere varamayız ve hayatta sizin fel-safenizde hayal edilebilenin ötesinde çok daha fazla şey vardır."

EŞ DEĞERLENDİRME (PEER REVIEW)*

R. Smith

*British Medical Journal, Eski Editörü,
İngiltere*

***Hazırlayanlar:** Konuşma özeti Hasan Yazıcı ve Salih Murat Akkın tarafından, ses kayıtlarından hazırlanmıştır.

Bilimsel dergilere gönderilen makalelerin inceleme sürecinin önemli bir ögesi olan eş değerlendirme (peer review) üzerinde anlaşılmış, kesin bir tanımı yoktur. Bu aslında pratik önemi olan bir sorundur. Bir derginin Washington'daki Ulusal Tıp Kütüphanesinin (*National Library of Medicine*) oluşturduğu *Index Medicus*'ta dizinlenebilmesi için bir peer review dergisi olması gerekliliği vardır.

Dr. Smith'e göre eş değerlendirme bir makale taslağının, bir proje başvurusunun ya da herhangi bir bilimsel yazının, metnin yayınlanmasında ya da projenin kabul edilmesinde nihai karar verici olmayan bir meslektaş (eş=peer) hakem tarafından eleştirel incelenmesidir. Burada bir dizi soru gündeme gelmektedir. Hakem konumundaki "meslektaş" aynı uzmanlık alanından veya tam olarak aynı araştırma alt alanından bir kişi mi olmalıdır? Kaç hakem bulunmalıdır? İnceleme ne kadar eleştirel olmalıdır? Seçilen hakem, uygulamada hemen hiç yapılmadığı gibi makalenin gönderme yaptığı kaynakların doğruluğunu da kontrol etmeli midir? Süreç, hakemin yazarların kimliğini bilmediği bir biçimde yani kör mü yürümelidir? Hakemin kimliği bilinmeyip yazarların kimlikleri mi açık olmalıdır? Nihai kararı kim vermelidir (hakem mi, editör mü, yoksa yayın kurulu mu)?

Hakem değerlendirmesinde standart bir sistem yoktur. Hemen her dergi biraz değişik bir sistem uygular. Burada unutulmaması gereken, çoğu durumda tek bir hakemin, çok sayıda hakem kadar etkili olabileceğidir. Bunun yanında gerekli inat ve sabır gösterilirse herhangi bir yazı eninde sonunda yayımlanacağı bir yer bulmaktadır.

Günümüzde çoğu hakem sisteminin örnek aldığı *British Medical Journal*'da (BMJ) 2004'te, R. Smith'in editörlüğü sırasında uygulanan sistem şöyledir: O dönem dergiye yılda yaklaşık 6.000 bilimsel yazı başvurusu olmakta, değerlendirme sonucunda ise yazıların sadece %5'i yayınlanmaktadır. Yaklaşık 1.000 makale doğrudan on kadar editörün oluşturduğu bir kurulca incelenip reddedilmektedir. Doğrudan reddetme nedenleri şöyle sıralanabilir:

- a. Yazının orijinal olmaması
- b. Aşırı uzmanlaşmış okuyucuya yönelik olması
- c. Dikkatsiz ve özensiz yazılmış olması, anlaşılır olmaması
- d. Prevalans çalışmaları
- e. Maliyet hesaplarını araştıran çalışmalar
- f. Temel bilim çalışmaları
- g. Hayvan çalışmaları
- h. Randomize kontrollü olmayan çalışmalar
- i. Randomize olup da düşük güç taşıyan tedavi çalışmaları
- j. Güncel geçerli tedavisi olan hastalıklarda yeni bir tedavi yöntemini salt plaseboya karşı sınavan çalışmalar
- k. Yayınlanmış çalışmalara dayanan sistematik derlemeler
- l. Olgu sunumları

Dr. Smith'in açıklamalarına göre gelen yazıların yaklaşık 2.500'lik ikinci bir grubu da 2 editör tarafından ayrı yapılan bir incelenme sonucu reddedilirken geri kalan 2.500 yazı ise başlangıçta bir hakeme gönderilmektedir. Derginin, sürekli denetlenen yaklaşık 5.000 kişilik bir hakem veritabanı bulunmaktadır. Hakemleri 4 puanlı bir sistem üzerinden denetleyen bir skorlama sistemi vardır. Buna göre, nitelsiz veya çok yavaş değerlendirme yapan hakemler bir daha kullanılmamaktadır. Konuşmacı burada, genellikle 40 yaşın altında ve tercihen epidemiyoloji ve istatistik bilenlerin daha iyi hakemlik yaptığını vurgulamıştır.

Orijinalliği, metodolojisi, sunumu, geliştirilip geliştirilemeyeceği, etik olup olmadığı vb. hakkında sorgulanarak yaklaşık 1.000 tanesi daha reddedilen yazılardan elde böylece 1.500 yayına aday makale kalmaktadır. Kalan bu yazılar da editörler arasında sunulup tartışılır ve bazıları özel bir kurulun görüşünün alınması için bir süre "askıya çıkarılır". Bu "askı" kurulu bir istatistik uzmanı ve bir editörün yanı sıra 2 pratisyen hekim ve birçoğu araştırmacı olan uzmanlardan oluşmaktadır. Sonuçta 300 makale yayın için kabul edilir. Bu kabul hemen

her zaman gerekli revizyonları yapmak koşuluyla olmaktadır.

Bu süreçte yazarlarca yapılan itirazlar da daima göz önünde bulundurulur. Dr. Smith'in açıkladığı uygulamada itirazların yaklaşık %20 gibi oldukça yüksek bir oranı kabul edilmekte, ikinci itirazlar ise göz önüne alınmamaktadır.

Editörü olduğu yıllarda BMJ'deki eş değerlendirme sürecinin böyle anlatan Dr. Smith konuşmasının bundan sonraki bölümünde süreç hakkında genel görüşlerini belirtmiştir:

- a. Eş değerlendirmede ana sorun bir yazının veya proje başvurusunun neden bir diğerinden daha üstün olduğuna ilişkin elde altın bir standart olmamasıdır.
- b. Geçmişte editörler yazıları alınca ya "bunu istemiyorum, bunu beğendim" der ya da her bir yazıyı 2 ayrı hakeme gönderirdi. Eğer iki hakem de evet derse yazı yayınlanır. Eğer ikisi de hayır derse reddedilir; ama eğer biri evet, biri hayır derse o zaman editör yazıyı üçüncü bir hakeme gönderir ve o ne derse onu yapardı. Dr. Smith'e göre bu, potansiyel olarak, neyin yayınlanacağına karar vermek için yazı tura atmaya benzer. Aslında eş hakem değerlendirmesinin iyi bir süreç olduğu hakkında yaygın bir kanı vardır. Ancak bu doğru değildir.
- c. Eş hakem değerlendirmesi sahtekarlıkları saptamaya yarar mı? Konuşmacıya göre tamamen yararsızdır. Sistem güven üzerine kuruludur. Bir dergiye gönderilip kabul edilmiş bir makalenin aynı dergiye birkaç yıl sonra tekrar gönderildiğinde daha önce yayınlanmış olduğunun farkına varılamayabileceği örnekler geçmişte yaşanmıştır.
- d. Öte yandan eş hakem değerlendirmesi sistemi bilimin tam merkezidir. Bilimsel yazı yayınlanacak mıdır? Finanse edilmesi onaylanacak mıdır? Bu kişi bir Nobel ödülü alacak mıdır? Eş hakem değerlendirmesi tüm bunların merkezindedir ve kuralları hala tam olarak yerleşmemiş bir süreçtir. Bilimsel anlamda kanıt temelli bir süreç değildir.

- e. Sürecin işleyişinin çok yavaş olduğu kesindir; günümüzde dahi bir makalenin yayınlanmasının bir yıldan fazla sürdüğü dergiler bulunmaktadır. İşlemler çok yavaştır ve pahalıdır. Çoğunlukla hakemlere ödeme yapılmaz, yine de çoğu hakem, rutin işlerini yaptıklarında saat başına yaklaşık 200 sterlin kazanabilecek nitelikte kişilerdir ancak yine de hakemlik için saatlerini ücret getirmeyen bir iş için harcarlar. Konuşmacıya göre eş hakem değerlendirmesi oldukça savurgan bir süreçtir.
- f. Bir diğer sorun yazarlarla ilgili ün, saygınlık, taşralılık, cinsiyet gibi önyargılardır.
- g. Eş hakem değerlendirmesinin özgün düşünce ve yenilik karşıtı olduğunu gösteren veriler de vardır. Bu yüzden harika bir fikrin reddedilme riski söz konusudur. Bunlar sistemin sevmediği yenilikçi şeylerdir. Ne yazık ki süreç kötü niyete açıktır; fikirler çalınabilir, diğerlerinin çalışmaları acımasızca derlenebilir. Burada Dramant'ın şu olağanüstü cümlesi önemlidir: "Eğer eş hakem değerlendirmesi bir uyuşturucuysa, asla piyasaya düşmeyecektir çünkü yararına ilişkin hiçbir kanıt yoktur ama olumsuzluğuna dair birçok kanıt sahibiz." Bu nedenle eş değerlendirme bilimin merkezi olan bilim üzerine değil güven üzerine kuruludur.
- h. Eş değerlendirme ne gibi yollarla geliştirilebilir? Makale değerlendirmesinin kalitesini ölçmek oldukça zordur, hatta olanaksızdır. Yine de eş hakem değerlendirmesini geliştirme amacıyla BMJ'de birçok farklı yöntem denenmiştir. Bunlardan bir tanesi, yazarların kimliğini hakemlerden gizlemektir. Fakat bu kör değerlendirme yazarları bilerek yapılan hakem değerlendirmesinden daha iyi sonuçlar vermemiştir. Bir diğer bir strateji de yazarların, hakemlerin kimliğini bilmesine izin vermek ve açık eş değerlendirme ile devam etmektir. Bu, hiç kimsenin görünmeyen bir yargıç tarafından yargılanmaması gerektiğinden giderek, gerçekten etik sebepler için uygulanmış ve hakemleri tanımlayarak eş değerlendirmenin güven ve sorumluluğu artırılmıştır. Bir başka adım

da, okurların yayınlanan makalelerle ilgili bütün hakem yorumlarını görmesine izin vermemiştir. Bu yöntemi sınavan çalışmalarda hakem yorumlarının kalitesinde düşüş saptanmamıştır.

- i. Eş değerlendirme hakemleri genellikle bu konuda eğitilmiş değildir. Ancak yapılan çalışmalar hakemlerin eğitimden hoşlandıklarını ortaya koymuştur.

Dr. Smith konuşmasının bunda sonraki bölümünde ileriye dönük şu görüşleri de dile getirmiştir:

"Çoğu doktor gerçekten de orijinal araştırmaları okumaz, çünkü araştırma okumak çoğu için bir anlam taşımaz. Birçok hekim araştırmaları değerlendirecek eğitime sahip değildir. Orijinal bir çalışmayı baştan sona okumak uzun bir süre alır; bu yüzden orijinal çalışmaların tümünün büyük bir veritabanında, elektronik olarak herkesin erişebileceği biçimde bulunması mantıklıdır. Videolar gibi tamamlayıcı materyaller ve uygun yazılımlarla birlikte tüm geçmiş veriler, dünya verilerinin tümü erişilebilir kılınmalıdır. Eş değerlendirme tamamen açık bir süreç olmalı, böylece çalışma bitirildiği an yayınlanabilmelidir. Makale önce askıya çıkar, ardından hakkında tartışma başlar, böylece bir kara kutu olmak yerine bilimsel bir söyleve dönüşür ve eş değerlendirme, azınlığın işi olmaktan ziyade Wikipedia gibi, birçoklarının ortak uğraşı olur. Ve şimdi olduğu gibi bazı makaleler önemli olarak ön plana çıkar fakat çoğu makale silinir ve kaybolur gider. Bu düşüncenin devrim niteliğinde olduğu düşünülebilir ancak 17. yüzyılda bilimin yayılmasında, üretilen bilimin, bilimsel bir toplantıda sunulması ve ardından her yönüyle tartışılması yoluyla gerçekleştiği anımsanırsa bu aslında bir tür eskiye dönüşür."

İTERAKTİF OTURUM I

M. İnanç, B. Çakır

İ.Ü. İstanbul Tıp Fakültesi
İç Hastalıkları Anabilim Dalı
Romatoloji Bilim Dalı, İstanbul

Hacettepe Üniversitesi Tıp Fakültesi
Halk Sağlığı Anabilim Dalı, Ankara

İteraktif Makale Tartışması (I)

M. İnanç: Sistemik lupus eritematosus (SLE) çeşitli organ sistemlerini etkileyen, alevlenme ve remisyon dönemleri olan heterojen bir hastalık olduğundan randomize kontrollü çalışmalar (RKÇ) açısından güçlükler göstermektedir. SLE’de son 40 yılda hiçbir yeni tedavi protokolü onaylanmamıştır ve az sayıda *outcome* (sonuç) ölçütü başarılı bulunmaktadır. SLE klinik çalışmalarında *Outcome Measures in Rheumatology Clinical Trials* (OMERACT) tarafından tavsiye edilen alanlar; hastalık aktivitesi, organ hasarı, sağlıkla ilgili yaşam kalitesi, ters etkiler ve sağlık ekonomisi analizleridir. SLE’de sonuç ölçütleri çoğunlukla gözlemsel, uzunlamasına verilerden elde edilmiştir.

SLE’de geliştirilmiş 5 aktivite indeksi bulunmaktadır. Bunlar “*British Isles Lupus Assessment Group Index* (BILAG), *European Consensus Lupus Activity Measurement Index* (EC-LAM), *Lupus Activity Index* (LAI), *Systemic Lupus Activity Measure* (SLAM) ve *Systemic Lupus Erythematosus Activity Index*’ dir (SLEDAI). Bunların bazılarının güncellemeleri de bulunmaktadır. Değişikliğe duyarlı ve aralarında da korelasyon gösterdiği saptanmış olan bu indeksler arasında önemli farklar vardır. Hangi indeksin RKÇ için daha uygun olduğu bilinmemekte ve bu indekslerin kullanılması eğitim gerektirmektedir [1].

SLICC/ACR hasar indeksi en az 6 aydır var olan, geri dönüşsüz organ hasarını ölçmektedir. Hasar indeksinin özellikle hasta alımında önemli olabileceği düşünülmektedir. SLE’de sağlıkla ilgili yaşam kalitesi ölçümü için SF-36 önerilmektedir. Skorların hastalık aktivitesi ve hasarla korelasyon gösterdiği bildirilmiştir. SLE için özgül sorgulama enstrümanları geliştirilmeye çalışılmaktadır. SLE’de hasta ve hekimin hastalıkla ilgili global değerlendirmelerinin farklı olabildiği bilinmektedir.

Son yıllarda SLE’li hastalarda yapılan RKÇ’lerden öğrendiklerimiz şöyle sıralanabilir:

1. Anti-CD40L monoklonal antikorlarla yapılan çalışmada (Faz 2) aktif ilaç grubu ile plasebo grubu arasında alevlenme açısından fark

bulunmamış ve RKÇ'ler tromboembolik komplikasyonların gözlenmesi nedeniyle durdurulmuştur. Tedavinin B-hücre altgruplarında belirgin biyolojik etkisi gözlenmiştir.

2. Anti-dsDNA antikorlarının düzeyinin, hastalıkta alevlenmeleri önceden belirlediği gösterilmiştir.
3. Seçilmiş hasta gruplarında alevlenme oranları belirlenmiştir, fakat alevlenmelerin önlendiğinin gösterilebilmesi için gerekli çalışma süresi kesin olarak bilinmemektedir. Yeni kontrollü çalışmalarda anti-CD20, anti-CD22 ve anti-BLYS gibi B hücre aktivitesini etkileyen monoklonal antikorların biyolojik etkileri de incelenmektedir.

Sonuçta, SLE'de RKÇ sonuçlarının değerlendirilmesinde güçlükler vardır. SLE'li hastalar gerek hastalık aktivitesi ve organ hasarı gerekse demografik anlamda heterojendir. Kısa süreli hastalık aktivitesi hastalık şiddetini yansıtmamaktadır. Kontrollü çalışmalar gereklidir fakat hastaların çalışma ilacına ek olarak aldıkları bazal tedaviler çalışma sonuçlarını etkilemektedir. RKÇ'ler SLE'de yeni ilaçların onaylanmasını henüz sağlayamamışlardır; bununla birlikte bu çalışmalardan önemli dersler alınmıştır.

İnceleyeceğimiz çalışmada [2] muhtemelen yeterli hasta sayısı sağlamak için bir merkezden yaşı 12-18 arasındaki pediatrik hastaları da katıyorlar. Laboratuvar incelemelerinin merkezi olarak yapılması çalışmanın bir artısı. Burada ELISA yöntemini seçiyorlar. Bu anti-DNA'lar belli bir grup lupus hastası için çok önemli antikorlar. ELISA çok spesifik kabul edilmeyen bir yöntem, ama bu çalışmada radyasyon gerektirmemesi, kolay yapılması ve kantitatif olması nedeniyle aktiviteyle ilişkisinin kolay anlaşılması dolayısıyla tercih edilmiş.

Birkaç kere göstereceğimiz bu akış şemasından başlayalım. Prednizon grubuna 180 hasta alıyorlar; bunların 154 tanesi gözlem dönemine giriyor. Araştırmacıların çalışmayı istedikleri grup serolojik aktif, klinik olarak iyi olan grup ve 40 kişi civarında. Bunları ilaç grubu ve plasebo grubu olarak ikiye ayırıyorlar, geri kalanını da non-randomize olarak izliyor-

lar. Bu hastaların bir bölümü protokolden düşüyor. Önemli bir grup çalışmanın ilgi alanına girmiyor; klinik alevlenme olup, serolojik aktivitesi olmayan bir grup bu; yani, bunların çalışmasına çok ters bir grup. Elli yedi hastada ise çok önemli bir şey olmuyor; bu grup gözlem dönemini iyi bir şekilde tamamlıyor. Şimdi tekrar başa, iki grubun karşılaştırmasına dönersek, burada gözünüze çarpabilecek, muhtemel basım hatası olarak değerlendirdiğimiz bir şey var; bilmiyorum onları görebiliyor musunuz? İnsan böyle büyük tablolarda bunları atlayabiliyor, demek ki editörler de atlayabiliyor. Burada eş değerlendirme (*peer review*) ve son editöriyel değerlendirme açısından bazı basit noktalarda önemli zaafılar gözüküyor. Bir tanesi yaşta, herhalde 35, bir tanesi hastalık süresinde, herhalde bir o da buradan düştü diye düşündük. Malum, Afrikan-Amerikan hastalarda hastalık şiddetli olabilir ama çok anlamlı fark gösteren bir grup yok. Bir grup daha az, bir grup daha çok; iki grubu bu şekilde eşleştiriyorlar. Bizim açımızdan önemli parametreler biraz daha aşağıda. Serolojik olarak gruplar birbirine yakın. Özellikle C3a üzerinde çok durmuşlar. Bu aslında bizim rutin klinik pratikte bakmadığımız bir biyolojik marker; biz C3 bakıyoruz sadece. Ona ikinci derecede önem veriyorlar, sonuçta hiçbir serolojik marker arasında fark olmadığını söylüyorlar ama rahatlıkla görebileceğiniz gibi bu plasebo grubunda nedense DNA titreleri biraz daha yüksek. Bu ilk bakışta zannettiğimiz kadar önemli olmayabilir çünkü ELISA yönteminde, antikor belli bir bağlanma düzeyine geldikten sonra bir plato çizmeye başlar, yani çok değeri olmayabilir. Orada belli bir sınırı geçtikten sonra yani yüksek olması yeterli olabilir; anlamlı bir fark değil diyorlar. Hasar indeksinin genelini düşünersek, burada iki katı gibi gözüküyor, ama genelini düşünersek 100 puan toplamı olan bir indekste, yarım olmuş, bir veya 2 olmuş çok önemi yok. Ortalamalara baktığınız zaman, yani hastalar arasında mühim bir hasar farkı yok gerçekten, aktivite farkı yok, serolojik skorlar açısından, global değerlendirme açısından da yine bir fark yok. Yani aktivite ve organ hasarı bakımından birbirine yakın hastalar, önemli bir

organ hasarı görmemiş bir hasta grubu, dolayısıyla, baştan söylediğim, iyice daraltılmış bir hasta grubu olduğunu görüyoruz. Bu gruba ilgili başka katkıda bulunmak veya soru sormak isteyen var mı? Yani görüyorsunuz, 20'şer hastaya kadar inmek zorunda kalınmış, çok özel bir grup hasta aradığımız anlaşılıyor. 180 hastadan başladık, aşağı yukarı %25-20 gibi bir gruba inmiş bulunuyoruz.

