

ROMATOLOJİDE KLİNİK ARAŞTIRMALAR VE NESNEL TİP VI, VII, VIII

EDİTÖR
HASAN YAZICI

S. AYDIN
B. ÇAKIR
A. DİNÇ
Ö. ERGÖNÜL
B. FELDMAN
V. HAMURYUDAN
G. HATEMİ
M. İNANÇ
E. LESAFFRE
S. SUISSA
N. SÜT
İ. ŞİMŞEK
K. TAŞÇILAR
B. ÜSTÜN
N. VASTESAEGER
H. YAZICI
Y. YAZICI

**ROMATOLOJİDE
KLİNİK ARAŞTIRMALAR
VE NESNEL TIP VI, VII, VIII**

MSD'nin koşulsuz eğitim sponsorluğu ile yapılmıştır.

Romatolojide Klinik Arařtırmalar ve Nesnel Tıp VI, VII, VIII

Editör

Prof. Dr. Hasan Yazıcı
İ. Ü. Cerrahpařa Tıp Fakóltesi
İç Hastalıkları Anabilim Dalı
Romatoloji Bilim Dalı (emekli)
İstanbul

Birinci baskı © Sigma Publishing, 2015

© Türkiye Romatoloji Derneęi

Yayın hakları Türkiye Romatoloji Derneęi'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, çoęaltılamaz ve yayınlanamaz. Kaynak olarak gösterilmek koşuluyla alıntı yapılabilir.

Yazıcı (Ed.)

Romatolojide Klinik Arařtırmalar ve Nesnel Tıp VI, VII, VIII
ISBN 978-605-5370-52-7
ISBN 978-605-5370-53-4 (Elektronik kitap)

Önemli Uyarı:

Bu kitabın içerięi, temel ve klinik arařtırmalarla sürekli geliřmekte olan tıp biliminin güncel bilgi düzeyi çerçevesinde oluşturulmuřtur. Editör ve yazarlar, kitabın hazırlandığı zaman diliminde, özellikle tanı ve tedavi uygulamaları ile ilgili olanlar başta olmak üzere tüm verilerin güncel olması ve hatasız basılması için özen göstermiřlerdir. Konuyla ilgili bilgilendirme amacıyla hazırlanmış olan bu kitapta yer alan bilgi ve öneriler okuyucunun saęlık 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 baęlayıcı deęildir. Bu 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 danıřmaları önerilir. Ayrıca kuřku uyandıracak pratik durumların yayınevine bildirilmesi rica olunur.

sigma publishing

Sigma Publishing Dan. ve Org. Dıř Tic. Ltd. řti
Viřne 2 Bölgesi, Çayır Güzeli Sok. No: 14 Zekariyaköy Sarıyer İstanbul
Tel: 00 90 212 352 1661 Faks: 00 90 212 352 1657
info@sigmapublishing.com www.sigmapublishing.com
Sertifika No: 12165

Baskı: Yek Matbaacılık San. ve Tic. Ltd. řti. 100. Yıl Mahallesi, MAS-SIT Matbaacılar Sitesi, 4. Cadde, No:122 Baęcılar İstanbul
Tel: 0212 430 50 00 | www.yekmatbaa.com
Sertifika No: 32287

Bu kitapta yayımlanan yazı ve resimlerin sorumluluęu yazarlarına aittir. Sigma Publishing herhangi bir sorumluluk kabul etmemektedir.

İÇİNDEKİLER

*Romatolojide
Klinik Araştırmalar
ve Nesnel Tıp VI*

Önsöz	V
Saha çalışmalarından beklenen ve olası sorunlar	1
<i>Banu Çakır</i>	
Romatolojide uzunlamasına çalışmaların analizi: Doğru yapılıyor mu?	13
<i>Emmanuel Lesaffre</i>	
Erken romatoid artritte uzun dönem prognozu öngörebilen ögeler var mı?	25
<i>Nathan Vastesaeger</i>	
Klinik çalışmalar ve klinik pratikte anti-TNF tedavisi için AS hastalarının seçimi	36
<i>Nathan Vastesaeger</i>	
Metaanaliz	43
<i>Önder Ergönül</i>	
Maliyet analizleri	56
<i>Necdet Süt</i>	
“Letter to Editor, Rheumatology”de son durum	65
<i>Hasan Yazıcı</i>	
Raed hasta kayıt veritabanı	69
<i>Vedat Hamuryudan</i>	
Biyolojik kayıt kütükleri	73
<i>Ayhan Dinç</i>	
Dergi Kulübü	81
<i>Koray Taşçılar</i>	
Dergi Kulübü	85
<i>Banu Çakır</i>	

İÇİNDEKİLER

*Romatolojide
Klinik Araştırmalar
ve Nesnel Tıp VII*

Randomize kontrollü klinik çalışmalar: Okuyucu, dikkat et!	91
<i>Banu Çakır</i>	
Romatoid artritte yapılan randomize kontrollü çalışmalar: Bize söyledikleri ve söylemedikleri	101
<i>Yusuf Yazıcı</i>	
Tedavi için gerekli sayı ve zarar çıkması için gerekli sayı ile ilgili sorunlar	105
<i>Samy Suissa</i>	
Sonsuz zaman yanlılığı	111
<i>Samy Suissa</i>	
Etki tahminleri ve şansın rolü	116
<i>Önder Ergönül</i>	
Olgularla tartışma	126
<i>Gülen Hatemi</i>	

İÇİNDEKİLER

Romatolojide
Klinik Araştırmalar
ve Nesnel Tıp VIII

Sistemik sklerozlu hastalarda “sınırdan” ortalama pulmoner arter basıncı	135
<i>Murat İnanç</i>	
ACR/EULAR romatoid artrit sınıflama kriterlerinin çok erken inflamatuvar artritli olan hastalara uygulanması çalışmasına yönelik eleştirel bir bakış	142
<i>Gülen Hatemi</i>	
Sinoviti saptamada kullandığımız farklı yöntemler ilerideki radyografik hasarı saptamada birbirine benzer sonuçlar verir mi?	147
<i>Sibel Zehra Aydın</i>	
Kanıtı dayalı tıp için ne kanıt var?	154
<i>Hasan Yazıcı</i>	
Dünya Sağlık Örgütü özürlülük değerlendirme programı 2.0	163
<i>T. Bedirhan Üstün</i>	
Klinik çalışmalarda Bayes metotları	174
<i>Emmanuel Lesaffre</i>	
Nadir görülen hastalıklarda klinik çalışma tasarımları	190
<i>Brian Feldman</i>	
Randomize kontrollü çalışmalarda hasta özellikleri ve merkezler arası farklılık	203
<i>Sebahattin Yurdakul</i>	
Romatoloji literatüründe “olgu sunumu” olgusu	205
<i>İsmail Şimşek</i>	
Romatoid artrit klinik çalışma tasarımları	209
<i>Yusuf Yazıcı</i>	
Dizin	213

ÖNSÖZ

Hasan Yazıcı

2006'dan bu yana düzenlediğimiz Romatolojide Klinik Araştırmalar ve Nesnel Tıp toplantıları, sevinerek söylüyorum, 10. yılına ulaştı. Son üç toplantının kitapçıkları, bir dizi nedenle, gerçekleşmemiştir. Elinizdeki kitapta bunları bir arada sunmaktan mutluyum.