Hasta sayıları nedeniyle istatistik analizlerden daha çok non-parametrik olanları seçmişler, bir sürü de analiz yapmışlar. İstatistikle uğraşan arkadaşlarımız bunu belki böyle bir çalışma için biraz fazla bulmuş olabilir, bunun üzerinde daha sonra duracağız.

A. Dinç: Hastalar iki gruba randomize oluyor, sonra bir grup ilaç alıyor, diğer grup almıyor. Sonra ikisini 90 gün takip ediyorsunuz.

M. İnanç: 3 aylık takip süresi var.

A. Dinç: Burada ilaç alan grubu bir ay korumuş oluyorsunuz siz, süre olarak baktığınız zaman. İlaç bittikten sonra bir 90 günü almak daha mı doğru acaba?

M. İnanç: Bazal tedaviler devam ediyor ama bunu gözden kaçırmayalım. Hatırlarsan belli bir dozun altında steroid, bazal tedaviyle karışmıyor. Ortalaması 3 mg, tedavi size yakın bir durum yani; plasebo grubunu kast ediyorum.

A. Dinç: Tedavi bittikten sonra 3 ay ikisinde de eşit olarak düşünülebilir mi? Yani burada bir problem var mı diye soracaktım. Çünkü birisinde ilaç alıp, yeni bitmiş bir dönem var, diğerinde plasebo kullanılmış. Aktivasyon yükselemesi gerçekleştikten sonra aslında birisinde 3 ay (bir ay tedavi) takip edilmiş gibi bir süreye geçmiş oluyor, diğerinde ise sadece 3 ay geçmiş oluyor.

M. İnanç: Toplam takip süresi 3 ay diyor.

B. Çakır: Yani iki grup da 3 ay izleniyor.

A. Dinç: Aktivasyondan sonra 4 ay geçmiş oluyor.

M. İnanç: İkisi de aynı.

A. Dinç: Tamam, ama ikisi aktivasyon olduktan sonra randomize olduğu için birinde 4 ay geçmiş oluyor.

Katılımcı: Makaledeki Tablo 2'ye baktığınız zaman 6 kişi 60 günden sonra aktivasyon göstermiş. Ayhan'ın dediği mantıkta da şu tabii, siz 3 ayın 90 gününde 30 gün tedavi veriyor, 60 günde atağa bakıyorsunuz, ama diğer grupta 90 günlük atağa bakıyorsunuz. En azından bunun da etkisi var diye bu şekilde analizin tekrarlanması uygun olur. Altmış günde ataklarda değişiklik var mı diye bakmak lazım.

B. Çakır: Grubu serolojik olarak aktif durumda dediğimiz anda ikiye bölüyoruz. Bir gruba plasebo veriliyor, bir gruba tedavi veriliyor ve bundan sonra 90 gün takip ediliyor. Ancak söylediğiniz çok doğru, bir grup ilk 30 gün, artı bir hafta artı bir hafta dozu azaltarak tedavi programını tamamladığı için, aslında o tedavi süresi içerisinde (1 ay) tedavi tamamlanmamış durumda. Ama şöyle bir iddia yok zaten, tedavi bittikten veya tedavi tamamlandıktan sonra şu kadar doz steroid denilmiyor. Hastalar steroid başladıktan sonrası itibarı ile takibe alınıyor; yani, tedaviye başladıktan sonraki 90 gün içinde tedavi alan gruptaki hastaların diğer gruba göre alevlenme olasılığı takip ediyor. Bu konuyu sonuçlarda tekrar konuşacağız ama steroid alan grupta klinik alevlenme süreci zaten daha uzun. Bunun sebebi de aslında hastaların bir kısmı zaten steroid alıyor ve steroid etkisi azaldıkça hastada atak olasılığı ortaya çıkıyor. Çalışmanın kısıtlamalarında tekrar bu konuyu tartışacağız ama bu gözlem çok doğru, belki o açıdan da çalışma sonuçlarının genellenebilirliği ve manası tartışılabilir.

M. İnanç: Gözlem süresi biraz daha uzun olsaydı şüphesiz çok daha iyi olurdu. Niye 3 ayda kestiler, doğrusu çok makul bir sebebinin ben göremedim. Belki 3 aydan fazla kör olarak kalmasını etik kabul etmediler. Oradaki dönem 3 ayla kısıtlanmış, bence de kısa bir süre ama bu protokolün kısıtlılığı diyelim. O zaman kısaca sonuçları da gözden geçirelim ki, elimizde ne sonuç var, ona göre metodoloji acaba nasıl daha iyi olabilirdi, onu düşünebiliriz. Burada, 3627. sayfada, alevlenmelerin tipleri var; hafif, orta ve şiddetli alevlenmeleri dikkate alıyorlar. 90 günün altındaki, randomizasyondan itibaren olan, alevlenmelere bakıyorlar. Aslında so-

nuş itibarıyla toplam alevlenmeler, prednizon grubunda %19, diğer grupta %35 ve arada anlamlı bir fark yok; ne Fisher'la ne Log-Rank testiyle. Prednizon aldığı zaman daha iyi olma rölatif riskini de 2.2 olarak hesaplamışlar. Çoklu alevlenmeleri dikkate alırlarsa, burada benim bilmediğim bir düzeltme var; Andersen and Gill diye, o zaman en fazla 2.36'ya çıkıyor, yani çok önemli bir fark yok aslında arada. Plasebo grubunda randomizasyon sırasında medyan anti-DNA düzeylerinin daha yüksek olduğunu onlar da yazmışlar ve Cox'la bunu düzeltmediklerini belirtmişler. Medyanları kullanarak bazalde yüksek anti-DNA ile bu plasebo grubundaki ciddi alevlenmeler arasında ilişki kurulabilir mi diye ayrı bir çalışma yapmışlar. Burada da prednizon alan 7 hastada hiç alevlenme yokken, plasebo alan 13 hastanın 5'inde alevlenme var. Burada p değerini log-Rank testle 0.07 bulmuşlar. Buldukları en anlamlı sonuç bu. Serolojik bulgular ve SLEDAI skorları üzerine tedavinin etkisi olduğunu yazmışlar. Bunların grafikleri yok, sadece sonuçta 3628. sayfada bir grafik var. Burada bu steroid verdikleri grup. Tabii beklediği gibi ortalama steroid dozunu burada azaltıyorlar ve öyle kalıyor, grafiğin bize gösterdiği bu. Diğer gruptaysa 3'lerde ama 10'a kadar yavaş yavaş bir artış gösteriyor, alevlenme olanlarda 32.5'a kadar çıkıyor. Yani belki de tartışmada da üzerinde durdukları gibi, şu kadar steroid kullanımı yaparsak sonraki aylarda kullanmamız gereken kümülatif steroid daha az, ama bunu yapmazsak, alevlenme olan hastalarda yine daha yüksek ve daha uzun süre steroid kullanılması gerekiyor demek istiyorlar; bu grafikten çıkarttığımı bu.

Yine sonuçlarda önem verdikleri şeylerden bir tanesi, gözlem grubundaki 21 hastada anti-DNA ve C3A yükselmesi olmamasına rağmen alevlenme görülmesi. Yani ilgilendikleri gruptan daha da büyük bir grupta aslında, geri kalan hastalarda bir alevlenme görülmüş oluyor. Bu da bize bu biyolojik markerlerin bütün alevlenmeleri göstermesi konusunda ciddi yetersizlikleri olduğunu gösteriyor. Yine tersten baktıysak, biyolojik markerleri yükseldiği halde, alevlenme olmayan önemli bir grup hasta da

var; en azından bu gözlem dönemlerinde. Şimdi isterseniz bu şekilde özetledikten sonra randomize kontrollü çalışma açısından, yani teknik açıdan bu çalışma ne ifade ediyor? Bununun hazırladığı sorgulama formuyla bunun üzerinde duralım. Sonra tartışmada ne kadar anlamlı bir tartışma yapmışlar onu konuşuruz.

B. Çakır: Teşekkür ederim. Hem dün konuştuğumuz konuların üzerinden geçelim, hem de bu makalede olmayan bazı eksik şeyleri nasıl yapmamız gerekir diye de konuşmak istedik. Öncelikle, çoğunuzun kullandığından eminim ama kullanmayan da olabilir; CONSORT ile ilgili iki tane konuya değindik. Biliyorsunuz CONSORT, randomize klinik çalışmaların nasıl okunması, nasıl değerlendirilmesi, nasıl yazılması gerektiği konusunda birçok uzmanın hemfikir olduğu bir kontrol listesi (Ek 1). Belki sizler de asistan eğitiminizde bunu kullanmayı yararlı bulabilirsiniz diye düşünüyorum. Çünkü klinik araştırmaları okurken birçok şey atlanabiliyor; özellikle klinisyenler, esas olarak olayın klinik yanıyla çok ilgilendikleri için, ister istemez metottan biraz kaçabiliyorlar, bazı şeylerin göz ardı edildiğini atlayabiliyorlar. Onun için bu kontrol listeleri çalışmalar yöntem açısından doğru yapılmış mı, kendimizi kontrol etmek açısından çok önemli. Nitekim biz de bu liste paralelinde bu makaleyi değerlendirdiğimizde, makalede bazı konuların eksik incelendiğini ya da eksik rapor edildiğini gördük. Dolayısıyla belki bu bir dergiye hakemlik yaparken de, incelediğiniz klinik araştırmada neler unutulmuş, göz ardı edilmiş ya da yeterince vurgulanmamış; hangi konulara yer verilseydi daha iyi olurdu şeklinde değerlendirmelerde bu tür listeler çok yararlı. Tabii okuyucunun bazı yerlerden ipuçları bularak bazı varsayımlarda bulunması mümkün ama bu tercih edilmiyor zira bu tür değerlendirmelerde insan faktörü gündeme geliyor.

Öncelikle ben akış şeması üzerinden gitmek istiyorum, çünkü bu makalede akış şemasında birkaç tane sıkıntımız var ve eğer CONSORT akış şemasına (Ek 2) paralel olarak incelediklersek, bir takım farklılıklar olduğunu göreceğiz. Aslında burada çalışma amaçları belirtilir-

ken birkaç değişik şey yapılmış. Ben makale-nin özet bölümünü okurken çalışmanın 3 ana objektifi olduğunu düşündüm. Tartışmayı okurken, özellikle tartışmanın ilk kısmını okurken de, bunlardan ikincisinin üzerinde yoğun-laşıldığını görüyoruz. Oysa, çalışma sonuçları bir amaç üzerinde yoğunlaşmış. Olasıdır ki, ça-lışma sonuçları hakkında bir başka makale da-ha yazmışlar, sonuçların bir kısmına da maka-lede vermişler. Tartışmada ise gündemdeki so-runları da biraz kaleme alalım deyip, sonuçlar-da sunulmayan bazı şeyleri de paylaşmışlar: Bu bizim çok arzu ettiğimiz şey değil tabii. Araştırma sunumlarında tartışma sonuçlar ışı-ğında ve kapsamında yapılmalıdır.

Alevlenmelerin SLEDAI skorlarının iki farklı kişi tarafından ve serolojik marker sonuçlarının bilmeden değerlendirilmesi herhangi bir *infor-mation bias*'ı ortadan kaldırmak yönünde güç-lü bir çalışma, bu iyi.

Çalışmada ikinci amaç, ki bu oldukça gizli kalmış bir amaç, çalışma sonuçları sunulurken, prednizon tedavisi verilmezse skorlarda ve ör-neğin markerlerde bir düşme oluyor mu, yani çalışmaya baktığımız zaman bu tedavi klinik alevlenme yapıyor mu diye planlanmış. Bu tab-loda öyle dikkat ederseniz; prednizon verilen ve verilmeyen grupta alevlenmeler var mı? Bu-radaki alevlenmeler aslında klinik alevlenme-ler anlamında, ikinci bir objektif olarak serolo-jik test sonuçları bakılmış ve üçüncü olarak da, ki asıl gündemdeki konu bu, bu markerler kli-nik açıdan, hastalığın klinik aktivitesi açısın-dan marker olarak kullanılabilir mi? Araştırma-nın akış şemasına bakarsak, CONSORT'un akış şemasında, bu kısım "uygun olan kaç kişidir" diye soruyor. Uygun olanlar 154 kişi, yani 180 kişiyi taramışlar ve onlardan sadece 154 kişi bu çalışma için uygun olmuş. Demek ki biz bu-roya 154 yazacağız. CONSORT akış şemasında, yana bir ok var (Ek 2). O grup da şu 113 kişi; biz bunları atıyoruz aslında, burada bu kadar detaylı bilgiye ihtiyacımız yok. Bunu niye ver-mişler? Ben ikinci objektif için verdiler diye dü-şünüyorum. Aynı zamanda tedavi alıp, bizim kategorimize uymayan gruplarda da bu mar-kerlerle klinik aktivite arasındaki ilişkiyi kon-

trol etmek için. Çok kısa olarak sonuç kısımın-da da yer var. Burası belki o anlamda fazla ol-mayabilir ama o objektifin daha net verilmesini beklerdim ben.

Bu kısımda da bir eksik var. Yine akış şe-mamıza bakarsak, kaç kişi randomize edildi? Birinci gruba 20 ve ikinci gruba 21 olmak üzere toplam 41 birey. Şurada büyük bir kısım eksik akış şemamızda; bunun üzerinden konuşaca-ğız biraz. Kişilerin gruplara atanması (allokas-yon) nasıl yapıldı, kimler grupta yer aldı, alma-dıysa niye almadı? Sonra biraz daha takip edi-yorsunuz, bu takipte 21 ve 20 kişiden kimler kayboldu, kaybolanlar niye kayboldu ya da gruplar arasında bir değişim var mı? Bunları hiç bilmiyoruz.

Yine altta eksik bir yer var, kaç kişi analize girdi? Dün çok kısaca konuştuk, isterseniz üs-tünden tekrar gidebiliriz. Analizde *intention-to-treat* prensibi var. Yani ne demek? 21 kişi orada, 20 kişi burada; ben bu kişileri aldım ve bunların bir kısmı izlem içinde kaybolmuş ola-bilir, diğer gruba geçebilir, ek tedaviler alabi-lir. *Intention-to-treat* analizinde kişiler, izlem-de gruplarda kalma durumlarından bağımsız olarak, kişileri ilk atandıkları grup içinde teda-vi edilmiş ve sonuna kadar bu grubun üyesi kalmıyacağına analiz ediliyor. Ancak, yazarların bu konuyu bize detaylandırmasında yarar var, yani prednizon grubunda mı kaçak var ya da öbür grup, plasebo grubunda mı kaçak var? Bu kişiler tedavi edici başka yöntemler kullanmış-lar mı, tedavinin yerine ya da ek olarak alter-natif bir tedavi almış olabilir mi? Bu konulara açıklık getirilmesi uygun olurdu.

H. Yazıcı: Daha fazla ilerlemeden araya gi-receğim. Bu yazıda temelde bir sorun var gibi-me geliyor benim. Şimdi en sonuna bakalım, bir subset almışlar ve bu subset klinik olarak stabil ve onları kortizon alan ve almayanlar olarak bölmüşler; tamam mı? Ve görmüşler ki kortizon bunlara iyi geliyor. Şimdi bu cümlede bir mantık hatası var, o da şu: Bunlar klinik olarak stabil olmayan, serolojik olarak stabil olamayan bir grup insan, serolojik olarak des-teklenen ama klinik olarak stabil olmayan bir grup insan almadıklarına göre, bunların sonuç-

larının bu gruba münhasırdır demeye pek hakları yok, yani iki işi birden yapmaya çalışıyorlar gibime geliyor.

B. Çakır: Öyle bir karışıklık var baştan sona.

H. Yazıcı: Biraz önce sizin dediğiniz, hani tartışmayı ekstrapole ediyorlar dediniz ya, bu ekstrapole etme çabasının da esas nedeni şu: Şimdi çift kör çalışma yaptınız. Çift kör çalışmayı bir limite grup insanda yaptınız ve limite grup insanda bunu gösterdiniz; çok güzel. Ama o limite grup insanda bu iş has demeye hakkınız yok çünkü o çalışmayı yapmadınız, böyle serolojik olarak belirlenmiş klinik olarak stabil olmayan insana böyle lupus dersenez, çift kör yaparsanız, belki onlara da iyi gelecek. Onu tartışmıyorlar hiçbir yerinde değil mi?

B. Çakır: Onu bilmiyorlar. Bilmedikleri için...

H. Yazıcı: Onu bilmediklerini de söylemek durumundalar ama. Söylemeleri lazım, çünkü onu söylemeyince adeta bu işleri yapmak, 180 hastadan buralara inmek falan biraz stratifikasyon gibi geliyor. Halbuki kortizonu geri kalanlara da verseler, onlarda da aynı şeyler çıkacak gibi geliyor bana. Yani makaleyi kritik ederken o mantığı gözetmek lazım.

B. Çakır: Şimdi, şöyle bir şey var, birkaç yerde üzerinde durmuşlar. Makalede, 3629. sayfanın son paragrafında diyor ki, "sonuçta bu grup, 21 hasta ve 20 kontrol, serolojik açıdan aktif". Onu da nasıl belirliyorlar: Son iki ay içerisinde işte şu düzeyde C3A'sı ve antids-DNA'sı yükselenler. Yani bunlar çok özelliği olan bir grup olduğu için bu sonuçlar genellenebilir mi, genellenemez mi? Ama burada tartışmada bulgular dışında o kadar detay anlatacaklarına, sizin dediğiniz gibi bunu vurgulamaları hoş olabiliirdi; ama gerçekten çok özelliği olan bir grupta çalıştıkları için kime genellenebilir bu sonuçlar, bu grup insana diye belirtmişler.

H. Yazıcı: Bu makaleyi okuyan kişi yavaş, yavaş şu mantığı çıkarıyor. İlk önce okur okumaz bu izlenimi çıkarttı. Ama çıkartmadan önce de düşündüm epeyce. Şimdi bu şöyle değil; mesela Behçet'te göz tutulumu olanları ayırdınız, göz çalışması yaptınız. Göz çalışmasında bir gruba siklofosamid verdiniz, siklo-

fosamid iyi geldi diyelim, o zaman hakkınız var, göz belli, ne olduğu belli. Ama burada onları ayırdıktan sonra iyi gelir imajı veriyorlar ki, bunları ayırmışlar. Tamam peki, 20-20 yapmışlar, o bilgi bizim için yeterli. Ekstrapole etmek burada işi tehlikeye sokuyor.

Katılımcı: Tabii çok doğru bir şey, eksternal validasyonu açısından aynı zamanda, bir grupta da önemli ölçüde alevlenme olduğunu göz önüne almamız lazım, 21 hastada. Burada çok teknik bir problemleri var, öbür hastalarda da bu etkili olabilir. Sizin dediğiniz gibi, bu olasılık kesinlikle var, ama orada tedaviyi ne zaman vereceklerini bilmiyorlar. Tabii burada şunu unutmamak lazım; kortizonu verdikleri süre kısıtlı. Hipotezlerine göre kısa süreli prednizon kullanımının bir şeyi engellediğini ispat etmeye çalışıyorlar. Ama şöyle diyebilirlerse belki, biz bu etkiyi yapabilecek, daha uzun etkili başka bir ilaç kullanacağız filan, o zaman tabii bu tamamen iflas eder. Burada serolojik aktiviteyi, biyolojik aktiviteyi bir marker olarak almış. Ve bu grupta öyle bir marker yok, ne zaman yapacağını bilmiyor o bir aylık tedaviyi.

B. Çakır: Şimdi ben şöyle düşündüm, belki cevap da olabilir. Biz yeni bir çalışmaya başladığımız zaman, kafamızda bir plan oluşurken bu var mı diye, bunu en riskli grupta deneriz sıklıkla. Bu grup markerleri yüksek, serolojik olarak aktif. Bu hasta grubu klinik olarak stabilse de, klinik alevlenmesi olması en çok beklenen grup. Bu kişiler tedavi edilmezse, sonrasında en kötü grupta olacağı için hem bu grubu tedavi edip böylece onların beklenen sekelilerini azaltmaya çalışıyoruz; eş zamanlı olarak eğer bu hasta grubunda steroide bağlı bir advers etki olacaksa bile bu grupta o tolere edilebilecek bir şey, çünkü bu grup hastalık açısından riskli bir grup, dolayısıyla ilacın istenmeyen etkileri olsa da tedavi üzerindeki istenen etkisi de önem taşıyacağı için, kar-zarar dengesinde pozitif yönde olacağız. Öte yandan, çalışma tamamlandıktan sonra devam edilerek, ki zaten yazı da bunu öneriyor, seçilmişlik açısından sınırları daha az keskin bir hasta grubunda benzer bir çalışma yapılması uygun olacaktır. Sizin dediğiniz gibi C3A'ları yükselmiş 2 hastaları daha var ama düzeyleri

istenilen kadar yüksek olmadığı için onları çalışmaya alamamışlar. Çalışma sonuçları alındıktan sonra, hasta popülasyonunu daha geniş tutarak tedaviyi diğer gruplara da yapmaları beklenir zaten. Keşke tartışma bu açıdan detaylandırılabilseydi.

N. Akkoç: Bunun dışında bir şey bu randomizasyonla ilgili; stratifikasyon kullanmışlar bu randomizasyon için. Oradaki kullandıkları da hastalık aktivitesi ve sitotoksiste, şimdi tabii bildiğimiz bir şey var, özellikle böbrek alevlenmesi açısından anti-ds DNA'nın titresi de önemli. Hakikaten baktığınız zaman plasebo grubundaki alevlenmelerin büyük çoğunluğu anti ds DNA'sı 416'nın üzerindeki grupta olmuş. Öte yandan bu başlangıç özelliklerine baktığımız zaman, Tablo 1'de gördüğümüz şey, örneklem küçük olunca randomizasyonda da problemler ortaya çıkabiliyor, çünkü bazı şeylerde aslında fark yok gibi gözüküyor ama, baktığınız zaman anti-ds DNA prednizon grubunda 391, plasebo grubunda 970; oldukça fark var. Aslında şunu da görmek isterdim açıkçası, burada fark yok demişler ama tamamen sayıyla ilgili olabilir fark görmemeleri. İkincisi de bu 416 limitini kullanmışlar; bu limiti geçen kaç hasta plasebo grubundaydı, kaç hasta prednizon grubundaydı, onlarda fark yok muydu? Onu görmek iyi olurdu.

B. Çakır: Bizim ilk 4 sorumuz da öyle. İster-seniz hazırladığımız soru seti ile devam edelim (Ek 3). Bir kere makaledeki Tablo 1'de önemli bir sıkıntı sizin de vurguladığınız şekilde; bunlar normal dağılıma uymamasına rağmen, tabloda değişkenleri tanımlar iken hep ortalama ve standart sapma kullanılmış. Bu doğru bir gösterim değil. Bizim bu değişkenler için minimum ve maksimum değerleri, değerler aralığını (ranj), medyan değerleri mutlaka bilmemiz lazım- merkezi eğilim ölçütleri olarak. Nitekim analiz yaparken medyan değerler üzerinden bölerek, mesela anti-ds DNA değerlerini medyandan ikiye bölerek incelemişler. Bu bağlamda, sonraki işlemleri doğru yapmışlar ama tabloda neden minimum, maksimum ve medyan değerleri yok, anlaşılamıyor, bu önemli bir eksik. Olsaydı belki bu soruların bir kısmı günde-

me gelmeyecekti. Bu konu için bir takım düzeltmeler yapmışlar: Yalnız o kadar çok analiz var ki, bu kadar analiz 40 kişi ile yapılmaz. Biz bütün bu yöntemleri biliyoruz, yaptık gibi sanki bilgilerini göstermeye çalışmışlar.

Şimdi bu ilk 4 soruya geçmeden önce, ikinci kısım yani materyal ve metod açısından nasıl bir değerlendirme yapabiliriz konusunu tartışmak istiyorum. Dünkü konuşmamızın devamı olarak CONSORT kontrol listesinde bizimle ilgili 4 tane madde var: 3 tanesi randomizasyonla ilgili, bir tanesi de körleme ile ilgili. Randomizasyonda sekansı nasıl oluşturursunuz, randomizasyonda allokasyonu nasıl yaparsınız ve de bunların alt maddeleri var. Hatırlarsanız dün iki konunun birbirinden ayıramayacağımızı söylemiştik; randomizasyon yapılacak ama sadece kimin nereye ayrılacağına karar vermek yeterli değil, bu kararı verdikten sonra, çünkü yazı tura örneği üzerinden gidersek, tura çıkan kişi gerçekten turanın işaret ettiği tedavi koluna mı alındı? konusu önemli. Bu allokasyon dediğimiz şey. Ondan sonra da aldığı ilacı alırken bunu biliyor muydu, bilmiyor muydu, bu da körleme. Bu üçünü (yani randomizasyon, allokasyon ve körleme özelliklerini) birbirinden ayıramayacağımız için bugün sorularımızı da bu şekilde tartışmak istiyoruz.

Devam edersek, yine belirtmek isterim ki randomizasyon doğru gerçekleştirilmeli: 41 ki-şimiz var ve bu 41 hastanın 20 ya da 21'inden hangisini tedaviye hangisini plaseboya koyalım? Dolayısıyla 1, 3, 5 mi tedaviye girecek ya da 1, 13, 18 mi girecek? Bu, randomizasyon gruplarının belirlenmesi. Burada sayı çok küçük olduğu için bir sınırlama yapmışlar. Dün söylemiştik, hasta sayısı 200'ün altında ise, belki kör çalışmalarda sayı çok önemli değil ama yazı tura atmak gibi basit yöntemlerin grupları eşitlemeyeceğini biliyoruz. Onun için burada bir sınırlama var. Bir gruba 20, bir gruba 21 alacağım, eşleştireceğim diye yazarların baştan kafalarında bir plan var, bunu gerçekleştirmeleri lazım. Bu başlangıç itibarıyla yapacağımız şeyin doğru olup olmadığı konusunda ki taraf tutmayı, biası engellemek için.

Şimdi tablo üzerinden tekrar gidelim. Bu prognostik faktörleri, istatistiksel olarak anlamlı olsun ya da olmasın romatolog olarak siz gündeme getireceksiniz ki bunun için bir düzeltme yapılsın. Hiçbir istatistikçi ve epidemiyolog tek başına hangi faktörlerin prognoz açısından önemli olacağına karar veremeyeceği gibi, hekim de olsa, o alanda çalışmadıysa, ilgili klinikten değilse, bu prognostik faktörleri alanın hekimi söylemeden bilmeyecektir. Başka bir ifade ile klinisyenlerin istatistiğe yön vermek açısından destek vermeleri önemlidir.

Araştırmada körleme ve *intention-to-treat* analizi yapmışlar ama neden yapmışlar, nasıl yapmışlar biraz da onları konuşalım istiyorum ben. İlk sorumuz random allokalasyon sekansları nasıl oluşmuş ve tanımlanmış mı? Biz bir blok randomizasyon kullandık ve bir stratifiye randomizasyon kullandık diyorlar. Blok randomizasyon demek bu şema üzerinden sınırlı randomizasyon yöntemlerinden birini kullanmışlar demek. Yani basit randomizasyon yapmamışlar, yazı tura atmamışlar. Blok randomizasyon grup içi ve kısa dönemde bir randomizasyon yapmak istiyorsak kullandığımız bir yöntem. Bu örnekte olduğu gibi iki grup söz konusu olduğunda gruplar bloklara ayrılacaktır. Bu örnekte bir de tabakalama var. Tabakalar neydi, hatırlarsak, indekse göre 4 ve 4'ün altı; aktiviteye göre 4'ün altı ve 4'ün üstü diye ikiye ayırmışlar. Bir de daha önce sitotoksik ajan alıp almadığına göre, yani 4 tane grup var. O zaman bu 40 kişiyi, sahip oldukları özelliklere göre 4'e böleceğiz. Her grubun içerisinde yani 10 civarında kişi beklerim ama 5 de olabilir; mesela o sayılar bize hiç verilmemiş, keşke verilseydi. Diyelim ki birinci grupta aktivite skoru düşük ve sitotoksik ajan alan varsayalım 10 kişi var, burada bir blok randomizasyon yapacağız. Ne yapmamız gerekiyor? Bu 10 kişinin 5'i tedavi, 5'i plasebo alacak tarzda 2 gruba randomize edilsin istiyoruz. Ama burada blok büyüklüğü değişebilir diyor. Mesela o grupta 15 kişi varsa, önce bir 6'lı blok yapıp, sonra 9'lu bir blok yapabilirsiniz ama sayıları eşitlemişler. Bu bizim için önemli, bu bilgiyi bize veriyorlar.

H. Yazıcı: Bir şey sormak istiyorum. Bir çalışmaya başladınız ve çalışmaya hasta alıyorsunuz diyelim, sadece bu çalışmada değil, yani gerçek hayatta nasıl oluyor? Çalışmaya giriş süresi bir yıl ve çalışmaya aldığınız insanların bir kısmının özelliklerini iyi biliyorsunuz, bir kısmınınkini de başlangıçtan pek bilemiyorsunuz. Şimdi bilemediğiniz zaman bu stratifikasyon ve bloklamayı, ikisi beraber veya tek tek, nasıl yapıyorsunuz? Bu insanların bunu da bir yerlerde anlatması lazım. Bunları 3 ay beklettiniz mi? Veya bütün uygunları seçtiniz, bunların arasında katılmayanlar da olabilir filan. O zaman blok veya stratifiye, *simple* dışında olduğu zaman, o yöntemin ne olduğu pek ortaya çıkmıyor.

B. Çakır: Çok önemli, şimdi bize şöyle bir şey söylemiş, “18 aylık bir başlangıç gözlem dönemim” var diyor. Bu 154 kişiyi, gruba katıldıktan sonra 18 ay gözlüyor. Ancak dediğiniz gibi bu 5 yıllık bir izlem, yani veriler 5 yıl içerisinde toplanmış. Çok önemli bir detay ve burada mutlaka verilmesi gerekirdi. Hastalar klinik açıdan takip ediliyor ve aylık olarak C3A değerlerine bakılıyor, aynı şekilde anti-ds DNA'lara bakılıyor. İki ay önceki değerinin üzerine çıkınca bu kişi randomizasyon için uygun hale geliyor. Ama bunların arası ne kadardır, kimdir bu insanlar, o arada işlemden bir bozukluk var mı, işte eğitim almak vs makale içerisinde açık değil.

M. İnanc: Yani önemli özelliklerinden bir tanesi bu makalenin. Burada anlık olarak seroloji kontrol ediliyor, ama rutin yapılan bir şey değil. Dolayısıyla bu hastaları özellikli bir alt gruba daha bölmeli aslında. Anlık seroloji takip edip yakaladığınız hastaları çalışmaya alabiliyorsunuz, iyice daraltmış oluyor. Hasta sayıları azaldı diyoruz ama başka türlü de bunu sağlamak mümkün değil.

N. Akkoç: Aslında baktığınız zaman burada bir randomizasyon nasıl yapılabilir ki?

B. Çakır: Muhtemelen biraz bekleterek.

N. Akkoç: Ama beklettiğin zaman olmaz ki. Beklettiğin zaman çıkıyor, aradaki süre. Yani 5'li blok yapıyorsanız, mümkün değil ki, strati-

fiye yapabilirsiniz ama blok yapamazsınız herhalde değil mi?

B. Çakır: Bir kere stratifikasyon yaptıktan sonra mutlaka blok yapması gerekiyor ki eşit olsun.

N. Akkoç: Hayır mesela stratifiye edip, her yeni hastayı yeni bir yere alabilirsiniz yani blok olmaz.

H. Yazıcı: Blok yapabilmek için tüm hastaların özelliklerini bir arada bilmek lazım. Eğer elinde bir kohortunuz var ise bunlara şu anda uygun mu diye bakıp sonra bir blok randomizasyon yapmak mümkün o anda. Ama zaman içerisinde tek tek geliyor ise blok randomizasyonu nasıl yaparsınız?

B. Çakır: Şöyle yaptıklarını varsaydım ben; birinci hastamız geldi. 4 tane stratum var: 1, 2, 3, 4. Varsayalım ki bu birinci hasta geldi ve birinci stratumun adamı, bir şekilde bir yöntemle, yazı tura atarak bu kişi ya tedavi alacak ya da plasebo. İkinci kişi geldi, bu ikinci stratumda, o da yine rastgele bir şekilde ya tedavi ya plasebo alacak. 3. kişi geldiğinde bu 1. stratumun kişisiyse, 1. grupta birinci tedaviye almıştı hatırlarsanız, o zaman bu 2. kişiyi plaseboya alacağız. Bu şekilde blok yapabiliriz. Yani hasta geldikçe bir tedavi, bir plasebo yapmış olabilirler. Yani 3 ay sonra geldim, 38. kişiyim grupta ama 1. stratumda düştüm, 1. stratumun 3. kişisiyim. O zaman 1. stratumun 1. kişisi tedavi almıştı, 2. kişi plasebo almıştı, ben 38. kişi olarak gelmeme rağmen o stratumda düştüm için benim tekrar tedavi almam lazım. Bu şekilde yapabilirler blok içerisinde. 1., 2., 3., 4., benim kendi blokum belli çünkü, stratumum belli. Bu benim söylediğim şey ön randomizasyona uyuyor ama ön demiyorlar. Onun için bu kısmı açık değil, bunu mutlaka detaylandırmak lazım. Diğer kısımlarını çok iyi yaptık diye göstererek bunu da iyi yaptıklarını bize düşündürdüler ama tabii bu yanlış olabilir, mutlaka açıklanması lazım. Akış şeması elimizde olsaydı, bilebilirdik. Şimdi akış şemamız yok, onun için bunu tayin etmek zor gerçekten. Bir de stratifiye randomizasyon 50 kişiden daha küçük gruplarda iyidir diye biliyoruz.

Soru: Blok kullanmak için örneklem büyüklüğünün şu kadar olması gerek diye bir şey var mı?

B. Çakır: Blok ya da diğerleri için, 200'ün üstündeyse hiç problem değil, stratifikasyon yapabilirsiniz. 200'ün altındaysanız ve körseniz yine problem değil, ama 200'ün altındaysanız ve körleme yoksa mutlaka kısıtlı bir yöntem kullanmanız lazım. Bu blok da olabilir. Stratifikasyon, kişi sayısı 50'nin altında olduğunda, prognostik faktör de varsa yararlı bir şey. Yani sayı yüksek olduğunda sorun yok ancak azaldıkça daha *restricted*, blok, tercihen stratifikasyon yapmak lazım.

Şimdi Tablo 1'e gidelim. Bu gibi tablolar, her klinik randomize çalışmanın başındaki ilk tanımlayıcı tablolardır ve grupların karşılaştırılabilirliğini göstermek açısından yapılır. Bu tür tablolarda verilen p değerleri gruplar arasındaki farkın (varsa) ne oranda şans ile açıklanabileceğinin göstergesi olarak sunulur, p'nin dolayısıyla 0.05'ten, eğer $\alpha=0.05$ kabul edilmiş ise, yüksek olduğunu görmek açısından bizim için önemlidir: grupların karşılaştırılabilirliğine, benzerliğine, işaret eder. Ama bu aşamada "tamam hadi ben devam edeyim dememek lazım". İşte o aşamada biz klinisyene soruyoruz, burada sizin için önemli klinik farklılıklar var mı? Mesela etnik gruplarda hastalık aktivitesinin, aynı şekilde markerlerin farklı olduğunu biliyoruz. Her ne kadar istatistiksel olarak anlamlı değilse de burada Afrikan-Amerikan grubunda bir farklılık var ama sayı çok küçük, 3-6 hadi onu almayalım. İkinci olarak sitotoksik ajan için tabakalaşmışlar. Üçüncü bir şey aktivite skoru; onun için de tabakalayıp randomizasyon yapmışlar. Bizim için bir tek sıkıntı yaratan, dediğiniz gibi, anti-DNA'yı burada yapmamış olmaları. Belki de yapamadılar, grup çok periyodik geldiği için belki de çok farkına da varamadılar. Ama onu da analizlerde yapmışlar.

H. Yazıcı: Benim son zamanlarda etrafta dolaştırdığım bir tane slayt var: Bu $p=1$ slaydı. Bir tane daha buldum çok kızmaya başladım buna, çünkü hakikaten yeni moda oldu, bilgisayarın içinde bu var. Şimdi bakıp da buraya,

bir grupta kaç kişi var? Diyelim, 5 kişi, öbür grupta 5 kişi var yanına p değeri=1.00 yazmak okuyucuya ve kendine saygısızlık.

Katılımcı: Burada da var, ne zaman söyleyeceksiniz diye bekliyordum.

N. Akkoç: Şimdi bu gruplarda biz tedavinin etkinliğini araştırırken p yani tip 1 hataya bakıyoruz yani hipotezimizi öyle çürütüyoruz. P bir de yan etkilerde kafama takılır hep. Aslında hipotezimiz bizim bir fark olmadığı, yani burada tip 1 hatadan çok tip 2 hata daha önemli değil mi?

H. Yazıcı: %5.

B. Çakır: Klinik çalışmalarda grupların randomizasyonu sonrasında bakılan p değerlerinin anlamı yok, hep bunları vurgulamaya çalışıyoruz, aslında koymamak lazım. P değerinin anlamlı olmasını istemiyoruz, zaten gruplar benzer ise p değerleri büyük olacak ki beklediğimiz de bu. Araştırmacılar sadece grup özelliklerini verip bırakabilir ancak, “gruplar farklı değil bulgusunu istatistiksel açıdan göstermek için” kullanılıyor. O zaman ben bu fark için ne yaptım, stratifikasyon mu yaptım, randomizasyonda ya da daha sonrasında analizimi mi değiştirdim? Eğer gruplar farklı ve p değeri anlamlıysa, o zaman p değerini verip, grupların bu farklılığı ile ilgili olarak ne yapıldı, nelere dikkat edildi ve benzeri konuların açıklanması lazım.