Editör olarak her konuşmayı en az birkaç kez okudum ve ufak tefek düzeltmeler yaptım. Yine mutluyum ki, özellikle sunum sonundaki interaktif tartışmalar yıllar içinde daha da heyecanlı ve öğretici olmaya devam ediyor.

Sunum yapan tüm yerli ve yabancı meslektaşlara, editörlük görevimde bana yardım eden Dr. İlker Kayı'ya, kitabı yayına hazırlayan Q Pharmaworld'den Dr. Rezan Cezan ve çalışma arkadaşlarına, yıllar içinde toplantıların gerçekleşmesine ve kitabın basılmasına koşulsuz eğitim desteği veren, başta Didem Gümüş ve arkadaşları olmak üzere, MSD ilaç şirketine, Sigma Publishing'e, içten teşekkür ederim.

Prof. Dr. Hasan Yazıcı
Ekim 2015

ROMATOLOJİDE KLİNİK ARAŞTIRMALAR VE NESNEL TİP VI

EDİTÖR
HASAN YAZICI

B. ÇAKIR
A. DİNÇ
Ö. ERGÖNÜL
V. HAMURYUDAN
E. LESAFFRE
N. SÜT
K. TAŞÇILAR
N. VASTESAEGER
H. YAZICI

SAHA ÇALIŞMALARINDAN BEKLENENLER VE OLASI SORUNLAR

Banu Çakır

*Halk Sağlığı Anabilim Dalı
Hacettepe Üniversitesi Tıp Fakültesi*

Beni davet ettiğiniz için teşekkür ederim. Türkiye’de epidemiyolojiye olan ilginin giderek artmasına seviniyorum. Aslında yöntem bilmek ne yaptığımızı anlamak için olduğu kadar ne yaptığımızı anlatmak ve başkalarının ne yaptığını anlamak için de çok önemli. Bu anlamda hocamla aynı düşünüyoruz, diye sevindim. Mutluluk verici, tekrar birlikte olduğumuz için de mutluyum.

Epidemiyoloji bir yöntem bilim, sağlık araştırmalarının yapılması, yorumlanması ve kullanılmasında yol gösteren bir yöntem bilim. Ancak oldukça yaygın kullanılan bir sınıflamada ki bu klinisyenler tarafından da sık olarak yapılabiliyor, sağlık araştırmaları epidemiyolojik çalışmalar ve klinik çalışmalar diye ikiye ayrılır oldu. Belki epidemiyolojik çalışmaların ya da epidemiyolojinin sadece sık olarak sahada kullanılan bir disiplin olmasından dolayı. Ya da klinikte yapılan çalışmalar ile klinik dışında yapılan çalışmaları ayırt etmek için bir kolaya kaçma. Oysa her türlü sağlık araştırmasında, ister sahada anket kullanılarak, ister ilaç ile yapılan müdahale çalışmalarında kullanılan yöntemler benzer ve tümü epidemiyoloji disiplini altında toplanıyor. Onun için kendi başlığımı, hocamla da konuşurken surveyler (saha araştırmaları) diye belirlemek istedim. Bugün de o isimle devam etmenin uygun olacağını düşünüyorum.

Survey kelimesi Türkçe’de de sık olarak kullanılıyor. Kelime karşılığına baktığınızda sözlükte “bakmak, incelemek, araştırmak, yoklamak, ölçmek, muayene etmek” olarak tanımlanıyor. Survey, sahada yapılan, toplumu temsil eden yeter sayıda kişiye gidilerek, belirli bir hastalığın yaygınlığını (prevalans), ilişkili risk faktörlerini irdelemeye yönelik kesitsel araştırmalar için kullanılıyor. Öte yandan, dünya üzerinde genel olarak çok geniş yelpazeyi tanımlamak için kullanıldığı gibi Türkiye’de de aynı şekilde kullanılabiliyor. Günümüzde, sahaya çıkılarak yapılan, anket kullanılarak veri toplanan tüm araştırmalara, toplumu temsil etsin etmesin survey denmesi de oldukça yaygın. Bu da surveylerin, bundan sonra saha çalışması diyeceğim, güçlü yanlarını ve potansiyelini olduğundan az algılanmasına neden oluyor. Oysaki saha araştırmalarının ciddi bir potansiyeli var, avantajları ve klinikte yapılan çalışmalara be-

lirli üstünlükleri. Ancak, her çalışmada olduğu gibi uygulamayı bilimsel yöntemlere uygun yapmazsanız çalışmanın zenginliğini elinden alıyorsunuz. Dolayısıyla bu tür araştırmalarda sık yapılan ve özellikle de ülkemizde sık yapılan hatalar üzerinden de konuşarak devam etmek istiyorum ki biz de aynı hatalara düşmeyelim. Survey kelimesini Türkçe’de saha araştırması, sahada yapılan çalışma yerine kullanıyoruz. İngilizce dilinde belki de epidemiyoloji disiplininin de daha iyi oturması nedeniyle farklı alt gruplar var.

Epidemiyolojinin çok büyük hocalarından biri olan John Last’ın belirli aralıklarla periyodik olarak yayımlanan, hepimizin ortak kullandığı sözlüğü vardır. O sözlükten survey tanımını getirdim. Aslında bu bağlamda bakarsanız çok genel bir şey gibi görünüyor. “Sistemati olarak bir veri toplanmasına dayalı inceleme sistemi”, sadece burada deneysel bir komponent yok. Toplumda yapıldığı zaman, birazdan detaylı olarak konuşacağız, yüz yüze anket yöntemleriyle uygulanabiliyor. Kişilerin kendi doldurduğu anketlerle de uygulanabiliyor. Türkiye’de her ne kadar yaygın değilse de telefonla ya da posta yoluyla yapılabilir. Her birinin avantajı, dezavantajı var. Hiç unutulmaması gereken bir şey, aslında surveyler ya da saha araştırmaları toplumla ilgili bir genelleme yapmak, toplumla ilgili bir şey söylemek için yapılıyor. Burada asıl belirleyici adım, seçtiğiniz örneğin toplumu temsil etmesi. Ne kadar iyi temsil ediyorsa toplumla ilgili o kadar iyi bir şey söyleyeceksiniz. Bunun bir alt grubu sahada yapılan surveyler. Yani toplumda sahaya giderek birebir yapılan çalışmalar, Türkiye’de çok yaygın olan hane halkı çalışmaları gibi. Burada kimler eksik, çeşitli kurumlarda yaşayanlar, belki bakımevleri, hastaneler, hapishaneler gibi. Onları alamıyoruz ama yine aynı yöntemlerle niceliksel, sayısal verilerle bir takım analizler yapmaya çalışıyoruz. Bunların alt grubunda da aslında benim bugün konuşmamı oturtmaya çalıştığım ve sizin için muhtemelen daha önemli olan kısım. Hastalık sıklık surveyleri denen ya da bizim yaygın olarak kesitsel araştırma da dediğimiz prevalans çalışmaları. Bunlar da toplum tabanlı saha araştırmaları. Saha araştırmasında hiç vazgeçilme-yen, temel özellikler var. Bir tanesi toplumun

bir alt grubundan, çoğu zaman, hepsine gitmek olası olmadığı için ya da sistematik hatayı en aza indirmek için bu kişilere ait niceliksel veri toplamak. Bu veri toplamada kullanılan araç da sık olarak anketler. Bu aracın özelliği, klinikte alışık olduğumuzdan farklı olarak, kişiden ya da sistemin içinde var olan veri tabanından soru sorma yoluyla, deney yapma yoluyla değil, sadece hastadan ya da başvurandan da değil, sağlam olan, semptomsuz ya da semptomu olan ancak başvurusu olmayan kişilerden de gözlemsel tabanlı bilgi toplama özelliği. İkinci bir özellik, bir örnek üzerinden çalışma. Örnek diyenler var, örneklem diyenler var, ben de kimi zaman birbirinin yerine kullanıyorum. Zaman içinde Türkçe’de doğrusunu bulacağız.

Bugün neleri tartışmak istiyorum? Neden böyle bir çalışmaya ihtiyaç duyuyoruz? Yararları neler? Sağlık çalışanları olarak, hekimler olarak, klinisyenler olarak bizim nasıl işimize yarar? Saha çalışması yapmak bizim hangi eksikimizi doldurur ve bunun “maliyeti” nedir? Maliyeti özellikle tırnak içine aldım, bu sadece finansal, parasal bir maliyet değil, aynı zamanda insan gücü, zaman, yaptığımız hatalar, hatalardan çıkan yanlış sonuçların maliyeti, hepsi birlikte. İlginizi çekeceğini düşündüğüm için, tartışmaya açmak için örneklem ile ilgili birkaç konu konuşmak istedim.

Veri toplamayla ilgili, daha önceki toplantılarda konuşmuştuk, birçok yan tutma tipleri, burada sıkça uygulanabiliyor, sıkça gündeme geliyor, onları tekrar gündeme getirelim. Veri toplama aracını anketlerde sık olarak nasıl kullanıyoruz? Biz kişiyle konuştuğumuz, bilgiyi doğrudan kişiden aldığımız veri toplama araçlarına anket, veriyi hali hazırda toplanmış bir kaynaktan, belirli bir sistemden, kayıtlardan aldığımız zaman bilgi formu diyoruz, anket ve bilgi formunu bu şekilde ayırt ediyoruz. Onun için hastanede daha çok kullandıklarımız, hastane tabanlı çalışmalarda, romatolojide sık olarak kullandığımız araç bilgi formları, ama sahada anket kullanıyoruz, çünkü veriyi kişiden doğrudan alıyoruz.

Anketlerin tamamlanmasında önemli bir konu, bugün başka katılımcılar da eminim gündeme

getirecekler, ‘cevapsızlık oranı’, çok ciddi bir sorun yani sonuçların değerlendirilmesinde çok ciddi bir sorun. Cevap vermeyenlerin kim olduğu, neden cevap vermediği, tüm seçilenler arasında cevapsızlık yüzdesi, çalışma sonuçlarının yorumlanmasında önemli bir hata kaynağı.

Önemli bir diğer hata kaynağı da analiz aşamasında, sonrasında da analizlerin yorumlanması aşamasında. Bugün saha araştırmalarında karşılaşılan bu tür potansiyel hata kaynaklarından bahsetmek istiyorum.

Öncelikle saha araştırmasının bize getirisi ne? Niye saha araştırması lazım ya da lazım mı? Saha araştırmaları lazım, çünkü kliniğe gelen hastalar üzerinden çalışma yaptığımız zaman çok özel bir grup insanı görüyoruz. Bir süredir Türkiye’de akut koroner sendromla ilgili bir çalışma yapıyorum, maliyet etkinlik çalışması. Şunu öğrendik ki, akut koroner sendromların zaten 1/3’ü hastaneye gelmeden hayatlarını kaybediyorlar. Bu 1/3’lük kısım içindeki kişilerin kim olduğunu, bunların hikâyelerini, risk faktörlerini, hangi aşamada kaybedildiklerini bilmek için mutlak surette eve gitmeniz gerekiyor. Yani bunu klinikten bulmanız mümkün değil. Şöyle yapabilirsiniz, sağlıklı kişilerden oluşan kohortlar dâhilinde bunu yapabilirsiniz ama bunun da maliyet ve diğer zorlukları var.

İkincisi, hep hastalıkla uğraşıyoruz ama zaten hastaları, hasta etmemeyi başarabilirsek biraz iş yükümüz azalacak. O zaman hastalıklar için sağlamları ya da yüksek riskli grubu bulmak bizim için önemli. Yüksek riskli grup ailesel hastalıklarda aile bireyleri olabilir, çevresel hastalıklarda çevrede yaşayanlar olabilir, dolayısıyla bu kişileri bulabilmek lazım. Bunun yapılabilmesi de yine sahaya gitmek ile mümkün. Hastalar bize gelmeden önceki dönemde evlerde yaşıyorlar. Onun için o hastaların yanında sağlam ve sağlıklı bireylere de ulaşmak mümkün.