Peki, sorularımızı irdelerken ilk sorumuzu yaptık, ikinci sorumuzu yaptık, 41 kişimiz var, 20 + 21, bunu özellikle sordum ben. Başka problem için de kullanabilelim bu kontrol listesini diye. Bire bir randomizasyon mu var? Neden böyle? Burada bire bir, 20’ye 21, niye var çünkü zaten ona göre yapmışlar, blok randomizasyon (*restricted*) yapmışlar. Peki, eğer bu blok olmasaydı ve bire bir olsaydı ne düşündünüz? Bu da çok önemli: Eğer burada blok randomizasyon olmasaydı, biz yazı tura attık ve 20’ye 21 deselerdi, ben işin doğrusu art niyetli düşündüm. Neden çünkü 41 içerisinde 20 ve 21, bu kadar bire bir olması mümkün değil. İnsanlar gruplar eşit olmalı mı diye kendilerini o kadar strese hissediyorlar ki eşit değilse ya rapor etmiyorlar ya da başka bir şey

yapıyorlar. Bu konunun hatırlatılmasıyla ilgili birkaç tane slayt hazırladım ben. Özellikle grup büyüklüklerini eşitlemek istediğimiz zaman restriksiyon gerekebilir. Böyle bir sınırlama getirdiğimiz zaman birazcık çalışmanın gücünü de arttırıyoruz kontrol sağladığımız için, ama eşit olmak zorunda değil.

Bu nokta çok önemli: Gruplardaki sayılar eşit olmak zorunda değil, yakın olmak zorunda. Ama illa eşit istiyorsak ve restriksiyon yapmıyorsak, her zaman kötü niyeti aklımıza getirebilmeliyiz diye düşünüyorum. Her zaman random olmayan bir takım manipülasyonların yapılmış olabileceğini aklımızda tutmak zorundayız.

H. Yazıcı: Çok kısa olarak gerçekten eşit olmayan örneklem büyüklüklerinin sorununu söyleyebilir misiniz bize?

B. Çakır: Bir kere sayı eşit olmadığı zaman, analizlerde sıkıntı olabilir. En önemlisi bu. Ya da bu örnekteki gibi prognostik faktörlerin farklı olduğunu bildiğimiz için sıkıntı olabilir. Diyelim ki 4 tane grubumuz var, bunların en altı, klinik aktivitesinin en düşük olduğu grup, bu gruptaki herkesi tedavi edip de, en üst gruptaki herkese plasebo verirse farkına varmadan çalışmanın sonuçlarını olumsuz etkileriz.

H. Yazıcı: Şunu öğrenmek istiyorum; çok klasik ilaç şirketlerinin yaptığı çalışmalarda, diyelim ki 150 kişiye A ilacı veriliyor, 150 kişiye A ilacının başka dozu veriliyor ondan sonra da 50 kişiye veya 75 kişiye de plasebo veriliyor. Şimdi bunu yap-yapma kavgası var. Yap diyenlerin düşüncesi nasıl, zaten benim plaseboyu etik bakımdan ufak tutmam lazım, tedavi etmeyeceğim insan sayısını minimumda tutmak istiyorum, orada mantık var. İşte para meselesi var. Zavallı kanserlilere ilaç veriyoruz meselesi var. Peki, bizi oranın aritmetik hatası hakkında, anlayacağımız şekilde aydınlatır mısın?

B. Çakır: Ben de bu düşüncedeki gruptayım. Birkaç yıl önce bu konuda çok fazla yayın da oldu. Grup için karar verilen sayının çalıştığımız sonuç açısından yeter bir sayı olması lazım. Ben kafadan 70’tir, 75’tir diyemem, ama çalıştığımız sonuç neyse, mesela advers etki

mi bekliyoruz, %2-3 çıkan bir advers etkiyse 75 yeterli olacaktır, ama binde bir çıkıyorsa 75 sayısı size hiç yeterli olmayacaktır; çıkmayacağını bile bile o ilacı/plaseboyu 75 kişiye uygulamanın anlamı yok.

H. Yazıcı: Benim elimde Liam Shaw'ın kitabı da var. *Power* hesabı. *Power* hesap cetvelinde eşit olmayan örneklem büyüklüklerinde daima sorun var, değil mi?

B. Çakır: Var, ama onun için düzenlemeye çalışılabilir, ama zaten limitlerini verir, eşit alırsanız şöyle alın, eşit olmayan alırsanız da en az şu kadar olsun diye. En azını belirtmek önemli ama sihirli bir sayı ne yazık ki yok. Aslında eşit olduğu zaman da sıkıntı var, özellikle uygulamada. Şöyle düşünün, size 10'lu bloklar geliyor, bu ne demek, 5 tanesi tedavi, 5 tanesi plasebo demek. Doktorların, özellikle klinisyenlerin olaylara hakim olma isteği bilinen bir gerçek. Örneğin hekim bilmesin diye tedavinin kodu opak zarf içinde verilir. Neden? Çünkü opak olmayan zarfı doktor mutlaka ışığa tutar ve içine bakar. Bu çok bilinen ve ispat edilmiş bir şey. Yani doktora diyorsunuz ki, "size bir zarf vereceğim, hastanın yanında açacaksınız, hatta hastanın üzerine ismini yazacaksınız. Banu Çakır şu nedenle uygundur, bu dosyadan çıkan ilacı ona vereceğim diyeceksiniz. Zarftan A çıkarsa tedavi, B çıkarsa plasebo vereceğim". Doktor zarfın içindekinin ne olduğuna bakıyor, eğer Banu Çakır için o tedavinin uygun olmayacağını düşünüyorsa A'yı vermiyor ve o zaman yaptığınız randomizasyon tamamen bozuluyor. Onun için zarfın opak olmasını tercih ediyoruz; allokasyon bozulmasını diye.

Randomizasyonda bir yerden sonra doktor farkına varıyor ve diyor ki, her 10'da bir veya 2'de bir plasebo geliyor. Bunu da nereden biliyor, özellikle ilaç yan etki yapmazsa plasebo oluyor. Onun için aktif plasebo diye bir kavram çıktı biliyorsunuz, yani plasebo, hiçbir şekilde içinde ilaç yok ama yardımcı yan etki yapacak bir ajan var, yani işe yaradığı gibi bir de size bulantı yaptırıyor.

H. Yazıcı: Onu kim yaptı önce biliyor musunuz? 1977 yılında kolşisin çift-kör çalışmasında plasebo içerisine fenoltalein koymuştuk.

B. Çakır: Evet. Bu önemli, doktorların randomizasyonu bozmasını engellemek açısından.

A. Dinç: Burada şöyle bir faktör var, onu söyleyeyim, lupusa özgü bir faktör. Lupusta benim gözlediğim hastalarda mesela tipik olarak yorgunluk çok sık olan bir şeydir ve steroid de mükemmel yanıt verir. Dolayısıyla lupusla uğraşan biriye bu çalışmayı yapan ki öyle, o klinikteki klinisyenler, steroid alan hastaları biraz düşünseler, çok rahat hastanın mutluluğunu ayırabilirler, çok net bir etkidir ve akuttur bu. Mesela bu çalışmada o tip semptomlardan da hiç bahsedilmemiş, hep böyle bir şey çıkabilir tedaviyle ilgili.

B. Çakır: Bu beni çok düşündürdü. Neden? Çünkü bu çalışmanın her şeyi çok güzel fakat allokasyonla ilgili, allokasyonun gizlenmesiyle ilgili hiçbir şey yok makalede ve doktor beyin söylediği o kadar önemli bir şey ki. Doktorların verilen tedaviyi ayırdetme olasılığının yüksek olacağı tahmin edilebildiği halde bu makalede allokasyonun nasıl gizlendiği konusunda bilgi olmaması çok ciddi bir eksiklik ve art niyet düşünmeseniz bile çalışmanın çok önemli bir limitasyonu. Bu yazı bana gelseydi, kendi adıma, allokasyon hakkında detaylı bilgi olmadan bu yazının yayınlanmasına izin veremezdim, nasıl verildiğini de çok anlamış değilim.

Soru: Anti-DNA genelde glomerülde yükseliyor ama burada böbrek tutulması olan hasta sayısını bilmiyoruz. Bazalde verilmemiş. Bir tarafta bir varsa diğer tarafta 9 böbrek hastası varsa yine bu sonuçları değerlendirmek zor.

M. İnanç: Şöyle, böbrek tutulumunu bir kriter olarak almamış da, geçmişte anti-DNA yüksekliği olanları aldığı için böbrek olsun olmasın aldık demektir. Ama şunu söyleyebiliriz önemli böbrek hasarı olan hasta yok. Çünkü kronisite skorları çok düşük.

B. Çakır: Bir şeyi eksik bırakmışız sorularda, dün de çok hızlı geçti, onun için onu vurgulayıp, allokasyon konusuna daha sonra geçmek istiyorum. Şimdi bu grup içerisinde akış şemasında da olmadığı için, prednizon uygulanan grup, plasebo uygulanan grupların her ikisi içerisinde de takip süresinin sonuna kadar giden ve gitmeyen insanlar olacaktır. O zaman

analizi çok farklı yöntemlerle yapabiliriz, prednizon almış ve sonucu belli N1'leri N2, C1 ve C2'yle mi karşılaştıracğız, N1'leri C2'yle mi yani karşılaştırmayı sadece plasebo alan complete'le mi yapacağız, ya da N1'leri plasebo olup sonuca kadar ulaşan ve ulaşmayanla mı karşılaştıracğız? Burada *intention-to-treat*'e göre analiz etmişiz, çünkü kayıplar olabilir ya da bu arada ekstra tedavi almış olabilirler. Doktorun bilgisi dışında gidip bir yerden ilaç kullanmış olabilirler ve bir takım kontrol edilmesi gereken faktörler olabilir. O zaman, *intention-to-treat*'te ne yapmışlar, N1 ve N2'yi prednizon alan grup tedaviye aynen devam etse de etmese de bir sayıp, bunu diğer grupla bir saymışlar, bu analizlerde bunun olması lazım.

Allokasyonda çok önemli eksikliği olduğu için biraz orayı açtık kendi sorularımızla. Allokasyonla ilgili 3 tane eksik tespit ettik. O zaman biz çalışma yaptığımızda, randomizasyon doğru yapılmış olsa dahi, bu seçimin doğru uygulanıp uygulanmadığını nasıl bileceğiz? Demin konuştuğumuz konular. Bazen biz rejimleri asıyoruz; birinci askıya, ikinci askıya. Bu son derece yanlış. Geçirgen zarflar son derece yanlış, üstü açık yapıştırılmamış zarflar son derece yanlış. Zarfların eşit ağırlıkta olması lazım, yani her türlü doktorun uyanmasından ve algısını etkileyebilecek şeylerden kaçınmalıyız. Onun için kullanılması en çok önerilen yöntemlerden biri arka arkaya numaralandırılmış, üstünde numarası olan, opak ve tamamen yapışkanı olan zarflar kullanılması. Tercihen de hasta geldiğinde üstüne ismini yazıyorsunuz ve hastanın bunu alacağını garanti edip, sonra zarfı açıp hastaya tedavisini veriyorsunuz. Bu çalışmada nasıl bir yöntem kullanıldığı konusunda bilgimiz yok. İkinci bir şey de, bunu dünkü ilk slaytta da göstermiştik, hasta gelecek, hastanın uygun olduğuna karar vereceksiniz. Böyle bir çalışma için hastayla konuşacaksınız, onayını alacaksınız, çalışmaya dahil edeceksiniz. Ondan sonra zarfı açacaksınız. Önce açarsanız, sonradan uygun olmadığını anladığımız bir grup hasta var ya, onlar atılır, bu uygun bir yöntem değil. O zaman rapor ederken nelere dikkat etmeliyiz, benim örnek büyüklüğüm ne,

seçilecek kişiler hangi özellikte insanlar, yani sonuçlarımı kime genelleleyebilirim, başlangıç özellikleri neler? Bu başlangıç özelliklerinin her birinde bir eşitsizlik varsa buradaki etkiler neler ve randomizasyon sistematik hatayla doktorun cevabını engelleyebilir, ama şans eseri her zaman hata yapabilirsiniz. Şans eseri taraf tutabilirsiniz. Onun için Tablo 1 önemli, bunun için p değerlerini veriyoruz. Hiçbir zaman istatistiğe tek başına güvenmiyoruz. Klinik değer bizim için önemli.

Birkaç tane allokasyonun gizlenmesine ait örnek çalışma getirdim, muhtemelen okumusunuzdur ama okumadıysanız mutlaka okumanızı öneriyorum. Lancet'in 2002 yılında, Ocak ayıydı galiba, bir epidemiyoloji serisi çıktı; tamamen klinisyenlere yönelik, üstelik bu konuda çok da yetkin olmayan klinisyenlere yönelik. Bunları onların içerisindeki randomizasyon kısmından aldım. Çok güzel, hap şeklinde verilen örnekler, değişik makalelerden alınmış, allokasyonu nasıl gizlediklerini gösteren güzel örnekler içeriyor. Bu makalede de beklerdik anlamında söylüyorum: Örneğin, ne yapmışlar? Random sayıları belirledikten sonra zarflar hazırlamışlar. Bu zarfları hastanenin eczacısına vermişler. İki merkezde de, hekim eczacıya telefon edip, zarfı istiyor ve o zarfı kullanıyor. Bu çok güzel bir gizleme yöntemi. İkincisi de, yine bir hastanenin eczacılık departmanında verilmiş ve kodları çalışma bitinceye kadar açmamışlar. Bu durum allokasyonun aynı zamanda körlemenin bir göstergesi. Burada güzel bir örnek var. Diyor ki "bloklü yöntem kullandık, her blokta 10 tane numara vardı, randomizasyon merkezinden bir kişi bu işten sorumlu oldu ve bu bir eczacı veya hemşireydi". Hiçbir zaman bu iş için doktoru kullanmayın. Doktor hiçbir zaman iyi bir anketör de olamaz, çünkü doktorlar olayı manipüle etmeyi severler. Böyle bir kural var, yani bu teorik değil, gerçekten açıklanmış bir kural. Eczacı ve hemşireler bu anlamda uygun kişiler: hem allokasyondan, ilacın hazırlanmasından ve ilacın infüzyonundan, hepsinden aynı kişinin sorumlu olduğu, infüzyon malzemelerinin başka bir merkezde hastaya uygulandığı ve uygulanan ilacın ne olduğunun hiç kimse tarafından bilinmediği çalışma-

lar var. Bunlar çok güzel örnekler. Bu nedenle bu şemayı hazırladık. Her yayını okurken şu sorular sorulmalı: Kişiler hangi aşamada dahil edilmişler, dahil edilme kriterleri, uygunluk kriterleri uygun muydu, bilgilendirilmiş olur alınmış mıydı ve çalışmaya ne aşamada gelmişler, ne tür bir yöntemle allokasyon yapmışlar? Biraz önce de söyledim; bunların hiçbirini bu çalışmada yok onun için burada vurgulamaya çalıştık.

Şimdi çok kafa karıştıran bir konu: Bir çalışmanın kör olması, allokasyonun gizlenmesinden (*allocation concealment*) nasıl farklıdır? Bu hep karışıyor. Rastgele sayıları tahmin edebiliyoruz. İşte 1, 3, 8, 15, 21. kişileri tedavi grubuna aldık, kalanları plasebo grubuna aldık gibi. Bir karar alacağız, bu kararın adı rastgele sekans seçimi. 1, 3, 15. kişilerin gerçekten bu tedaviyi almaları ve alıncaya kadar da hangi ilacı aldıklarının hiç kimse tarafından bilinmemesi, allokasyonun gizlenmesi. Buraya kadar her şeyi doğru yaptık ama bundan sonra çalışmanın sonuna kadarki aşama ve uygulama aşamasındaki körleme: Eğer ikisinden sadece birini seçecekseniz, allokasyon gizlemenin doğru yapılması lazım, körlemenin değil. Körleme yapılmasına rağmen double blind hatta triple blind bir çalışmada allokasyon gizlemesi yanlışsa hala hata yapabilirsiniz, ama allokasyon gizlemesi doğru ise ve kör değilse, işte bu yan etkiler hastanın durumunu etkilediği için, daha az hata yaparsınız. Bizim randomize kontrollü çalışmalarda sık olarak aradığımız özellik çalışma kör mü, değil mi? Oysa bundan daha önemlisi allokasyon dediğimiz şey yani hastalara ilacın verilmesi, gruplara ayrılma aşaması ya da gruplara gerçekten hangi tedavi ya da plasebonun verilmesi aşaması kör olmaktan çok daha önemli, bunu vurgulamak istiyorum. Ama kör olması da önemli tabii. Körlemeyle ilgili de herkesin farklı bir yorumu vardır. Bir dönem maskeleymeyi de Türkçe’de çok kullandık. Ama genel kanı artık maskeleymeyi kullanmamak, kör diyelim çünkü maskeleyme açık mı, yarı kapalı mı demek, karmakarışık oluyor. Körleme için de herkesin uygun bulduğu bir isimlendirme sıralamasını buraya koydum. He-

pimiz aynı şeyden aynı şeyi mi anlıyoruz? *Open blind*’la *open label* aynı şey. Özellikle ilaç kullanımı varsa biz *open label* diyoruz. Koriyucu hekimlik uygulamalarında özellikle açık çalışma da diyebiliriz ya da korumaya yönelik klinik çalışmalarda tek kör de. Sadece bir tek grup kör, dolayısıyla bu grubun tercihen hasta olduğunu varsayıyoruz. Bu çalışma çift kör, çalışmanın içerisinde bir yerde bu hekimleri nasıl körlediğini söylüyor. Bu arada ilginç, hastaların nasıl körlendiği konusunda hiçbir fikrimiz yok. Yani bu prednizon ve plasebo nasıl uygulandı, nasıl verildi, insanlar bunu anladı mı anlamadı mı? hiç bilmiyoruz. Çift körde hem hastalar hem araştırmacılar ya da bunu değerlendirenler kör oluyor. Örneğin, hastanın EKG’sini okuyan ya da sonucunu bildiren kişi, genellikle araştırmacıyla sonucu değerlendiren aynı kişi olduğu için çift kör diyoruz, ama bunlar ayrıysa bile yine çift kör demeye özen gösteriyoruz. Eğer üçlü kör diyorsak analizi yapan kişi de kör demek. Bu arada buna dörtlü körleme diyenler de var ama genel görüş üçlü olması gerektiği konusunda.