Surveyler asıl olarak sosyal bilimcilerin kullandığı, bize de oradan gelen bir uygulama tarzı. Biz bunu niçin kullanıyoruz? Sağlıkla ilgili her türlü davranış, inanç, görüş, risk alma davranışı, risk

faktörleri, semptomlar her ne ise bunları sahada çalışmamız kolay. Önemli bir özellik bunu yaparken toplumu temsil etmek istiyor muyuz? Eğer temsiliyetten bahsedeceksek, örneklemimizin olasılık tabanlı bir örneklem yöntemi ile seçilmiş olması lazım. Şu örnek çok yanlış: Ben sokaktan geçen 330.000 kişiye bunu sordum, cevabı bu, dolayısıyla Türkiye’de bu diyemiyoruz. Her ne kadar bunu, tarafsız seçmeye çalışsanız da en azından sokağa çıkabilmek için yürüyebilmeniz lazım, sağlıklı olmanız lazım vs.

Araştırmalarda “Bir okula gittik öğrencilerle konuştuk, bir AVİM’de alışverişe gelenlere sorduk, kahvehanelerdekiyle görüştük, eve gittik tüm aile bireyleri ile konuştuk...”, bu durumlar sadece sahaya gidildiği için saha çalışması kabul edilemez. Belirli kuralları var. İlk kurallarından biri, eğer kesitsel bir prevalans çalışması yapacak ve önceden belirlenmiş bir neden-sonuç ilişkisi çalışacaksa, örnekten elde edilen değerler ile topluma ait bir tahminde bulunacaksa, bunun ön koşulu olasılıklı bir örneklem yöntemi kullanmak.

Saha araştırmalarında standardize ölçüm yöntemleri kullanıyoruz. Bu anket dahi olsa soruyu nasıl sorduğunuz, sorudan herkesin aynı şeyi anlaması önemli. Soruların sırası, sorulma tarzı, kişilere hatırlatma ya da probing uygulanıp uygulanmadığı önemli. Aksi durumda bilgi toplamaya bağlı taraf tutma ekarte edilemez.

Bazen de, özellikle bizim gibi kayıt sistemleri iyi olmayan ülkelerde, hastalık yaygınlığını başka şekilde bulmak mümkün değil, prevalans hesapları için mutlaka sahaya gitmemiz gerekiyor.

Peki, saha araştırmalarının maliyeti ne? Hastalığın yaygınlık durumuna göre gidilmesi gereken kişi sayısı artıyor, farklı adreslere gidiliyor, kişiler ile görüşmek için tekrarlayan ziyaretler yapmak gerekiyor. Bir örnek alacak iseniz (kan, idrar vb.) bunun taşınması vb. de var. Dolayısıyla sonucun eldesi aşamasında ciddi bir para harcıyorsunuz. Buna paralel olarak görev alacak (anketör, saha denetçisi, vb.) insan gücü ve zaman da maliyete katkı yapıyor. Tabii seçeceğiniz örnekteki insan-

ların coğrafik yayılımı, özellikleri, bir haneden tek bir kişinin ya da herkesin seçilebilir olması vb. birçok faktör bu maliyeti doğrudan ya da dolaylı olarak etkiliyor.

Saha araştırmaları çoğu durumda araştırmacının ilgi alanı ve amacına göre özgün konular üzerinde yapılıyor. Şu anda Sağlık Bakanlığının bir hastalık yükü çalışması hazırlığında olduğunu biliyorum. Yine birkaç tane kronik hastalık seçilmiş ama diğer kronik hastalıklar yok. Haneye gittiğiniz zaman tek bir durumu sorgulayabileceğiniz gibi eş zamanlı birden çok durumu da inceleyebilirsiniz. Ancak örnek hacminin belirli bir yaygınlık için hesaplanması nedeni ile, eğer birden çok durumu inceleyecekseniz bu durumların toplumda görülme sıklıklarının benzer olması gerekir. Öte yandan, sadece bir durumu irdeleyecekseniz saha araştırmalarında veri başına maliyeti arttırıyorsunuz. Tabii şu da doğru değil, 5 yılda bir yapılan nüfus ve sağlık araştırması var. Ona romatoloji sorusu da ekleyelim diyemezsiniz. Tamamen alakasız konuları bir ankete yerleştirmek örnek hacminin yeterli olup olmamasından öte soru sayısını arttıracak, cevap verenin motivasyonun azaltacak, verinin toplanma süresini uzatacaktır. Dolayısıyla, saha araştırmalarında eş zamanlı birden çok durum incelenebilirse de, bu sayının gerekçelendirilmesi önemlidir.

Saha araştırmalarında veri toplamının kesitsel olduğu, amaca yönelik bir kerelik cevap verdiği, toplanan verinin zaman içinde değişebileceği unutulmaması gereken bir konu.

Gönül istiyor ki sağlık alanında kapsayıcı, tam, doğru, sürekli bir kayıt sistemimiz olsun, iyi yönetilen süveyans sistemleri olsun, biz de sık olarak survey yapılmasına ihtiyaç duymayalım. Tabii bu tür sistemlerin kurulmasının da çok ciddi maliyetleri var ve sonuçları almanız için uzun soluklu çalışmalara ihtiyaç var. Bu arada kesitsel araştırmalar veri ihtiyacını kısa sürede ve amaca yönelik olarak karşılayabilir ve etken-sonuç ilişkisi konusunda da bilgi sağlayabilirler. Kesitsel araştırmaları daha önce de konuşmuştuk; etken-sonuç ilişkisini durumsal olarak incelediği için temporalite (zamansallık) çalışmaz. Oysa ki

süveyans sistemleri sürekli ve düzenli olursa bu durum da incelenebilecektir.

Bir saha araştırmasında basamaklar neler? Önce bir örnek seçiyoruz. Bu örneği nasıl seçtiğimiz önemli, örneğe dâhil edilmesi gereken minimal kişi sayısı önemli. Örnekte kimlerin cevap verdiği (cevapsızlık oranı) ve nedenleri gibi bilgilerin alınması önemli. Veriyi kimden aldığımız ve hangi yöntemle veri topladığımız önemli. Tabii görüşmelerin özellikleri var ve analizler. Saha araştırmalarında da etik konular önemli. Dikkate alınması gereken etik sorunlar var. Her ne kadar biz klinik çalışmalarda etik olan için aydınlatılmış bilgi formu için çok çaba sarf ediyorsak da bu belki daha da önemli çünkü burada bilgiyi evine gidip, kendi başvurusu olmayan bir durumda alıyorsunuz. Yani, size doğrudan başvuran, yardım isteyen kişilerden hastalığı ile ilgili bilgi almak zor ise, beklemediği bir anda kapısını çalıp, aklında olmayan bir konu hakkında bilgi almak daha zor.

Saha araştırmalarında önemli bazı potansiyel yan tutma nedenleri var. Bunların içinde birkaç tanesini birlikte tartışalım istiyorum. Bir potansiyel hata kaynağı seçime bağlı taraf tutma. Burada örneklemin nasıl seçildiği, cevapsızlık oranı, cevap vermeyi etkileyen faktörler önemli. Diğer bir hata kaynağı sahada veriyi toplayanlar ve uygulamaları ile istatistiksel analizi yapanların birbirinden habersiz olması, iletişimin tam olmadığı durumlarda analizlerde yapılması gereken bazı düzenlemelerin (ağırlıklı analizler, eksik verinin düzenlenmesine ilişkin konular vb.) eksik yapılması veya göz ardı edilmesi. Saha araştırmaları bir ekip çalışması. Bu ekipte mutlaka hekim olacak, klinisyen olacak. Tercihen bir sosyolog belki antropolog olacak, bir epidemiyolog olacak, bir biyoistatistikçi olacak ve bunlar birlikte çalışacaklar. Ama sık olarak öyle olmuyor. Soruları bir grup hazırlıyor, anketteki sorular sahada başka bir grup tarafından (çoğu zamanda ihale ile dışarıdan bir araştırma firması elemanları ile) toplanıyor, veri toplandıktan sonra analizi yapan ekip başka. Yorumlar ilgili klinisyenlerce analiz özellikleri göz önüne alınmadan yapılabiliyor. Belki hekim olarak siz soruları verme aşamasında görüyorsunuz, bir de analizler çıktığı aşamada. Oysa ki bu arada hataya açık çok fazla şey var.

Eğer toplumdan bir örnek seçeceksek bu topluma ait bir örneklem çerçevesi olması lazım. Türkiye’de toplumla ilgili kullanabileceğimiz örneklem çerçevelerimiz neler var? Türkiye’de şu anda belediyelerle işbirliği ile iç işleri bakanlığı ile yapılan adrese dayalı kayıt sistemi var. Toplum tabanlı çalışmalarda kullanabileceğimiz nüfus sayımlarımız var, aile hekimlerinin listeleri güncellendiği oranda ve tamamlandığı oranda bizim için bir örneklem çerçevesi olabilir. Ya da başka ufak listeler de olabilir. O zaman bir çerçeve önemli. Bizim çerçeve dediğimiz şey aslında bizim örneğe alabileceğimiz insanları içinde bulunduran herkes. Bu çerçevenin tam olması çok önemli, bu çerçevenin güncel olması çok önemli. Öyle oluyor ki adrese dayalı kayıt sistemi diye gidiyorsunuz ev yerine inşaat çıkıyor ya da orada hiç ev yok çıkıyor. Ya da orada bir işyeri çıkıyor, oradan bir hane bulmanız mümkün olmuyor. Ya da bir çerçeveniz var, siz Ankara’daki aile hekimliği ve Ankara’daki sistemi değerlendirmeye çalışıyorsunuz, aile hekiminin başka yerden gelen hastası varsa onu aldığınız zaman bu sefer çerçeve dışından bir kişiyi almanız gerekiyor. O zaman bu çerçevenin güncel olması, herkesi kapsaması lazım.

Örneklem çerçevesinin kapsayıcı, tam ve güncel olması koşulu sağlandığında, bu çerçeveden seçeceğimiz örneğin sayısı (büyüklüğü) ve bu sayıda kişinin nasıl seçileceği önemli. Tabii buradan örneğin nasıl seçildiği de önemli ve her çalışmada olduğu gibi saha çalışmalarında da bunun maliyet etkili bir şekilde yapılıyor olması lazım.

Değişik toplantılarda yine konuşmuştuk ama eğer bu insanlar sizin çerçevenizse buradaki her bireyin olasılıklı seçiminden bahsettiğimiz zaman şunu söylüyoruz. Buradaki her bir birey, gördüğünüz her bir bireyin çalışmaya alınması için çalışma başlangıcında bilinen ve sıfır olmayan bir seçim olasılığı var. Bu eşit olmak zorunda değil ama herkesin örneğe girme şansı var. Eğer bu örnekten arkadakiler girmeyecekse zaten biz toplumun bir kısmını yansıtmıyor olacağız. O zaman olasılıklı örneklemde özellikle altını çizmeye çalıştım, eşit olması gerekmiyor.

Örnekten elde edilen tahmin değerlerinin toplumu temsil etmesinin altında yatan mekanizma

farklı kişiler tarafından farklı tanımlanabiliyor. Epidemiyolojinin “tasarıma dayalı” bakış açısına göre olasılıklı bir örneklem seçiyorsanız zaten bu örnek hedef grubu temsil eder, tahmini değerleri hedef grubu tanımlamak için kullanabilirsiniz. “Modele dayalı” yaklaşımda ise, ki son 30-40 yıldır daha yaygın olan bir yaklaşım bu, örnek bir üst topluluktan, büyük sanal evrenden gelir. Bu evren zaten rastgele dağıldığı için değişik popülasyonlarda benzer özellikler olur. O zaman bu örnekten aldığınız verileri bir model ile kontrol ederek sonuca ulaştığınızda, örnekten çıkan sonuçlar toplum için de geçerli olacaktır. Tabii burada seçilen modelin doğru olması, modellemenin zorunluluklarının karşılanmış olması gerekli. Kabul ettiğiniz yaklaşım ne olursa olsun, örnekten elde ettiğiniz nokta değerler toplum için söyleyebileceğimiz “en iyi tahmin” değerleridir.

Örneklem seçme yönteminin ne olduğu analizi yapan tarafından mutlaka bilinmelidir. Örnek seçiminde en basit ve kolay yöntem basit rastgele örneklem yöntemidir. Örneğin, 7 kişiden oluşan bir grupta rastgele 3 kişi seçilecekse, rastgele sayılar tablosundan 1-7 arası 3 sayı rastgele seçilir (örneğin 2, 4,5), bu sayılara karşılık gelen kişiler çalışmaya dâhil edilir.