Körleme neden önemlidir? Tamamen insan faktörünü ortadan kaldırmak için. Bu çalışmada güzel bir şey yapmışlar, fizik muayeneyi yapan kişiyle, serolojik markerleri ölçen kişiyi farklı kılmışlar. Aktif plasebodan bahsettik, çift kör uygulaması yine bazı yerlerde gündemde olabilir. Eğer prednizonun yerine ona benzer bir şey koyamıyorsanız, iki tane dummy kullanıp iki farklı şekilde kullanabilirsiniz. Bunları da yine vurgulamak çok önemli. Bir tane de güzel körleme örneği getirdim; plasebo kontrolü çift kör bir çalışma yapmışlar. Dolayısıyla ne doktorlar, ne hastalar plasebonun aktif ajandan farkını anlayabilmişler. Kapsül verilmiş, kapsüllerin tadı aynıymış ve aynı şekilde kullanılmışlar. Biyolojik analizler de, yani değerlendirme de randomizasyon kodu açıklanmadan sonrasına kadar tamamlanmış; dolayısıyla güzel bir kör çalışma. Yine güzel bir örnek, bir yoğun bakım ünitesinde yapılan bir çalışma bu; çalışma boyunca hasta, hemşire, araştırmacı ya da hiç kimse bilmiyormuş. Aynı zamanda burada istatistiksel analizler de kör

olarak uygulanmış, yani bu bir üçlü kör çalışma. Sonuna kadar, randomizasyon konusundaki bir takım çalışmaları yapmışlar ve demişler ki randomizasyonun ya da körlemenin açıldığını gösteren en ufak bir bulgu varsa mutlaka bunu dosyalayın demişler. Bu yapılmadığı için onun da doğru olduğunu söylüyorlar.

H. Yazıcı: Doktorlara olan sataşmayı, istatistikçilere çevirmek istiyorum şimdi. Banu, yeni biten bir çalışma var, yurtdışında birileriyle yapılan limite bir çalışma ve tabii ben tutturdum, mutlaka ve mutlaka sonuçlar da kör değerlendirilecek diye. Amerikalı istatistikçiden, eğer gruplara kör olarak bakacaksak, o zaman alacağım ücret artar, diye bir yanıt geldi. Lütfen şuna üçlü kör demeyin, zaten analiz kör olarak yapılır, grubu bilmezsin, çift kördeki olay da odur. Üçlü kör deyince, fiyat listesinde yükselme oluyor.

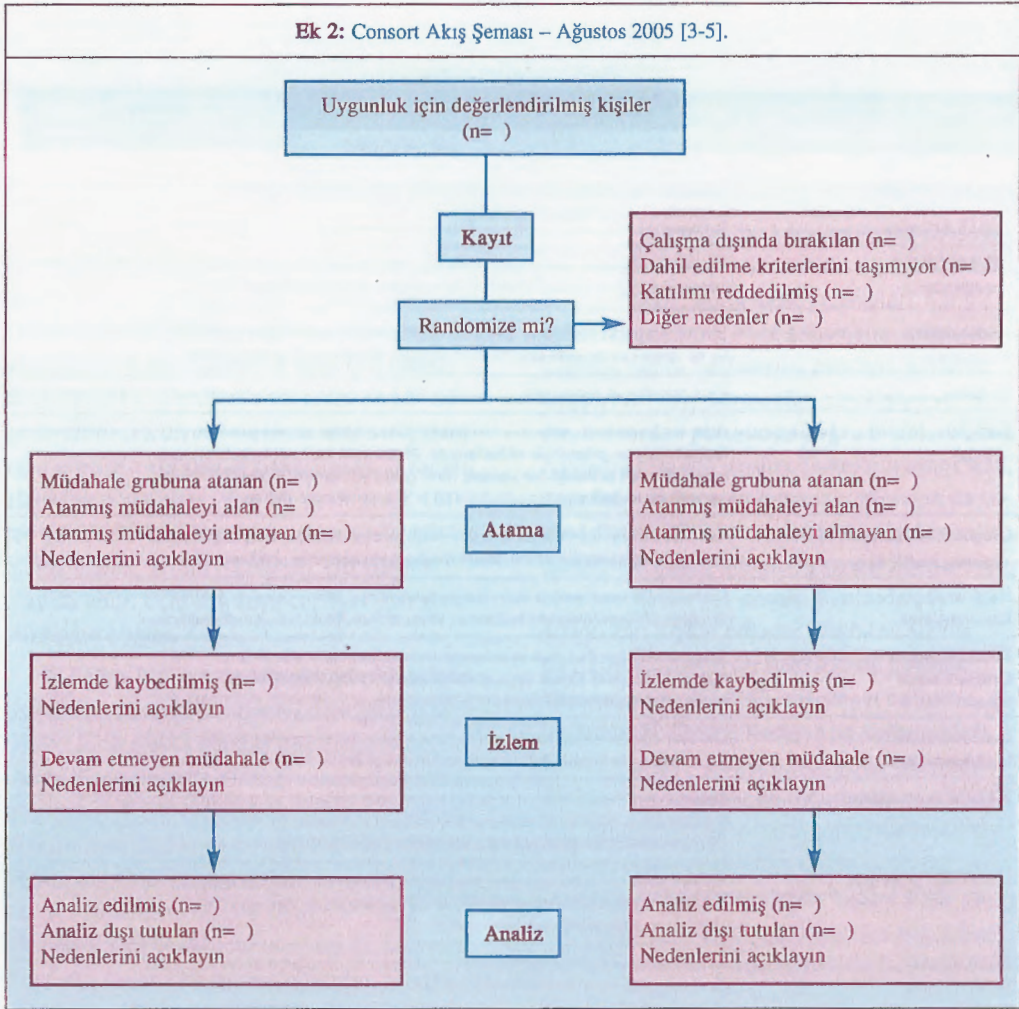
B. Çakır: Bu konuda kitaplar yazılmış, istatistiklerle nasıl yalan söylenir diye, yani istatistikçi kötü niyetli olmasa bile araştırmacının içinde bir sonuç çıksın isteği vardır. Ben şimdi bir dermatolog arkadaşla çalışıyorum. Patoloji preparatlarına bakıyor ve deri lezyonlarıyla ilişkilendirmeye çalışıyor. “Bir de şunu denecek acaba anlamlı bir şey çıkar mı?” diyor yani o kadar istekli ki. Ben de diyorum ki, bu yan tutmadır. “Hayır, bu benim sonradan aklıma geldi” diyor. Ama o bizim ilk aşamada aklımıza gelmiyorsa, sonradan gelmemesi tehlikeli. Çünkü ne kadar çok ilişki çalışırsak er geç birinde bir özellik yakalarız. Ancak, neye bakacağımızı çalışma başında belirlemeliyiz. İstatistikçi hangi ilişkinin çalışıldığını, hangi ilacın etkisinin araştırıldığını bildiğinde ister istemez ilişki bulamadıkça analiz sayısını artırıyor, kullanılan yöntem artıyor, ya da “ben bunu parametrik yapmıştım, non-parametrik mi yapсам acaba farklı bir şey mi çıkar” diyor. Bu çalışmada da var, aslında onu gündeme getirmek istiyorum; parantez içinde Fisher’s testte bu çıktı, Log-Rank testle bu çıktı diyor. Farkındaysanız Log-Rank testi gücü biraz arttırabileceği için p değeri biraz daha anlamlı. Orada bir zorlama var aslında çünkü Fisher’de 0.08 çıkmış, bir de Log-Rankle bakıyor 0.07, belki 5 çıkarsa,

sadece onu yansıtacak. Ama eğer mantık gücü arttırmaksa ve bu çalışma kohort, randomize klinik bir çalışma ise, o zaman analizde niye baştan Log Rank test yapmadınız?

Bu da son slaydım. Yeni bir alışkanlığımız ile ilgili: Kör bir çalışma yaptığımızda değerlendirme yapanlar, bu kör yaptığımız çalışmada, körleme yeterince iyi midir diye, bakmamızı istiyor artık. Bunu nasıl yapmışlar? Kör bir çalışmadan sonra, önce çalışmanın araştırmacılarına, sonra hastalarına ayrı ayrı, körleme uygun mu? diye sormuşlar: 126 kişiye sormuşlar, bunlardan plaseboyu doğru bilenlerin, ben hastama plasebo verdim diyenlerin oranı %18, ama %69 da karar verememiş, yani ona da tedavi diyememiş. Aynı şekilde 68 kişiden 16’sı plaseboyu doğru bilmiş, 13 tanesi bilememiş, %71’i de fikrim yok demiş. Bunlarda şöyle bir sıkıntı var: Kişiler bilseler bile bunu sizinle paylaşmak istemiyor; özellikle de hekimler. Yani çalışma çift kördü, çalışmayı bitirdiniz, diyorsunuz ki, “hangi hastaya ne uyguladığını biliyor musun?” Aslında ben bunu çalışmada körlemenin ne kadar iyi olduğunu tayin etmek için yapıyorum, ama bilmiş olsa bile bunu sizinle paylaşmak istemiyor. Çünkü bildiği zaman belki de şaşırtmışımdır hissini sizde yaratmamak için. Bizim sorularımız kabaca bunlardı. Eğer kendi çalışmalarınızla randomizasyon nasıl olur, faktöriyel designlar, ön randomizasyonlar gibi daha detaylı bilgi isterseniz, Laçin’in kitabı çok güzel bir kitap, okumak açısından öneririm, ama dediğim gibi özellikle eğitim aracı olarak da bu seriyi söyledim, birden çok var. Bu da kitap olarak hepinize önerilebileceğim bir kitap. Kullanır mısınız bilmiyorum. Ben Elwood’u yakın bir zamanda keşfettim. Klinisyenlerle, özellikle de kanıta dayalı tıp yapanlarla çalışırken çok hoşuma giden, okuması çok kolay, örnekleri güzel bir kitap. 2007 basımı. Kanıta dayalı uygulamalarda ya da işte makale saatlerinizde asistanlarınızla, yeni uzmanlarınızla paylaşırken çok kolay okunabileceğine inandığım bir kitap. Teşekkür ediyorum.

Rapor Bölümü ve Başlık	Madde	Açıklama	Raporlanan Sayfa No
BAŞLIK VE ÖZET	1	Katılımcıların müdahalelere nasıl dağıtıldığı (örn. "rastgele dağılım", "randomize" veya "rastgele atanmış").	
GİRİŞ Arkaplan	2	Bilimsel arkaplan ve mantığın açıklaması.	
YÖNTEMLER Katılımcılar	3	Katılımcılar için uygunluk kriterleri ve verilerin toplandığı yerler ve düzenlemeler.	
Müdahaleler	4	Her bir grup için planlanan müdahalelerin tam detayları ve gerçekte nasıl ve ne zaman uygulandıkları.	
Hedefler	5	Özel hedefler ve hipotezler.	
Sonuçlar	6	Açık bir biçimde tanımlanmış birincil ve ikincil sonlanım noktalarının ölçümleri ve uygulanabilir olduklarında, ölçümlerin kalitesini güçlendirmek için kullanılan herhangi bir yöntem (örn. çoklu gözlemler, ölçüm yapacakların eğitimi).	
Örneklem büyüklüğü	7	Örneklem büyüklüğünün nasıl belirlendiği ve uygulanabilir olduklarında, herhangi bir ara analiz ve çalışmayı durdurma kriterlerinin açıklaması.	
Randomizasyon – Dizi oluşturma	8	Herhangi bir kısıtlamanın detayları da dahil olmak üzere, rastgele dağılım dizisinin oluşturulmasında kullanılan yöntem (örn. bloklama, tabakalandırma)	
Randomizasyon – Gruplara atama	9	Rastgele dağılım dizisinin uygulanmasında kullanılan yöntem (örn. numaralandırılmış kaplar veya merkezi telefon) - müdahaleler dağıtılana kadar dizinin gizlenip gizlenmediği netleştirilerek.	
Randomizasyon – Uygulama	10	Dağılım dizisini kimin oluşturduğu, katılımcıları kimin kaydettiği ve katılımcıları ilgili müdahale gruplarına kimin atadığı.	
Körleme (maskeleye)	11	Katılımcıların, müdahaleleri yöneten kişiler olup olmadığı ve sonuçları değerlendirenlerden grup atamalarının gizlenip gizlenmediği. Eğer yapıldıysa, körlemenin (blinding) başarısının nasıl değerlendirildiği.	
İstatistiksel yöntemler	12	Birincil sonlanım noktaları açısından grupları karşılaştırmakta kullanılan istatistiksel yöntemler; ek analizler için kullanılan yöntemler, örneğin alt-grup analizleri ve ayarlanmış analizler	
SONUÇLAR Katılımcı akışı	13	Her bir aşama boyunca katılımcı akışı (mutlaka bir diyagram önerilmektedir). Özellikle, planlanan tedaviyi alan, çalışma protokolünü tanımlayan ve birincil sonuç için analiz edilen, rastgele atanmış katılımcıların sayısını her bir grup için raporlama. Nedenleriyle birlikte planlandığı biçimde çalışmalardan protokol sapmalarını açıklama.	
Katılımcı Seçimi	14	Katılımcı seçimi ve izlem dönemlerini tanımlayan tarihler.	
Temel veriler	15	Her bir grubun temel demografik ve klinik özellikleri.	
Analiz edilen	16	Her bir analize dahil edilen her bir gruptaki katılımcı sayısı (payda) ve analizin "tedavi amaçlı" olup olmadığı. Uygulanabilir olduğunda, sonuçları mutlak sayılarla belirtin (örn. %50 değil 10/20 şeklinde).	
Sonuçlar ve tahmin	17	Her birincil ve ikincil sonlanım için, her bir grubun sonuçlarının özeti ve tahmini etki boyutu ile doğruluğu (örn. %95 güven aralığı).	
Yardımcı analizler	18	Önceden tanımlanmış olanlar ile hipotez yaratmaya yönelik olanları gösteren, alt-grup analizleri ve uyarlanmış analizler dahil, yapılan diğer herhangi bir analiz raporlayarak çeşitlilik gösterme.	
Olumsuz gelişmeler	19	Her bir müdahale grubundaki tüm önemli olumsuz gelişmeler ve yan etkiler.	
TARTIŞMA Yorumlama	20	Çalışma hipotezlerini, potansiyel önyargı veya beklenmeyen durum kaynaklarını ve analizlerle sonuçların çeşitliliğine ilişkin tehlikeleri dikkate alarak, sonuçların yorumlanması.	
Genellenebilirlik	21	Çalışma bulgularının (dış geçerliliği) genellenebilirliği.	
Ayrıntılı kanıt	22	Mevcut kanıt bağlamında sonuçların genel yorumu.	

Ek 2: Consort Akış Şeması – Ağustos 2005 [3-5].



1. Çalışmada random allokalasyon sekansını belirlemede hangi yöntem kullanılmış? Detaylı ve uygun tanımlanmış mı?
2. Toplam örnek hacmi nedir? Randomizasyon oranı nedir?
3. Eğer bire-bir randomizasyon oranı kullanılmış ise, bunun sebebi ne olabilir? (Toplam örnek hacmi? Teknik hangisi? Nasıl rapor edilmiş?)
4. Burada sınırlama var mı? (*restricted randomization* mu?)
5. Bu çalışmada tabakalı randomizasyona ihtiyaç var mı? Nasıl incellersiniz?
6. Tedavi ve kontrol grupları karşılaştırılabilir özellikte mi? Nasıl incellersiniz?
7. Bu çalışmada prognostik faktörler nelerdir? Eğer prognostik faktörler açısından gruplar benzer değilse neler yapılmış/yapılabilir?
8. (*Permuted block randomization* tasarımı kullanıldı ise) Blok büyüklüğü nedir? Değişkenlik gösterir mi?
9. Klinik çalışmalarda "*minimization*" ile "*randomization*" arasındaki fark nedir?
10. Hastalar hangi aşamada gruplara ayrılmış? Bu neden önemli?

Soru 11 and 12'yi yanıtlarken, allokalasyonu kimin hazırladığı, katılımcıları çalışmaya alanın kim olduğu, katılımcıları gruplara ayıranın kim olduğunu, randomizasyonun yapılmasından uygulanmasının sonuna kadar geçen (aşağıdaki tabloda özetlenen) süreçleri ayrı ayrı ele alarak yanıtlayınız.

Randomizasyonun yapılması	Uygulanması
Rasgele bir dizi oluşturulması	Katılımcılar çalışmaya alınırken: • Uygun olmalarını kontrol etme • Çalışma hakkında bilgi verme • Bilgilendirilmiş onam alınması • Çalışmaya dahil etme
Hastaların gruplara ayrılmasının sağlanması (örn. kodlanmış ilaç kutuları, zarflar vb.)- bunların içinde ne olduğu bilgisinin, kişileri gruplara kabul eden (ayırıcı) kişi(ler)den saklanması tercih sebebidir.	Tedavi seçeneklerinden uygun olanı kişiye verme (örn. bir sonraki zarfı verme)
	Müdahalenin uygulanması
11. Çalışmada allokalasyonun yeterince gizlendiği görüşünde misiniz? Neden? 12. Allokalasyonun gizlenmesi hakkında makalede yeterince bilgi verilmiş mi? 13. Klinik çalışmalarda "<i>blinding</i>" ile "<i>allocation concealment</i>" birbirinden nasıl farklıdır? Hangisi çalışmada taraf tutmak açısından daha yüksek potansiyele sahiptir? 14. Yazarlar çalışmada kim kör, bu nasıl sağlanmış, açıklıyor mu? Tedavilerin benzerlikleri nasıl? Eğer varsa bu çalışmada kör olmayan kimler? 15. Yazarlar körlemenin başarısı konusunda bir bilgi vermiş mi? Bu durum nasıl incelenebilir? Körleme başarılı mı, değil mi konusunda yazarlar makalede yeterli bilgi sağlanmış mı?	

KAYNAKLAR

1. Starnd V. Lessons learned from clinical trials in SLE. *Autoimmun Rev* 2007; 6: 209-14.
2. Tseng CH, Buyon JP, Kim M, et al. The effect of moderate-dose corticosteroids in preventing severe flares in patients with serologically active, but clinically stable, systemic lupus erythematosus. *Arthritis Rheum* 2006; 54: 3623-32.
3. <http://www.consort-statement.org/consort-statement/citing-and-using-consort/>
4. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010, 8:18.
5. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, for the CONSORT Group. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trial. *BMJ* 2010; 340: c869.

İTERAKTİF OTURUM II

N. AKKOÇ, Ö. ERGÖNÜL

Dokuz Eylül Üniversitesi Tıp Fakültesi
Romatoloji-İmmünoloji Bilim Dalı,
İzmir

Marmara Üniversitesi Tıp Fakültesi
Enfeksiyon Hastalıkları ve Klinik
Mikrobiyoloji Anabilim Dalı, İstanbul

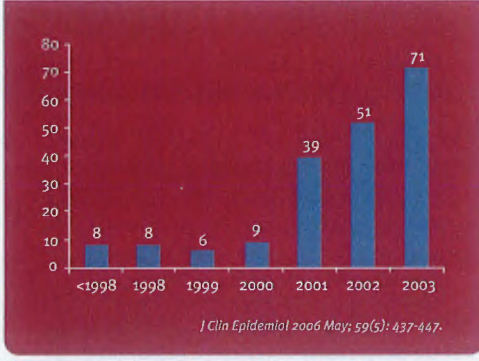
İteraktif Makale Tartışması (II)

N. Akkoç: Bu oturumda sizlerle *propensity* analizini tartışabileceğimiz bir makaleyi değerlendireceğiz. Şekil 1’de 1990 sonrasında *propensity* analizinin kullanıldığı çalışmaların sayısının yıllar içerisinde nasıl arttığını görüyoruz (Şekil 1).

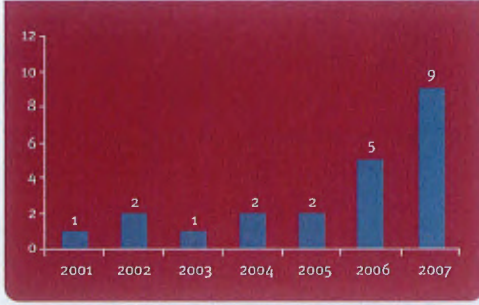
Ben de romatolojideki çalışmalar için çok kabaca böyle bir değerlendirme yaptım. Bu alanda *propensity* analizinin kullanıldığı ilk makale 2001’de çıkmış, ki bizim bugün tartıştığımız da bu makale [1]. Gördüğünüz gibi 2007’ye kadar bu makalelerin sayıları gittikçe artıyor (Şekil 2).