Önemle vurgulamak istediğim bir şey, özellikle bizim Türkiye’de kullandığımız tüm istatistik paket programları, siz aksini belirtmezseniz, örneğin basit rastgele seçildiğini varsayıyor. Bu önemli bir sorun. Eğer biz küme örnekleme seçiyorsak ki Türkiye’de şu anda yasal düzenleme gereği bütün saha çalışmalarının örneğini Türkiye İstatistik Kurumu’nun seçmesi isteniyor. Türkiye İstatistik Kurumu örneği kendi seçmezse çalışma sonuçlarının Türkiye’yi temsil için kullanılmasını kabul etmiyor. Türkiye İstatistik Kurumu da çalışmalarda hane halkı belirlemesinde küme örnekleme seçiyor hatta çok aşamalı seçiyor, tabakalı küme örnekleme. Bu durumda analizde bu seçim özelliği için kontrol sağlanması gerekiyor.

Bir Katılımcı: Peki herhangi bir hastalığın sıklığına bakmak için, bir sahada diyelim. Bunun için istatistik kurumuna mı gitmem lazım?

Evet. Çünkü toplumu temsil edeceksek, o toplumdaki tüm bireylerin bir çerçevesi olacak, o çerçeve içerisinde kişiler (veya haneler) rastgele seçilecek ki o ülkeyi ya da o bölgeyi temsil etsin. Şimdi adrese dayalı kayıt sistemi Türkiye İstatistik Kurumu'nun elinde var. Siz gidiyorsunuz, diyorsunuz ki ben Silivri bölgesinde şu bölgeyi temsilen bir örnek seçmek istiyorum. Örnek büyüklüğünüzü kendiniz belirleyebilirsiniz ya da onlar size destek oluyorlar, onlar size kümeleri veriyor. Sizin küme başına seçmeniz gereken kişi sayısına göre 100'lük, 200'lük, 400'lük kümeler veriyor. Küme içinde de bir başlangıç noktası veriyor. Buradan itibaren sahada devam ediyorsunuz. Eğer siz bunu yapmazsanız ve Türkiye çalışması diye yayımlarsanız Türkiye İstatistik Kurumu bu çalışmanın Türkiye genellemesini kabul etmiyor.

Aslında etmemesi kötü niyetle değil, insanların herhalde kötü niyetini engellemek için böyle yapıyorlar. Çünkü öbür türlü şöyle bir şey oluyor. “Biz Türkiye’yi temsilen bir çalışma yaptık, on altı ile gittik. On altı ilden 60.000 kişi seçtik ve Türkiye’yi çalıştık” diyorlar. Bunu rapor ediyorlar. Eğer örnek doğru seçilmezse bu raporlar çoğu zaman sorunlu olabiliyor. Çünkü örnek seçimi kapalı bir kutu. Örneği seçiyorsunuz ve ben buradan sonucu aldım, diyorsunuz. Bu kutunun içeriğini doğru yapıyorsanız doğru, yanlış yapıyorsanız yanlış. Türkiye İstatistik Kurumu bunu kötü niyetle yapmıyor muhtemelen ama başka bir şey için bunu söylemek gereksinimi duydum. Örneği küme yöntemiyle seçiyor. Ama Türkiye İstatistik Kurumu seçmese bile biz çoğu zaman küme örnekleme kullanmak zorundayız. Çünkü bizim bir listemiz yok, yani aile hekimliğinde bu liste var, 3000 tane hastanız var, 3000 kişinin ismi var, listeden seçeceksiniz. Bu durumda basit rastgele ya da sistematik rastgele örnekleme yapılabilir. Ancak bu çerçeveler her zaman güncel olmuyor.

Başka bir durum diyelim ki Silivri’de belki bunu harita üzerinden örnek seçeceksiniz. Ancak, bu durumda harita üzerinde her sokakta kaç kişi yaşadığını bilemiyorsanız, hız hesaplaması yapmanız zor. Belki listeleme yapacaksınız, evleri gezip kaç tane ev olduğunu bulup onun üzerinden seçeceksiniz. Bunlar hep küme örnekleme uyuyor.

Küme örnekleminde kabul ettiğiniz durum, kümelerin genel itibarı ile birbirine benzer olduğu ancak her küme içinde heterojen bir yapı olduğu. Öte yandan, bu her zaman doğru değil. Sağlık olayları incelenirken genetik ve çevresel faktörler aynı küme içinde yaşayan insanların diğer kümelerde yaşayan insanlara kıyasla daha “benzer” olmasına neden oluyor. Buna desen (tasarım) etkisi deniliyor. Bu konuya çok dikkat etmeliyiz.

Diyelim Silivri’den 5000 kişiye gideceğiz, Türkiye’de ortalama hane büyüklüğü 4. Öyleyse 1200 eve gidip 4’er kişi topluyorsunuz. Oysa romatizmal hastalıkların genetik bir alt yapısı varsa ya da en azından çevreyle bağlantısı varsa o zaman bu kümeler (hane içindekiler) birbirinden bağımsız insanlar değil. O zaman bu basit, rastgele bir örneklem değil, sizin yaptığınız istatistiksel analizlerde kümeleri belirgin kılacak ve tasarım etkisi için kontrol edecek bir uygulama yapılması lazım. Aksi halde hem prevalans hesaplamalarına ait güven aralıklarında hem de ilişkileri çalışırken elde edilecek p değerlerinde hata yaparsınız.

Saha çalışmalarında örnek seçerken örneklem hataları birkaç nedenle olabiliyor. Bir tanesi çerçevenizde ilgili kişiler yoksa, yani çerçeve kapsamına dâhil edilmemiş ise, o kişilere gidemiyorsunuz, zaten baştan bir hata yaptınız. İkincisi eğer bir uygunluk örneği (convenience sample) kullandıysanız, “Ben gittim, bu okulu tanıyordum, bu mahalleyi tanıyordum, bu insanlarla konuştum,” dediğinizde toplumu temsiliyeti bozan bir hata yapıyorsunuz. Bunları düzeltseniz bile, örneğin büyüklüğü ve seçimi ile ilgili şans eseri bir takım değişiklikler olabileceğini de unutmamak lazım. Bunlar için kontrol etmek lazım.

Türkiye’de sağlıkla ilgili yapılan çalışmalarda, özellikle kentsel alanda, ciddi bir cevapsızlık problemimiz var. Bu cevapsızlığın birkaç nedeni var. Bir tanesi gittiğimiz zaman örneklem çerçevemizdeki yapıların yıkılmış olması nedeniyle ya da örneğin haziran ayında giderseniz tarlaya, bağa bahçeye gittikleri için gittiğiniz yerde kişiye ulaşamama riskiniz var. İkincisi kapıya gittiği-

nizde tanımadığınız birinin kapısını, beklemediği bir anda çalılırsunuz ve bir istekte bulunuyorsunuz; zamanını ve bilgisini istiyorsunuz. Çoğu kapıdan kibarca, hatta zorla uzaklaştırılabiliyorsunuz. Burada anketörün rolü, uyandırdığı güven önemli.

Başka bir durum, gittiğiniz kişiler çalışmaya katılmak isteyebilirler ama katılamayabilirler. Çok hasta olabilirler, bir problem olur vs. Bunların hepsi cevapsızlık oranına sebebiyet verir. Bunun için farklı yöntemler var, bir tanesi çalışmaya başlamadan önce olası cevapsızlık oranına göre yedek seçme. Siz 5000 kişiye gitmeliyseniz ve tamamlanma oranı %50 bekleniyorsa o zaman 10.000 kişiye gideceksiniz. Ama şu da, bizim intörnlerin yaptığı gibi, doğru değil. “Hocam ben zaten 5000’e gidecektim, siz beni 10.000’e gönderdiniz ben 6000 kişiye gittiğimde 5000’i tamamladım, diğer 4000’e gitmiyorum,” diyor. Bu doğru değil. Yola neyle çıktıysanız onunla tamamlayın. Çünkü başlangıçta bu kişileri benzer bir seçilme oranı ile seçtiniz ancak ilk gittikleriniz ile sonra gideceğiniz bölgeler fark gösteriyor ise geride bıraktığınız 4000 kişi minimal örnek hacmini etkilemez ama temsiliyeti bozabilir. Diğer kayıp kişileri aramak yoluna gidebilirsiniz ama bunun maliyeti çok arttırdığını düşünmeniz lazım.

Beni yine rahatsız eden bir şey istatistiksel düzenlemeler. Eğer cevapsızlık oranı varsa, dünya için de geçerli ama özellikle Türkiye’de bir grup ‘benim cevapsızlık oranlarım var, bu kadar ama olsun ben bunu size düzenlerim’ dediğinde bu ciddi sıkıntılar yaratıyor. Bir örnek vermek istiyorum, 2008’de Türkiye’de tütün kullanım araştırması yaptık. Eve gidiyorsunuz, örneğin evden Banu Çakır’ı bulmak üzere gidiyorsunuz, bu kişiyi evde bulamıyorsunuz, yine gidiyorsunuz, bulamıyorsunuz. Üçüncü kez gittiğinizde de bulamıyorsunuz. Diyorsunuz ki ben 3 kere gittim, benim anlaşılmam gereği ben artık Banu Çakır’ı bulamayacağım, bir sonraki kişiye gidiyorum. Bunda hiç kötü bir şey yok, taraf tutma yok ama bu şekilde bulduklarınız aslında Banu Çakır’ı temsil etmiyor. Çünkü o kişinin evde olmamasının sebebi her neyse o sigara içme davranışını da değiştiriyor

olabilir. Biz çok iyi biliyoruz ki kadınlarda çalışma durumu ile sigara içme durumu arasında bir ilişki var. Sigaraya erişim, alım gücü, sosyal statü, kabul edilirliliği vs çalışan kadın ile çalışmayan kadın için farklı. Şimdi biz evlere gidip, 3 kez bulamadığımız her kadın için yedeği tamamlamak için sonraki hanede evde bulup cevap veren (çoğu durumda çalışmayan) kadından bilgi aldığımız zaman sigara içme durumunu evde oturan, çalışmayan kadınlar lehine hatalı buluyoruz. Üstelik bir de kadınlara ulaşma yüzdemiz %100 olmadığında, diyelim %90 olduğunda bir de post survey adjustment yapıldığında kadınların sigara içme yaygınlıkları olduğundan az çıkabiliyor. Bu ciddi bir sıkıntı. Bu konuda uyarıyı yapması gereken istatistik ya da örneği seçen kişi değil, çalışılan durum ile ilgili karıştırıcı faktörleri bilen kişinin de bu durumda hem seçim hem de analiz aşamasında ilgili kişiyi uyarması lazım. Onun için her türlü istatistiksel düzenleme, uyarılama yapıldığı zaman, mutlaka o işin içinde klinisyenin de olması lazım. Fikir üretebilmemiz lazım. Birlikte hareket etmemiz lazım.

Farklı yöntemlerin, postanın, telefonun, yüze anketin farklı görüşme tekniklerinin farklı özellikleri var. Ama bir soruyu sorarken, bu da çok önemli ve destek verilmesi gerektiğini düşünüyorum. Sağlık Bakanlığı olsun, kurumlar olsun, üniversiteler olsun. Bir anket formu hazırlanırken değişik uzmanlarla çalışıyorlar, aslında bunun sebebi, uzmanların sorulacak sorunun, doğru bir ölçüt olduğuna destek olmaları. Yani şimdi bir tartı aletinin niye ölçüp ölçmediğini biliyorsunuz. Soruyu çok doğru bir şekilde sorabilmek, iyi bir ölçüt olması için çok fazla kişiden destek almak lazım. Bunun güvenilir olması, yani her seferinde bu sorunun aynı cevabı vermesi lazım ve bu sorunun neyi ölçmek istiyorsak onu söyleyebilmesi lazım. Bu konuda ciddi çalışmalara ihtiyacımız var. En önemlisi de şu ön test dediğimiz şey. Yani bir anket yaptığımız zaman bunu mutlaka bir yerde test etmemiz lazım. Önce küçük gruplarda, daha sonra büyük gruplarda, soru anlaşıldı mı, sorunun yeri uygun mu? Bazen sadece sorunun yeri uygun olmadığı için ankette insanları kaybedebiliyoruz. Hassas konulardaki sorular anketin baş kısımlarında yer aldığında