Genellikle bu makaleler romatoid artrit ve lupusla ilgili. Romatoid artrit ile ilgili makaleler çoğunlukla radyolojik progresyon ve sakatlığı ölçüyor ve kayıt sistemlerinde kullanılıyor.

Bizim tartışacağımız makale ilk yayınlanan, *propensity* skorunun romatolojide kullanıldığı makale. Makalenin amacı DMARD ve/veya steroid kullanan hastalarda, 5 yıllık sakatlık gelişme riskinin tedaviyle ne ölçüde değiştiğini göstermek. Metod için Norfolk Registry’den 384 hasta alınıp 5 yıl boyunca takip ediliyor ve bu hastaların tedavi detayları ve HAQ skorları yıllık olarak kaydediliyor. Daha sonra başlangıçtaki faktörlerde DMARD’lar ve/veya steroidlerin başlanmasıyla ilişkili olarak başlangıçtan sonra 12 ay içinde oluşan değişikliklerin modellenmesi için lojistik regresyon kullanılıyor. Bu modele dayanarak da bir *propensity* skor geliştiriliyor; aslında *propensity* skor her hasta için tedavi grubuna girme olasılığını ifade ediyor. Kayıt sistemindeki bir hasta için *propensity* skoru 0.75 ise, bu o hastanın %75 tedavi alma şansı olduğu anlamına geliyor. Sonra da bu *propensity* skor kullanılarak hastalık şiddetindeki farklılıklar ayarlanıp tedavi alan ve almayan hastalarda sakatlık odds’ları ($HAQ > 1.00$) ikinci bir modelle karşılaştırılıyor. HAQ’lar gruplandırıldıktan sonra, 5. yılda tedavi alan ve almayan gruplarda, bir de değişik tedavi grupları arasında, değişik zamanlarda tedaviye başlayan gruplar arasında bir farklılık olup olmadığına bakılıyor.



Şekil 1: Yıllara göre Pubmed ve Science Citation Index'te karşılaştırılan faktörlerin kontrolünü sağlamak amacıyla propensity skoru kullanılan medikal araştırmaların sayısı



Şekil 2: Yayınlanma yıllarına göre Pubmed'de propensity skoru kullanılan medikal araştırmaların sayısı.

Sonuçta bulunanlar şöyle: Ayarlanmamış analize göre tedavi alan hastalar almayanlara göre daha kötü bir sonuca sahipler. Ancak sonuçlar propensity skoru kullanılarak hastalık şiddeti için düzeltildiği zaman, erken dönemde, yani bulguların başlamasından sonraki 6 ay içinde tedavi alan hastalarda, 5. yılın sonunda gelişen sakatlık, hiç tedavi almayanlara benzer bir sıklıkta oluyor. Buna karşılık 6 aydan sonra tedaviye başlanırsa bu kişilerde 5. yılda sakatlık gelişme riski iki kat daha fazla olarak belirleniyor. Sonuç olarak propensity skorunun, gözlemsel çalışmalarda 'confounding by indication' ı düzeltmek için faydalı bir yöntem olduğu söyleniyor. Bu çalışma aynı zamanda, erken dönemde DMARD veya steroid başlanan hastalarda, 5. yılda bir sakatlık gelişme riskini, klinik olarak tedaviye ihtiyacı olmadığı düşünülen hastalarinkine benzer bir düzeye azalttığını da gösteriyor.

Ö. Ergönül: Böyle bir girişten sonra dağıttığımız föydeki toplam 11 tane soruyu takip ederek tartışmayı açalım. Bu yayında en önemli şeylerden bir tanesi aslında hipotezin çok net ifade edilmesidir. Gerçekten de materyalleri değerlendirirken, çalışmanın hipotezi ve amacının ne olduğunu, bir de çalışmanın nereye oturduğunu, hangi ihtiyaçlardan kaynaklandığını anlamak çok önemli, çünkü sonuçlarda olumlu bir veri çıkıp çıkmayacağını bu açıdan değerlendirmek çok anlamlı olacaktır. Sizce, yani romatolog olarak böyle bir şeye ihtiyaç var mıdır, nedir bu mesele, kendi klinik pratiğinizde karşılaştınız mı?

H. Yazıcı: Bence önce hipotezin tam ne olduğunu ortaya koymak lazım, sence tam olarak nedir?

N. Akkoç: Aslında şöyle, başlıktan gidelim "reduced disability at five years with early treatment of inflammatory polyarthritis". Yani bir dizi romatolojik hastalık için, burada romatoid artrit için erken tedavinin etkisi var mı?

H. Yazıcı: Şimdi işin felsefesine de biraz girmek gerekirse, tüm çalışmanı varsayıma göre düzenleyeceğine göre, buradaki varsayım "romatoid artriti erken tedavi etmek iyidir" diyelim. Şimdi sorun şu aslında: Varsayımların

pozitif olarak söylenmesini öngörüyoruz ki sonra düzenlemeyi onu çürütmek yönünde yapalım. Bunun için iyi değil mi *propensity* analizi? Yani romatoid artritte erken tedavi iyidir dersek, kontrollü çalışmamızı yaparız. Bu pozitif olarak ortaya koyduğunuz varsayımı çürütmek için *propensity* analizi modelini ona göre ayarlamışlar mı burada?

N. Akkoç: İşte onu da irdedeceğiz burada. Şöyle ki, *propensity* analizi burada uygulanacak olan-klasik lojistik regresyona ek bir modifikasyon getirerek daha doğru bir sonuç alınmasına yönelik ek bir şey aslında.

H. Yazıcı: O zaman, romatoid artritte erken tedavi etmenin, erken tedavi etmemekten iyi geldiğini göstermek lazım. Eğer onu çürütebilirseniz o zaman erken tedavi etmek iyi mi gelir? Yani *propensity* analiziyle onu değerlendirebiliyor muyuz?

N. Akkoç: Kısım. Ama kaldığımız yerden devam edersek *propensity* analizi daha çok negatif sonuçlar, dolayısıyla birçok çalışmada gösterilememiş fark olduğu için kullanılıyor. Bu farkın gösterilememesinin nedeni bu hastaların zaten ağır hastalar olması. Dolayısıyla böyle bir düzeltmeye ihtiyaç var, düzeltmeyi yaptıktan sonra bunu gösterebiliriz. Aslında *propensity* analizi kullanmalarının nedeni de bu.

Ö. Ergönül: Amacımız inflamatuvar artritlerde erken tedavinin iyi olup olmadığını göstermek olduğunda, bunların hepsini bir yana bırakalım, *propensity* veya başka bir şey düşünmeyelim, böyle bir sonuca gitmek için acaba nasıl bir düzenleme var? Nasıl bir tasarım var? Siz olsaydınız ne yapardınız? Belki bunları tartışmakta fayda var. Kohort mu, olgu kontrol mü yapardınız veya ne gibi sıkıntılar olabilirdi? Gerçekten de böyle çalışmalar çok yapılmış, tatmin edici olmadığı için sonucunda *propensity* analizi eklenmiş. Diyelim inflamatuvar artrit ile gelen hastaların bir kısmına tedaviyi erken başlادınız, bir kısmına geç başlادınız. Böyle bir kohort olabilir mi, yoksa olgu kontrolden giderek yani tersten giderek, önce çoğunluğu elde edip, tedavi başlanmış, ne zaman başlanmış gibi kafamızda geriye dönüş olabilir mi?

N. Akkoç: Peki, randomize yapılabilir mi; randomize kontrollü?

Katılımcı: Bence buradaki sorun sürenin uzunluğu, bu çalışmalarda yeterince kayıp olmadan 5 yılı tamamlayamıyoruz. Benim bildiğim elimizde 5 yıllık izlem süreli çalışma sayısı oldukça az; Sirocco gibi bir iki tane var. Bu, eldeki açık kohortun kullanılmasını sağlıyor. Bu amaca yönelik olarak, yani verinin artmasını sağlıyor. Bence buradaki temel nokta bu olabilir; randomize kontrollü bir çalışmayla, uzun dönemdeki sakatlık doğurucu bir sonucu ortaya koymak hemen hemen mümkün değil. Böyle bir çalışmayı finanse etmek de mümkün değil, hele karışık ilaç grubuyla finanse etmek hiç mümkün değil. Onun için bir kohort olacağı kesin ama burada kohortlar genellikle karmaşık bir hasta grubundan oluşacağı için araştırmacılar sonuçta ne tip bir tedaviden yarar görecektir alt grubu belirlemeye çalışıyorlar aslında. *Propensity* skorunun getirdiği şey bence bu.

Ö. Ergönül: Peki, randomizasyonda ne gibi sıkıntılar olabilir?

Katılımcı: Yani formal bir randomize kontrollü çalışmadan mı söz ediyorsunuz? Bir kere tedavi standardizasyonu büyük bir sıkıntı. Bu kohortu ben de biraz incelemiştim. Oldukça erken, romatoid artrit kriterlerini, sınıflandırma kriterlerini doldurmayan, bazı hekimlere göre klinik olarak da romatoid artrit denilmeyecek hastalardan oluşuyor. Dolayısıyla tedavi standard değil, dolayısıyla bunlara şunları verin denemeyecek bir hasta grubu. Bu biraz hastalığın erken oluşundan, biraz da hasta grubunun karışık olmasından olabilir. Yani, karışık tanılarından oluşan bir grup bu, onun için şimdi burada kontrollü bir çalışma yapmak, yani homojen bir grup elde etmek çok zor.

N. Akkoç: Peki, klasik düzeltme yöntemlerden biri kullanılamaz mı? Düzeltme için *propensity* dışında başka bir istatistiksel yöntem kullanılamaz mı?

H. Yazıcı: İşte bence esas sorun bu. Yani ben *propensity* analizini hiç duymamış olsaydım, bana bu veriler gelseydi, ben ne yapardım diye düşünüyorum. Tamam, bazılarında HAQ'ı düşük olduğu için, bazılarında bir sürü başka

nedenle tedavi başlanmamış. Ben alırdım bunları, şöyle bir bakardım, ondan sonra bilgisayar vb kullanmadan, bunları kendi içinde, alanlar ve almayanlar olarak benzer bir hale getirmeye çalışırdım. Ondan sonra ne olmuş diye onların arasında bir alt analiz yapardım.

N. Akkoç: Stratifikasyonu mu kast ediyorsunuz?

H. Yazıcı: Evet, kendi aralarında bir alt grup analizi yapardım ve o alt grup analizlerine göre bir yerlere gitmeye çalışırdım. Benim *propensity* analizi öğrenelim dememin nedeni, felsefe bakımından farkının ne olduğunu anlamak.

Ö. Ergönül: Dünkü örneğimizde sigara ve MI ilişkisine bakmıştık. Orada iki değişken vardı. Hipotez oral kontraseptifin MI'ye neden olup olmadığı ile ilgili idi. Cümle olarak söylemek gerekirse hipotez, "oral kontraseptifler MI riskini arttırır" idi. Bir yandan da aralarında sigara içen ve içmeyenler vardı. O zaman ne yapıyorduk? Sigara içenleri ve içmeyenleri ayırıp bir daha bakıyorduk. Az sayıda bağımsız değişken olduğu için o iş daha kolaydı ama burada çok sayıda değişken var. Bunu tek tek yapmak oldukça güç bir işlem. Toplantı sonunda *propensity* analizinin stratifikasyondan nasıl bir farkı olduğuna geleceğiz.

İsterseniz şimdi kohort üzerine konuşalım. Dünyada en iyi bilinen kohort 1950'lerde Amerika'da başlayan Framingham Kalp Çalışması'dır. Bu çalışma, Boston yakınlarında bir kasabada, 1950'li yıllarda, 10.000 kişiyi dahil ettikleri bir çalışma. Kohort olduğu için, önce bütün insanları kaydediyorlar; kayıt sistemi çalışması. Aynen burada olduğu gibi bir kayıt sistemi var, tabii ki daha sonra bu insanların bir kısmı MI geçirecek, inme geçirecek, bir kısmı ölecek. Beklersek bunları göreceğiz diye giden bir çalışma. Kişilerin yapmış oldukları faaliyetler, yeme alışkanlıkları, içki tüketimi, egzersiz gibi noktaları, bir kitapçık gönderilerek dolduranlara belli bir para ödemek şartıyla 6 ayda bir bildirmelerinin sağlandığı ve Boston'da ayrı bir binada, içerisinde yüzlerce insanın çalıştığı bir analiz bürosuyla yürüyen bir kohort. Hala da dünyanın en eski kohortu olarak biliniyor

ve tartışmasız en üstünü, ama tabii romatolojideki hasta sayısı daha az.

Katılımcı: Framingham kohortu o bölgede, o toplumda yaşayan, böyle kriterleri olan her türlü hastayı içerdiği için önemli ve bunları uzun sürede izlemeyi başardığı için çok da bahsedilmiş bir kohort. Tahmin ediyorum 20'ye yakın yayın var bu kohorttan çıkmış olan. İskandinav sağlık sisteminde buna benzer kohortlar var. İsveç'te, Finlandiya'da, bir de Hollanda'da böyle erken artrit kohortları var. Onun dışında Avrupa'nın güneyinde ve doğusunda bunu yapmak çok zor, çünkü çok yer değiştirmeyen bir popülasyon gerekiyor.

Soru: Örnek çalışmada, başlangıç grubundan 384 deneğe inilmiş ve bu kohortta analiz yapılmış. Kontrollü çalışmadaki gibi *intention-to-treat* yapılamaz mıydı?

Ö. Ergönül: *Intention-to-treat* tasarımı yeri gelmişken değinmek gerekir. Küçük bir grup çalışmadan ayrılmış. Toplam 37 hasta, %7.9'u ölmüş, 35 hasta yani %7.5'i çalışmayı terk etmiş ve %2.3'ünde ne olduğu belli değil; yani ayrılan toplam %17. Yüzde 87.3'ü ise duruyor. *Intention-to-treat* gibi, yani protokol dışında kalan hastalar da değerlendirilmiş.

H. Yazıcı: Şimdi baktığınız zaman aslında bu çalışma inflamatuvar poliartritin prognozu na bakmıyor, romatoid artrit prognozuna bakmaya başlıyor. Çünkü bunların bir bölümünü çıkartıyorsunuz ve sonunda elde ettiğiniz salt romatoid artritler. Halbuki dünyada, mesela Hollandalılar'ın yaptığı metotreksat çalışması gibi, erken inflamatuvar poliartritte 30 kişilik iki grupta, bir gruba metotreksat veriyorsun bir gruba vermiyorsunuz. Bakıyorsunuz ki verdiğiniz grubun romatoid artrit gelişmesi açısından prognozu bir senede çok daha düzgün. O zaman tedaviye göre inflamatuvar poliartritin 5 yıldaki prognozuna değil, romatoid artrit prognozuna bakıyor. Katılıyor musunuz buna?

N. Akkoç: Aslında makaleyi okurken benim kendi görüşüm şu olmuştu: Bunlar romatoid artrit çalışması yapmak istediler ama yapamadılar çünkü romatoid artrit olarak yaparlarsa örneklem büyüklüğü çok küçülecekti.

Ö. Ergönül: Bir de yeri gelmişken; *intention-to-treat* analizini “niyet ettik çalışmaya” şeklinde çevirmiştim, hakikaten niyet ettik çalışmaya gibi muğlak bir şey. Başladık topluyoruz, bakalım ne olacak gibisinden. Fakat şunu unutmayalım ki, aslında o kavramlar randomize çalışmalara ait kavramlardır. Bu çalışmaysa, bir gözlemsel çalışma. Burada randomizasyona benzemeye çalışan, dün quasi randomizasyon olarak bahsettiğimiz yarı randomizasyonlu çalışma. Aslında yazarlar şunu söyleyebilirler: Neden *intention-to-treat* analizinde farklılık bulup göstermedik? Bu gözlemsel bir çalışma zaten, onun için bunu kullanacağız. Randomizasyon olursa, *intention-to-treat* ve per protokol analizin her ikisinin birlikte verilmesi gereklidir.

Katılımcı: *Propensity* analizi için olması gereken bir hasta sayısı var mı? Mesela lojistik regresyonda her bir şey için 10’ar hasta olması gibi.

Ö. Ergönül: Yine aynı şey geçerli, en az 10 gibi, ama alt grupta toplam olgu ne kadar olacak diye bir şey konamaz, bu kovaryatların sayısına göre değişir. Her bir kovaryatta 10 gibi düşünürsek sonuçta çarparak sayıyı bulabilirsiniz.

Konuşmacı: Hocamın söylediği aslında, 5 yılın sonunda diyelim ki 600 hasta var bunların bir kısmı RA ile sonuçlandı, bir kısmı psoriyatik artritle, bir kısmı bağ dokusu hastalığı ile. Mesela onların 5 yılın sonunda, tanısı belli olduğunda retrospektif olarak, belli durumlarda ne olmuş, bağ dokusu hastalıklarında ne olmuş ya da psoriyatik artritte ne olmuş gibi, tekrar alt grup analizi yapılarak hastalık gruplarına spesifik veriler elde edilir.

A. Dinç: Bizim sorunumuz RA’daki veya diğer hastalıklardaki ileri evredeki etkinlik değil ki. Sorunumuz erken bir evrede bize hasta geldiyse ve biz buna tanı koymadıysak, bu hastaya tedavi başlasak mı başlamasak mı sorumlusunun cevabını almak. Yani bu farklılaşmamış aslında, öyle düşünmek lazım.

N. Akkoç: Çok doğru, ayırt edilerek söylenmesi lazım. Ama burada şu da söylenmeli: Belki romatoid artritte erken tedavi verilmesi

önemli olabilir ama psoriyatik artritte önemli olmayabilir. Dolayısıyla o ayırımın da sonradan yapılması gerekebilir.

M. İnanç: Bu biraz da, Hasan Hocanın dediği gibi, çalışmayı yapanların felsefi yaklaşımıyla ilgili bir konu. Onların genel bir hipotezi var; Deborah Simmons ve ekibini kast ediyorum. Onlara göre bir insanın psoriyatik artrit olup olmaması veya şu olup olmamasının hiçbir önemi yok. Onların bir akış şeması vardır; şöyle yani, inflamatuvar poliartrit başlar, etrafı çevresel ve genetik faktörlerden etkilenir, psoriasis de buna dahildir, yani psoriyatik artrit çok ayrı bir antite kabul etmeyen bir gruptur.

Ö. Ergönül: Prospektif ve retrospektif kohort derken bir süreci düşünüyoruz. Sürecin bir yönü var ama zamanla çok ilgisi yok. Söyle ki, kohort çalışmalarda etkenden sonuca doğru giden bir yön var. Olgu kontrol çalışmalarda ise sonuçtan yola çıkarak, etkene doğru giden bir yön var. Ama bu birisi prospektif olacak, diğeri retrospektif olacak anlamına gelmiyor. Bu çalışmalarda ilk başta beklenen veya insanlarn yazalım diye düşündüğü şey çalışmanın prospektif kohort mu veya retrospektif olgu kontrol mü olduğudur. Yöntemi yazarken böyle bir ifade kullanmaya gerek yok, bu yöntemi incelerken zaten ortaya çıkan bir şeydir. Yani biz çalışmalarımızı yazarken, klasik bildiğimiz epidemiyolojik kavramlara göre buna bir ad koymak zorunda değiliz. Örneğin buradaki çalışmada da, okunduğunda bunun aslında bir prospektif kohort olduğu çok net olarak anlaşıyor, ama girişte, bakınız bu bir kohorttur gibi klasik tanımlamalarla ifade etmemişler. Evet, çalışma sonuçta bir prospektif kohort. Şimdi, yavaş yavaş, istatistik yöntemleri veya metod kısmını gözden geçirelim. *Propensity* modelleme nasıl yapılmış? Dün konuştuğumuz kavramsal yaklaşımın pratikteki bir uygulamasına gelmiş oluyoruz.

N. Akkoç: Belki ilk kısmı şöyle formülize edebiliriz: *Propensity* modelleme ile randomizasyon arasındaki önemli fark, randomizasyonda aslında gruplar arasında anlamlı ve anlamsız değişkenleri kontrol edebilirsiniz. Bura-

da ise sadece gözlenen değişkenler üzerinden bir “*confounding by indication*” yapabiliyorsunuz. Dolayısıyla aslında yaptığınız *propensity* analizinin en önemli kısmı doğru değişkenleri seçebileniz; hem ilişkisiz bir değişkeni *propensity* analizine sokmamanız, hem de mevcut değişkenlerden hastanın prognozunu daha doğrusu tedavi kararını etkileyecek bu tür değişkenleri işin içine katmanız. Dolayısıyla birinci aşamada yapılacak şey şu soruların karşılıklarını bulmak: Değişkenler yeterli midir, yoksa başka çıkartabileceğiniz değişkenler var mı veya otörlerin buraya eklemesi gereken değişken olabilir mi?

Ö. Ergönül: Aslında burada konuya yine başlarken dün Hasan Hoca ve Nurullah Bey’in de vurguladığı “*confounding by indication*” terimini sormak istiyorum. Bu konuda rahat mıyız? *Confounding by indication* dendiğinde kafamızda kesin bir şey oluyor mu?

H. Yazıcı: Burada istatistikçilerin, varlığını yeni öğrendiğim bir günahı var. Şimdi, burada ki *confounding by indication* gayet açık; dün konuştuğumuz gibi tedavi endikasyonu, ağrı vs. Ancak bazı kitaplarda düzgün herhangi bir *selection bias*’a, *confounding by indication* da deniyor. Mesela *work-up bias*’a da *confounding by indication* deniyor. Burada *confounding by indication* şöyle: Ağrı varsa tedavi veriyor. Halbuki başka türlü *selection bias*’lar da pekala olabilir, mesela klasik bir örnek Mayo Kliniği. Çok ağır hastalar Mayo Kliniğine uzak diye gidemiyorlar, dolayısıyla en iyi istatistikçiler Mayo Kliniğinden geliyor. Bu da *selection bias* ve *selection bias*’lara da genellikle bazı istatistikçiler *confounding by indication* diyorlar. Halbuki esasında doğrusu belki de sadece ve sadece *work-up bias*’a *confounding by indication* demek. *Work-up bias* şöyle; diyelim Behçetlilerde beyin MR’ı çekiyorsunuz. Sadece başı ağrıyanlara MR çektiğinizde, doğal olarak o Behçetlilerin genel olarak MR’ını kapsıyor, herkese MR çekmeniz lazım. O zaman sağlıklı sonuç elde ediyorsunuz. Tabii, bunun sonunda da geline sonuç *lead time bias*, yani genel bir ad yerine olanı söylemek daha doğru; *selection bias* deyip, *selection bias*’ın tipini söylemek belki daha iyi.

Ö. Ergönül: Evet tabii klinik olduğu için aslında *selection bias* gerek klinik gerek sosyal, yani geniş her alanda kullanılabilecek bir bias. Bu, *selection bias*’ların bir alt tipi. Klinik bir çalışma olursa *confounding by indication* olarak ifade ediliyor. Aslında klasik bir *selection bias*, pek çok alanda örneğini görüyoruz. Dediğiniz gibi *work-up bias* da aynı şekilde. Benimünkü örneğimde, Swan-Ganz kateteri takılması gibi. Veya Kırım-Kongo olgularında ribavirin verilsin mi verilmesin mi diye tartışılıyor. Çok ciddi, son safhaya gelmiş, doğudan sevk edilmiş, artık üçüncül merkeze gelmiş hastaya, hastalığın son noktasında, ribavirin yararlı olmayabilir. Aslında hekimlikte son derece yaygın bir şey, çünkü hekimler genelde karşılarında etkene maruz kalan veya kalmamış ama hasta olan hastaları görüyorlar, oysa etkene maruz kalan veya kalmayan ama hasta olmayan grubu görmüyorlar ve zaman içerisinde bu durum bizde bir mesleksi deformasyon yaratıyor. İşte tam da bir *confounding by indication* sorunu ortaya çıkıyor. Biliyorsunuz, biz bunu dört gözlü bir tabloda ifade ediyoruz. Diyelim ki sütunlarda hastalık var/yok olsun, satırlarda da etken var/yok olsun; toplam dört gözlü bir tablo. Biz bu dört gözlü tablonun en fazla ilk sütununu, yani hasta olarak karşımıza gelen grubu görüyoruz zaman içerisinde, ama etkene maruz kalmayıp da hasta olmayan veya etkene maruz kalıp da yine hasta olmayan insanlar doktora başvurmadıkları için onları göz ardı ediyoruz. Bu sıkça vurgulanan bir durum. Aslında ormanın bütününe bakmak gerekir. Biz zaman içerisinde hep ağaçları görmeye alışıyoruz ve oradan yola çıkarak, hatta bir olgu üzerinden yola çıkarak da, çeşitli ilişkiler kurmaya çalışıyoruz. Oysa o dört gözlü tablonun tümüne bakmadan çalışma yapmak zor.

N. Akkoç: Bu bölümle ilgili yorumu olan var mı? Ekleme, çıkarma var mı?

N. İnanc: Mesela akut fazda şiş eklem gibi birbirini etkileyecek şeylerin de konulmasının etkisi olabilir mi, Önder Bey? Örneğin sizinünkü örneğinizdeki gibi, onları elememiz gerekir mi burada?

Ö. Ergönül: Burada zaten muhtemelen o tür faktörleri olabildiğince elemişler, yani olabildiğince kesişmeyen faktörler alınmış ama elbette örneğin, sizin alanınız ancak tahmin edebiliyorum, şiş, hassas, şiş ve hassas eklemeler gibi faktörler birbiriyle ilişkilendirilebilir aslına bakarsanız.

N. Akkoç: Burada bir şey söylemek istiyorum; *propensity* analiz ile yapmaya çalıştığımız şey kişilerin en doğru şekilde tedavi aldığını bulmaya çalışmak. Dolayısıyla bunu yaparken kesişim olması çok önemli bir şey değil. Eğer sizin modelinizi daha iyi yapıyorsa siz istediğiniz kadar kesişim dahil edebilirsiniz ve aslında bu istenmeyen bir şey de değil. Önemli olan şu ki; siz burada bir model buluyorsunuz ve bu modelle kimin tedavi alma ihtimalinin en iyi olduğunu nasıl bileceğinizi araştırıyorsunuz aslında.

Ö. Ergönül: Bir de burada teknik olarak Nevsun'un sorusuna yanıt şu ikinci paragraf aslında. "Hosmer ve Lemeshow kullandık" diyorlar. Düünden hatırlarsak, Hosmer ve Lemeshow, ortaya koyduğumuz bu bağımsız kovariyatların sonucu hangi kombinasyonu en iyi temsil ediyor? Yani bunları bir dizi analiz yapmışlar, mesela birini çıkarmışlar birini koymuşlar, tek tek koymuşlar, gruplar halinde koymuşlar, bunlar arasında Hosmer ve Lemeshow katsayısı, testi en yüksek olanı sonucu en iyi belirler diye bir sonuca varmışlar. Ancak Hosmer ve Lemeshow testi son sözü söylemez. Ben, Hosmer ve Lemeshow veya başka istatistiklerden ziyade, biyolojik ilişkinin tanımını önemserim. Yoksa bunları hiç bilmeyen, aradaki ilişkiyi kuramayan bir istatistikçi, Hosmer ve Lemeshow testini mecburen yapar ve hangisinin katsayısı yüksekse onu kabul eder.

Soru: Şurada "*joint factors generated by principle components factor analysis*" diye bir söz var. Orada birbiriyle lineerlik, ko-lineerlik gösteren faktörleri de ayırıp buraya koyduk mu demek istiyorlar?

N. Akkoç: Kovariyatlardan lojistik regresyon yapılıyor, ilişkili olanlar tekrar kontrol ediliyor ve bunlardan en çok ilişkisi olanlar sıralanarak, kendi içerisinde birbirleriyle ilişkili modeller alınıyor ve bu modellerin Hosmer ve Lemeshow testiyle ne kadar uyduğuna bakılıyor.

H. Yazıcı: Yani matrisin değişkenliğini en fazla açıklayan faktörler alınıyor.

Ö. Ergönül: Kesinlikle hocam, aynen söylediniz gibi en açıklayıcı olan modeli alıyorlar.

Katılımcı: Şöyle bir şey söylemek istiyorum; o kohortta bu klinik özelliklere sahip insanların yer almış ve bu klinik özelliklere sahip olmasına rağmen bunlara DMARD başlanmamış olması gerekiyor ki bu test doğru olsun. Örneğin bir kısım hekim RF'si pozitif, romatoid nodülü olan, şiş eklem sayısı yüksek hastaları görmüş, buna rağmen DMARD başlanmamış. Bunları bulacaklar ve bunları diğerleriyle eşleştirip analiz yapacaklar.

N. Akkoç: Şimdi bir sonraki soru şuydu: Çalışmanın ortaya koyduğu sonuçlar nelerdir ve *propensity* analizi sonuçları ne ölçüde değiştirmiştir? *Propensity* analizi olmadan yaptığımız zamanki sonuçları hiç tedavi almayan kişilerle mukayese ettiğimiz zaman, değişik zaman dilimlerindeki odds oranları eğer ilk 6 ay başlanmışsa 2.06, 6-12 ayın içindeyse 6.15, 12 ayın üzerindeyse 3.81. Bu size, eğer *propensity* analizi yapmazsanız *confounding by indication*'ın önemli olduğunu gösteriyor, çünkü 5 yılın sonundaki sakatlık erken tedavi alan grupta dahi almayanlara göre daha kötü görünüyor. Eğer düzeltme yaparsanız -burada aslında metoda geçtik, önce sadece *propensity* skor için düzeltme yapmışlar- o zaman erken tedavi yaptığınızdaki odds oranı, almayan gruba göre kıyaslanabilir duruma geliyor. 6-12 aya çıkanlarda veya 12 ay üzerindekielerde hala yüksek, yani şu 6 aydan erken tedavi olanlar, burada 6-12 ay ve 12 ay üstü odds oranları hala 2 civarında. Daha sonra bir analiz daha yapıyorlar, çünkü başlangıç değişkeni olmamasına rağmen, *propensity* analizi tamamen başlangıç kovariyatlar üzerine kuruyorlar. Daha sonra görüyorlar ki analizleri yaparken eğer bu ilk 12 ay içerisinde, hastaneye başvuru varsa, bu da hastanın tedavi alma şansını etkileyebilir. Buna göre de bir düzeltme yaptıkları zaman, erken tedavi alan grupta odds oranı 0.71, 6-12 ayda 1.98, 12 ay üzerinde 2.03. Bu bir ölçüde, bu tip bir analiz yaparken, *confounding by indication*'ın göz önüne alınması gerektiğini gösteren bir sonuç. Aslında makalenin en önemli taraflarından biri budur diye düşünüyorum.

Katılımcı: Bir nokta üzerinde durmak istiyorum. Bu çalışmada hastalar 90-92 yılları arasında toparlanmış. Aşağı yukarı 5 sene izleme süresi var. Demek ki 2000'lere ulaşan veri yok burada, yani ele gelen en geç veriler 96-97'ye ait. Burada aslında şöyle enteresan bir şey de öğrenmiş oluyoruz; çalışma içinde bulunduğu ülkenin pratik şartlarını da bize yansıtıyor. Ben İngiltere'de çalıştığım için çok iyi biliyorum, romatologların %80-90'ı romatoid artrit tanısı alanlara bile steroid vermezlerdi o dönemde. Görüyorsunuz, 6 yılın altında böyle hastalara DMARD verilme sayısı artmış yani hastaların altıda biri, yedide biri. Bu o ülkenin pratiğini yansıtıyor çünkü, Türkiye'de biz bu çalışmayı (ki Türkiye romatologlarının çoğunluğu akademik ağırlıklı çalışıyor) şimdi yapmaya kalkışsak böyle bir çalışma yapamayız. Hastaların çok daha homojen bir tedavi aldığını görürüz. Halbuki o dönemde orada steroid almayan, DMARD almayan hasta bulabiliyorlardı. Şimdi yapmaya kalksalar onlar da aynı sonucu alırlar, çünkü şimdi işin içine başka markerlar, anti-CCP vs girecek. Yani bu çalışma aynı zamanda o dönemin pratiklerini bize yansıtıyor.

H. Yazıcı: Şimdi burada önemli bir şey var; sayfa 1035'e geldiğimizde. Düzeltmeden yapılan analiz tamam; onun ne olduğu çok açık. Düzelttikten sonra, *propensity* skorlarını hesapladınız ve işte bir tane en yüksek grup var, ondan sonra bir tane daha grup vs. Diyelim ki bunların kontrol grupları -ki buna referans grup diyor burada-, hiç almamış olanlar var, yani iki iki değerleri buna giriyor, öyle değil mi? Peki, onlar tam oraya nasıl giriyor? Yani günün sonunda mesela referans grubun sayısını ben bilmiyorum burada...

Ö. Ergönül: Makalenin tablo dördüne gelelim. Burada aslında *propensity* analizi gruplarının *quintillerinin* nasıl oluştuğunu anlatıyor, yani *quintile*, persantil meselesi, toplam 5 gruba bölünmüş, hafiften, ağır seyre doğru, 1, 2, 3, 4, 5 diye giden bir grup. Tedavi edilen ve edilmeyen kesimlerin nasıl seyrettiğini görüyoruz. Hakikaten *quintiller*, 1, 2, 3, 4, 5 diye gittiğinde tedavi edilenlerin sayısı 9, 19, 32, 52, 71 diye gidiyor. *Quintillerin* her biri basamak basamak, en hafiften en ciddiye doğru seyrediyor. Bunu nasıl koyuyor peki tekrar regresyo-

na? Şimdi bunlar nasıl elde edildi? *Quintillere* hastalar yerleştirildi, bir klasifikasyon yapıldı. Uzmanlar bir araya geldiler, baktılar regresyon oluşturdular, bu regresyon sonucunda her birine bu 2 bu 3 diye bir skor verdiler; tamamen bir ordinal veri oluşturdular aslında. Şöyle ki, toplam 380 tane satır var, her bir hasta bir satıra kaydedilmiş. Yeni bir sütun açtılar ve bu sütunda her bir hastanın denk geldiği *quintile* neyse onu yazdılar. Nihayetinde yeni bir ordinal, sürekli değişken gibi işlem görebileceği, regresyonda, bir sistem oluşturdular ve bunun üzerinden sistemi lojistik regresyonla yeniden ele aldılar. Mesele bu. Bu durumda lojistik regresyon şöyle yorumlanabilir: 5 kategorimiz var, her bir kategoriden diğerine, yani mesela 1'den 2'ye geçerken odds oranı kaç kat değişiyor gibi yorum yapılabilir. 5 birim var, 5 birim içerisinde her birim artışta, bir *quintile* artışta %20'lik persantil artışında ne ölçüde farklılık var, yorumunu yapabilirler rahatlıkla.

N. Akkoç: Hocam, aslında *propensity* skor yaptıktan sonra birkaç tane yöntem kullanılmış, burada en çok kullandıkları stratifiye. Aslında stratifiye ettikten sonra her grup içerisinde hastaları karşılaştırıyorlar, ortalama tedavi etkisi ondan sonra hesaplanıyor. Çünkü bu birinci *quintildeki* hastalar birbirine benzer *propensity* skorlarına sahipler, dolayısıyla birbirlerine benzer prognostik faktörlere sahipler, aslında standardize prevalans hesaplamak gibi.

H. Yazıcı: Biraz daha irdeleyelim. Bu tabloya gelene kadar tamam. Bu tabloya geldikten sonra, 4-5'i birleştirmek mümkün ki bunlar ağır skorlar. 4-5 tedavi alanlar. Bunu yapmıyorlar mı?

N. Akkoç: Hayır, ayrı ayrı değil, yapıyorsanız -ki zaten bu çalışmanın kısıtlamalarından biri bu, o yüzden aldım- örneklem büyüklüğü bu bakımdan önemli. Bakın burada 71 tane hasta var, burada 5 tane hasta var; yani 71 hasta ile 5 hastayı karşılaştırmak sonucun çok güvenilir olmayacağını gösteriyor. Aynı şekilde birinci *quintilde* de %5 hastayla %40 hastayı karşılaştırmışlar. İdeal bir *propensity* analizinde, bu sayıları büyük olduğu zaman her stratum için hesaplıyorsunuz, daha doğrusu program hesaplıyor. Mesela bir Mantel-Haenzsel yapabiliyorlar, bir kombine odds oranı veriyor.

H. Yazıcı: Bunu sormuyorum. En sona gelelim, yani en ağır hastaları alalım; tedavi alanların en ağırılarını. En ağırılarının arasında gayet güzel ikiye iki gidebilecek. 5 senenin sonunda ki bilgiyi oradan bulabiliyorsunuz. O kadar kesin değil ama kesine biraz daha yakın. Niye? Birleştirmiyorsunuz, teorik olarak yapabilirsiniz, hatanızı biraz daha büyütürsünüz, ama yönünü görürsünüz, dolayısıyla n'leri 30'a çıkarttığınızı zaman, *confidence intervalleri* daha düşük verebilirsiniz.

N. Akkoç: 5'ten daha fazla böldüğünüz zaman hata büyüyor.

B. Çakır: Biraz teknik olacak ama yeterince teknik konuştuk onun için açmakta zarar yok diye düşünüyorum. *Quintilleri* indikatör mü alıyorlar, *dummy* gibi mi alıyorlar? Benim anladığım indikatör gibi alıyorlar çünkü Mantel-Haenzsel ile birleştiriyorlar. Ama bu *propensity* analizinin mantığına ters değil mi? Bizim beklentimiz gerçekten tedavinin etkisinin aslında ayrı bir değişken olabileceği ve endikasyon nedeniyle tedavi değişebileceği için sonucun değişebilmesi ise ve eğer biz farklı şiddette endikasyonların altında tedavi sonuç etkisinin farklı olabileceğini düşünüyorsak mantiken bunları *dummy* kabul etmek lazım. Yani bunları Mantel-Haenzsel ile bir araya getirmek. Aslında grup içerisinde yani birinci *quantildeki* tedavi ile sonuç etkisinin, ikinci *quantildeki* tedavi ile sonuç etkisine benzer olduğunu varsayıyoruz. Bu bizim baştaki yola çıkışımıza ters düşmüyor mu? Ama öte yandan da, size sormadan düşündüm, makaledeki tablo 6 gibi bir yerde, 6 aydan önceki dönem içerisinde tek bir değer vermek çok zor; Mantel-Haenzsel'i verdikleri için tek veriyorlar. Yani kullanımla ilgili böyle bir teori tartışması var mı, merak ediyorum.

N. Akkoç: Tek bir yöntem yok aslında. Mantel-Haenzsel çok da yapılmıyor. Değerlendirmede stratifikasyon bir yöntem, diğer yöntem de mesela bir *propensity* skoru var, tekrar lojistik regresyon yapıyorsunuz ve *propensity* skorunu da içine katıp o şekilde hesaplıyorsunuz.

H. Yazıcı: Bir soru daha. Peki, onu yapmadık, sadece bunları olduğu gibi bıraktık. Bu ol-

duğu gibi bıraktığımızda, o 4 *degree of freedom* ki-kare yapabilirsiniz buna.

N. Akkoç: Yine de beşe bölemiyorsunuz aslında. Çünkü bunu dağıttıktan sonra bir şey bakmanız gerekiyor o da şu, buradaki her stratumda kovariyatların dağılımına bakmanız gerekiyor, kovariyatların dağılımının eşit olması lazım her şeyde. Eğer kovariyatların dağılımı eşit değilse, o zaman sonuçlar güvenilir değil. Zaten buradaki sayıların düşük olması o yüzden. Başlangıçta sağlamaya çalıştığınız değişik stratumlar oluşturarak aynı değişken distribyusyonuna sahip gruplar oluşturabilmektir.

H. Yazıcı: Buraya gelene kadar, o kovariyatlar bakımından olabildiğince düzeltmeye kalktınız ve buraya geldiniz zaten.

N. Akkoç: Ama onu kontrol etmemiz lazım burada; doğru düzelttik mi diye.

Ö. Ergönül: Burada Nurullah Beyin söylediği tarzda yaparsak, iki kesimli, aslında, her bir grup kendi içerisinde değerlendirilebilir. Böl-dük yani biz bunu beşe böldüğümüz zaman her bir grubu kendi içerisinde değerlendirdik. Ancak burada tabii en başta sözü edilen başka kovariyatlar var. O kovariyatların, buradaki her bir grubu, örneğin beş kişiyi temsil edebilir olması lazım. Dolayısıyla daha da küçük bölümlere bölüneceği için, bu kovariyatların değeri düşer; mesela orada 5 olması gibi.

N. Akkoç: Bakın *propensity* skoru yaptık buraya gelmek için ama yaptığımız şeyin amacı şu aslında: Biz bir formül oluşturduk, bütün prognostik faktörler için bir planlama yaptık. Bunu yapmaktaki amacımız şuydu: Biz farklı tabakalamalar yapalım ve bu farklı tabakalarda birbirine benzer prognostik faktörlere sahip hasta grupları oluşturalım.

H. Yazıcı: Tamam çok güzel. Onların dağılımı beklendiği şekilde tedavi almayanla alan arasında...

N. Akkoç: Bu yapılmamış aslında, burada bahsetmiyorum.

H. Yazıcı: Makaleden bahsetmiyorum. Bu cetvele geldikten sonra bu cetvelde benim hemen yapmak istediğim ya uçlara yahut da toplama, ya ikiye iki, dört serbestlik dereceli bir tane ki-kare yapmak.

Ö. Ergönül: Yapabilirsiniz.

Katılımcı: O stratifikasyonu yapmak için aslında kovariyatların eşit olması gerekli değil mi?

N. Akkoç: Dağılımın eşit olması lazım

Katılımcı: Olayın klinik anlamını da beraber düşünürsek aslında bu tablo bize aynı zamanda sınırlarını da veriyor bence. Çünkü sona bakalım, burada tedavi serbest olduğuna göre, aslında çalışmanın genel mantığını düşünürsek, insanlar 76 hastanın 71'inde doğru karar vermişler. Yani hastaların büyük çoğunluğunda *propensity* skor yüksekse, ciddi bir tedavi yapılmış değil mi? Öbür uca bakarsak, eğer *propensity* skor düşükse hekimlerin çoğu zaten 97 hastanın 88'inde doğru bir tedavi seçimi yapmışlardır. O zaman bu yöntem bu uçlarda bizim çok işimize yaramaz. Zaten böyle bir yöntem olmadan da gereği yerine getiriliyor ama ortaya bakarsak, yani *propensity* skoru daha ortada olanlarda bir karar problemi var. Belki işte bu çalışma bize o grup arasındaki farkı göstererek yol gösterebilir.

Ö. Ergönül: Zaten şöyle bir şey, hatta birleştirebilirsek, birinci ve ikinciye birleştirip, 4-5, hatta 5-8 olsun, 3 tane grup da yapılabilir. İsterseniz hani, anlamlı bir sonuç geliyorsa, bunları kollaps yapabilirsiniz aslında dağılıma göre. Ama denemişlerdir mutlaka ve bence şuna gelmişlerdir: Aslına bakarsanız 71-5 bize çok düşük görünüyordu ama tedavi almamış 5 tane denek kalmış olması, bölmelerine rağmen çok da kötü değil bence.

A. Dinç: Bu uçları *outlier* olarak değerlendirdiğimizde, hepsi bir merkezden geliyorsa orada birisi yanlış yapmış demektir. Ona da bakmak lazım.

N. Akkoç: Şimdi bahsedeceğim, öyle bir yöntem var: *Outlier*'ı tamamen diskrimine edebilirsiniz. Zaman kısıtlandı, isterseniz toparlamaya çalışayım. Bu *propensity* skorlarda yaptığınız şey ilk önce tedavi gibi bir olasılığı hesaplıyorsunuz, daha sonra regresyon yapıyorsunuz, bu regresyonda da hastanın tedavi alma şansını etkileyebilecek kovariyatları kullanıyorsunuz. Bu *propensity* skoru yaptıktan sonra da, bunu nasıl kullanacağınızın değişik yöntemleri var; bu makalede kullanılan stratifikasyon yöntemi veya *matching* yapmak gibi.

Matching iki türlü yapılır: Bir tanesi random olarak tamamen tedavi grubunu yazıyorsunuz, ondan sonra her hastanın *propensity* skoruna göre ona en yakın diğer grubu seçiyorsunuz veya tamamen bu stratifikasyonu yaptıktan sonra *outlier* olan kısmını dışarı atıyorsunuz. Tedavi etkisini hesaplıyorsunuz, tekrar regresyon yapıyorsunuz; lojistik regresyon. Regresyon yaparken *propensity* skorunu da tekrar değişken olarak işin içine katıyorsunuz, buna göre düzeltme yapıyorsunuz. *Propensity* skoru tartışmalarında çalışmanın en büyük problemi, bana göre DMARD'lar ve steroid dediğimiz tedavilerin çok heterojen olması. Hastaların bir kısmına metotreksat veriliyor, bir kısmı sulfasalazin, bir kısmı steroid alıyor. Dolayısıyla *propensity* analizi yapsanız da yapmasanız da buradan çıkacak şey kolay değil. Bir ikincisi hem ilaçlar, tedavi stratejileri birbirinden farklı, hem de hasta homojen değil. Bir kısmı romatoid artrit bir kısmı psoriyatik artrit; bu önemli bir kısıtlılık. Şimdi bu kovariyatlardan aslında değişken içerisine katmamalarına rağmen, makaledeki tablo 5'te şunu görüyoruz ki tedavi alan grupta romatoid artrit görülme sıklığı ACR kriterlerine göre bir metodla %60, tree metod ile %78, halbuki tedavi almayan grupta %38'e %55; bu istatistik olarak anlamlı. Dolayısıyla tedavi düşündüğümüz zaman da en önemli şeylerden bir tanesi eğer hasta romatoid artrit kriterleri dolduruyorsa buna tedavi verirsiniz, doldurmuyorsa vermezsiniz. Halbuki *propensity* skor analizinde hastaların zaten %50'si başlangıçta romatoid artrite sahip ama buna rağmen *propensity* skoru hesaplanırken romatoid artriti hesaba katmamışlar, yanlış olabilir veya güvenilirliğini azaltabilir.

Soru: Ama orada zaten *propensity* skoru içindeki *outcome* 12 ay içerisinde DMARD başlanması değil miydi? *Propensity* skorunu belirleyen şey bu zaten.

N. Akkoç: *Propensity* skoru 12 ay içerisinde tedavi alma olasılığını söylüyor.

Soru: Tamam yani sonuçta bu bana hastalığın ağırlığını ifade ediyor. Aslında öyle görüyorum ki ağır dediğiniz hastalar daha çok romatoid artrit ve daha çok DMARD alıyorlar, yani bu çok şaşırtıcı bir şey değil.

N. Akkoç: Bir modelin tam olması için bütün önemli kovariyatlara sahip olmak lazım. Eğer bütün kovariyatları koymazsam eksik olabilir sonuçta; romatoid artritin oraya konması gerekiyor.

Ö. Ergönül: Bir başka şey, burada zaman faktörü söz konusu, mesela romatoid artritin seyri zamana ayrılabilir. Romatoid artrit diğerine göre daha hızlı ve progresif seyrederse o zaman sistem çöker ya da mesela DMARD'la ilgili yeni ilaç çıkmıştır, mesela anti-TNF'ler son 5 yılın ilaçları.

N. Akkoç: Burada yapılması gerekenlerden bir başkası *propensity* skorlarında *overlap*'a bakılması gerekliliği. Bu grafiği yapmamışlar. Verdikleri *inter-quantillerden* ben çizdiğim zaman, görüyorsunuz *overlap* son derece zayıf; şu %25'i, şu da %75'i gösteriyorsa *overlap* son derece zayıf. Halbuki *propensity* skorunun metodolojisinin uygun olması için, güvenilir olması için gruplar arasında iyi bir *overlap* gerekiyor. Yoksa işte şurada gördüğünüz gibi bir kısım hasta hiç tedavi almayabilir. Bir başka sorun şu: *Propensity* skoru hesaplarken ilk 12 aya göre hesaplıyorlar, fakat daha sonra nedense bunu 6-12 ay, 6 ay diye bölüyorlar. Muhtemelen baştan yapmayı düşündükleri analizden bir sonuç çıkmadığı için daha sonra böyle bir şey yapıyorlar, çünkü *propensity* analiz, ilk 12 ay için hesaplanıyor.

H. Yazıcı: Bunu çok güzel vurguladınız. O zaman *propensity* skoruyla analiz etmek için ideal durum, endikasyonları çok kesin olmayan tedavilerin geriye dönüp irdelenmesi.

N. Akkoç: Bir başka kısıtlılık da şu: Burada *propensity* skoruna aldığınız bütün değişkenler başlangıçtaki değişkenler. Halbuki tedavi kararını daha sonra veriyorsunuz ama bu bazale dayanmak zorundasınız. Bir de açıkçası şu çok komik geldi bana: 6 ayda erken tedavi verirsiniz, geç tedaviye göre sakatlık daha iyi ama eğer erken tedavi verirsiniz sonuçlarda hiç tedavi vermediğin kadar iyidir diyor. Yani bu şu anlama geliyor: Hiç tedavi almayanların verileri ile erken tedavi alanların verileri iyidir. Bu bizim klinik pratiğimizde nasıl bir etkisi olabilir? Olamaz çünkü biz hastayı gördüğümüz zaman hangi hastanın tedavi alıp almayacağını

bilemiyoruz. Bu çalışmadan ilk okuyuşta, erken tedavi verirsiniz, bunlarda gidiş daha iyi gibi bir sonuç çıkıyor ama gördüğünüz gibi tedavi verilmeyen bir grup hastanın da kliniği burada çok iyi gidiyor. Dolayısıyla bence okuyucu olarak, açıkçası çok fazla kısıtlılıklı olan bir çalışma; herhalde *propensity* analizini ilk defa kullandıkları için.

Ö. Ergönül: Şöyle toparlayalım: Dün Banu'nun söz ettiği, iç ve dış geçerlilik anlamında, bir de Nurullah Beyin son söylediği gibi, bir makale olarak, istatistik anlamda titiz bir çalışma sonuçta, ama nasıl bir sonuç çıkaracağız?

Soru: Aslında çok bilinen bir hipotezi irdelenen bir çalışma ve senesi 2001. Bunların toplanması 90'lar. Geriye dönüp baktığımız zaman bize çok normal geliyor, ama o dönem özellikle Avrupa'daki romatolojiye baktığımız zaman, erken tedavi konseptinin yeni yerleşmeye başladığı zamanlar; 90'ların sonu, 2000'lerin başı. Dolayısıyla bu yazının önemini biraz o açıdan değerlendirmek lazım. O dönemde Hollanda'da çıkan yazılar da böyle, erken tedavi ile *questi-onnaire*, daha sonra radyolojik erozyonla ilgili yazılar da var. Bu, onların başlangıcı olan bir yazı. Dış geçerliliği bence böyle yapılmış sayılabilecek bir yazı aslında. Yöntem olarak değil ama konsept olarak dış geçerliliği yapılmış sayılır. Şu anda biz bu erken tedavi ile sonuç parametrelerinin daha iyi uyduğunu düşünüyoruz

H. Yazıcı: Bu bana biraz RA'da BEST çalışmasının sonuçlarını çağrıştırdı. BEST'in sonuçları, ne yaparsa yapsın ikinci yılın sonunda aynı yere geliyordu, yani bir tedavi vermek fena değil gibi de... Tedavi vermek iyi çünkü onun bir kısmını, Nurullah'ın altını çizdiği yeri oldukça ciddi romatoid artrit tedavi edenler kullandılar, yani bir işe yarıyor mu diye. Çok teşekkür ediyorum, en içten teşekkürler. Fevkalade yararlandık.

KAYNAK

1. Wiles NJ, Lunt M, Barrett EM, et al. Reduced disability at five years with early treatment of inflammatory polyarthritis: results from a large observational cohort, using propensity models to adjust for disease severity. *Arthritis Rheum* 2001; 44: 1033-42.

A	
ACR kriterleri	37
Açıklayıcı çalışmalar	8, 9
<i>Allocation</i> bkz. Gruplara atama	
Andersen-Gill düzeltmesi	54
“Askı” kurulu, eş değerlendirme	48
B	
Bağımlı / Bağımsız değişken	16
Bayes kuramı	38
Behçet çalışması	38
— kriterleri	38-40
BEST çalışması	79
Bilgilendirilmiş olur	43-46
— araştırmalarda	46
<i>Blinding</i> bkz. Körlleme	
Blok randomizasyon	3, 9, 10, 58-60, 65, 67
<i>British Isles Lupus Assessment Group Index</i> bkz. SLE aktivite indeksleri, BILAG	
C	
Chapel-Hill kriterleri	37
<i>Confounding by indication</i>	24, 70, 74, 75
<i>Confounding factor</i> bkz. Karıştırıcı faktör	
<i>CONsolidated Standards of Reporting Trials (CONSORT)</i>	7, 8, 12, 54, 55, 65, 66
— akış şeması	66
— kontrol listesi	65
<i>Counterfactual</i>	24

Ç

Çok değişkenli analiz	16
-----------------------	----

D

Değişkenlerin etkileşimi	22
Duyarlılık	38-41

E

Eğilim analizi bkz. <i>Propensity</i> analizi	
Eş değerlendirme	47-49, 52
Etik komite onayı	45
EULAR kriterleri	38
<i>European Consensus Lupus Activity Measurement Index</i> bkz. SLE aktivite indeksleri, ECLAM	
<i>Explanatory trials</i> bkz. Açıklayıcı çalışmalar	
<i>External validity</i> bkz. Geçerlilik, dış	

F

Finlandiya Kanseri Veritabanı	29, 30, 33
Fisher	2
— testi	54, 64
Framingham Kalp Çalışması	72

G

Geçerlilik	6, 21, 38, 41
— dış	6, 65, 79
— iç	6, 8, 12, 79
Genellenebilirlik	6, 8, 12, 53, 56, 65
Gruplara atama	4, 9, 63, 67
Grupların karşılaştırılabilirliği	11

H

Hakem	47
Helsinki Deklarasyonu	43-45
Hosmer-Lemeshow testi	21, 24, 75

I

<i>Index Medicus</i>	47
<i>Informed consent</i> bkz. Bilgilendirilmiş olur	
<i>Intention-to-treat</i>	6, 8, 11, 55, 58, 62, 72, 73
<i>Interaction</i> bkz. Değişkenlerin etkileşimi	
<i>Internal validity</i> bkz. Geçerlilik, iç	

K

Kanser kayıt sistemleri	27
Kaplan-Meier analizi	28
Karıştırıcı faktör / değişken	1, 12, 22-24
Kesişim noktası	18
Kırım-Kongo kanamalı ateşi	26, 74

Kısıtlanmış / Sınırlı randomizasyon	2, 3, 8, 9, 59, 60, 67
Kohort analizi	28, 29
Körleme	4, 11, 13, 57, 58, 63, 65, 67
Kuvvet hesabı	bkz. Power hesabı

L

<i>Left truncation</i>	28
<i>Likelihood ratio</i> bkz. Olabilirlik oranı	
Lineer regresyon	16, 17
Log-Rank testi	54, 64
Lojistik regresyon	15-22
— transformasyon	18
<i>Lupus Activity Index</i> bkz. SLE aktivite indeksleri, LAI	

M

<i>Management trials</i> bkz. Yönetim çalışmaları	
Mantel-Haenzsel testi	77
Modelleme, periyot analizinde	33

N

<i>National Library of Medicine</i>	47
<i>Number needed to treat</i>	12
Nürnberg Kodu	43, 44

O

Odds oranı	18-22, 38
Olabilirlik oranı	21
OMERACT	51
<i>Open blind / Label</i>	61
<i>Outcome Measures in Rheumatology Clinical Trials</i> bkz. OMERACT	
Outlier	78

Ö

Özgüllük	38-41
----------	-------

P

P değerleri	4, 19-22, 59, 60, 62, 63
<i>Peer review</i> bkz. ROC Eş değerlendirme	
Periyot analizi	27-37
— — modellenmiş	33-36
— — standart	35
— yaşam tabloları	29
Plasebo	44
Power hesabı	61
<i>Pragmatical / Practical trials</i> bkz. Pragmatik çalışmalar	
Pragmatik / Pratik çalışmalar	8, 9
<i>Propensity</i> analizi	23-26, 69
— skoru	23

Q

<i>Quintile</i>	76
-----------------	----

R

<i>R square</i>	21, 24
<i>Random sampling</i> bkz. Rastgele örnekleme	
Randomizasyon	1-13
Randomize kontrollü çalışmalar (RKÇ)	51
Rastgele örnekleme	5
<i>Receiver operating curve</i> bkz. ROC	
<i>Restricted sampling</i> bkz. Kısıtlanmış randomizasyon	
ROC eğrileri	39

S

Saarland Kanser Kayıtları	35
Sağkalım analizi	16, 27
— eğrileri	30-33
Selection bias	74
Sınıflandırma ve tanı kriterleri	37-42
SLE aktivite indeksleri	
— — — BILAG	51
— — — ECLAM	51
— — — LAI	51
— — — SLAM	51
— — — SLEDAI	51
— — — — skorları	54, 55
<i>Stratified randomization</i> bkz. Tabakalı randomizasyon	
<i>Systemic Lupus Activity Measure</i> bkz. SLE aktivite indeksleri, SLAM	
<i>Systemic Lupus Erythematosus Activity Index</i> bkz. SLE aktivite indeksleri, SLEDAI	

T

Tabakalı / Stratifiye randomizasyon	3, 7, 9-12, 20, 23, 58, 64, 65, 76
Tanı kriterleri bkz. Sınıflandırma ve tanı kriterleri	
Tedaviye niyet (analizi) bkz. <i>intention-to-treat</i>	
Tek değişkenli analiz	16

W

Wikipedia	49
<i>Work-up bias</i>	74

Y

Yönetim çalışmaları	9
---------------------	---