insanlar çekiniyorlar, cevap vermek konusunda tereddütte kalıyorlar, çalışmadan kopuyorlar ve sizin ne dediğinizle ilgilenmiyorlar, onlar başka bir yerde kalmış oluyorlar. Klinisyenlerin anket uygulamada zorlukla karşılaşmaları çok doğal. Çünkü hasta, klinisyene kendi isteği ile geliyor, yardım almaya geliyor ve biz ona bir şey soruyoruz. Hasta zaten bize söylemek için geliyor. Yani paylaşmak için geliyor, hatta mümkünse başka şeyler de söylüyor, susturmaya çalışıyoruz. Ama anket için eve gittiğiniz zaman, önünüzde bir duvar var, siz o duvarı yavaş yavaş açarak söylemesini istediğiniz şekilde yaklaşıyorsunuz. Bunun için anketörleriniz çok önemli. Anketörü çok özenle ve dikkatle seçmek lazım... Bir kere bu kişinin istekli olması lazım. Her zaman böyle ortamlarda olmuyor. Kişi kahvedeyse gidip kahvede görüşmeniz gerekiyor. İşyerindeyse, işyerinde görüşmeniz gerekiyor. “Bugün git, yarın gel” diyor. “Gece gel” diyor. Çünkü herkesin özel bir hayatı, planları var. Anketörlerin istekli ve çalışma konusunda motive olması lazım. Bu işi yapabilir olması lazım, gideceği toplumdan soru alabilecek gücü olması lazım, eğitimi olması lazım. Ama onun dışında sizin de onu anket konusunda eğitmeniz lazım. Anketörün “Ben bu çalışmayı niye yapıyorum, bir art niyet var mı yok mu, bilgileri nasıl paylaşacağız?” bilmesi lazım ki o da güven duysun, güven uyandırabilsin. Ama konuyu ve neyin incelendiğini çok da bilmemesi lazım, eğer çok fazla bilir, karşıdaki kişiyi o yönde etkilerse bu sefer alacağı cevaplar da değişebilir. Onun için hep diyoruz, hekimler hiçbir zaman iyi bir anketör olamazlar. Çünkü mutlaka manipülatif ruhları açığa çıkıyor. Oysaki anketör kendine ne iletilmişse o çerçevede yapmak üzere seçilmiş insanlar. Dikkat etmemiz lazım. Mutlaka sahada düzenli denetlemeler yapmanız lazım. Örneğin işyerinde anket yapıyor ise çevrede kimsenin olmaması lazım, etkilenmemesi lazım, vereceği cevap için yargılanmaması lazım. Ve mutlaka sorunun geçerliliğinin kontrol edilmesi lazım. Birçok firmayla çalışıyoruz, anket formu getiriyorlar ve biz sahada çalışma yaparken, anketörün kahvede kendi kentine anket doldurduğuna tanık oluyoruz. Yani 3 tane anket yapıyor, mantık da şöyle, 3 tane anket yaptıysam bunun

diğer 3 tanesi de aynı olacak, yani anketleri replike ediyorsunuz. Ama bu çok yanlış bir şey. İşte küme örnekleminin daha kötüsü, çok fazla benzer sonuçlar çıkıyor, onun için gidip yerinde mutlaka anketörü test etmeniz lazım.

Anketöre ne kadar ücret ödeniyor ve anketöre neler için ücret ödeniyor? Anketörü öncelikle sigortalamanız lazım. Sigortasız çalıştırdığınız zaman hem etik olarak yanlış hem de ciddi sıkıntıları var, sigorta ücreti ödüyorsunuz. İkincisi yaptığı anket başına ya da gün başına bir ücret ödüyorsunuz. Ben sık olarak yaptığı anket başına ödüyorum çünkü gün karışabiliyor. Gittiğiniz bölgedeki kişinin cevapsızlık oranı, tekrar gitmeniz gereken durumlar vb. Ama bunu da yine anlamak için konuşmak lazım. Anketörün kalacağı yer için bir ücret ödüyorsunuz. Eğer siz götürmüyorsanız taşıt için ücret ödüyorsunuz. Değişiyor. Yapılan anketin soru sayısı çok önemli. Yani normalde halktan bir kişinin ankete aralıksız yanıt verebilme süresi yaklaşık 20-25 dakikadır. Ama okumuşluğunuz arttığı sürece daha çok zaman ayırabiliyorsunuz. Ama biz Türkiye’de iki saate kadar olan anketler yapıyoruz. Normal toplumda, yani ilkökul mezunlarıyla hatta hiç okumayan yazmayanlarla anket yapıyoruz. Dolayısıyla bu anketörlerin eğitimi ve bu kişilerle yaklaşımı, kontağı kurması o şeyi uzatması çok zor. Dolayısıyla maliyet açısından anketin sayısı çok önemli. Gideceğiniz yer önemli. Bazen özel anketör lazım, yani anketörleri sık olarak lise mezunları arasından seçiyoruz ama sağlık anketlerinde tıp öğrencileri destek verebilir. Hali hazırda çalışmaya pratisyen hekim olabilir. Dolayısıyla her anketör için çok değişken. Şöyle söyleyim, Ankara için günlük küçük bölge 20 liraya da mal oluyor, günlüğü 250 liraya da anket yapılıyor.

Saha araştırmalarında karşılaşılan olası hata kaynakları ile devam edelim. Şimdi anketiniz geldi, siz olasıdır ki verileri kendiniz girmeyeceksiniz. Birlikle çalışacaksınız. O biriyle yine birlikte çalışmak lazım. Neden? Çünkü bu veri geldikten sonra bir veri tabanına aktarılıyor. Ve buradan analiz yapılıyor. Bu aşamada değişik şeyler oluyor. Anketör veriyi yanlış girmiş olabileceği gibi anketörün yazısını okuyamadığı için kişi veriyi

yanlış da girebiliyor. Ama onun dışında değişik kodlamalar kullanıyoruz, değişik şeyleri birleştiriyoruz, mesela diğer sıklıkların isterseniz açık girip sonra kodlayabileceğiniz gibi, isterseniz baştan kodluyorsunuz. Bunların hepsinin yararı var ama buradaki en önemli, hiç unutmamamız gereken şey yaptığımız her gruplama verinizden bir şey alıyor, yani siz büyük bir zemin alıyorsunuz bunun içinden bir gruplama yapmak zorundasınız. Bu gruplama yapılırken mutlaka işin içinde olmanız lazım. Kodlamalar, verinin girilmesi, bunu bir başkasına bırakmamanız lazım. Sonrasında diyorlar ki biz veriyi edit ettik, temizledik, size temiz veri getiriyoruz. Temiz veri şu, bakıyor boyu 2 m 15 cm, olmaz diyor, atıyor. Bu doğru bir şey değil. Mutlaka her aşamasında birlikte paylaşmanız lazım. Atılmalı mı atılmamalı mı, temizlensin mi? Bazen diyorlar ki, sorular çok saçmaydı, güvenilir değildi. Kontrol soruları koyuyoruz biz anketlerin içerisine. Yani 10-15 soru sonra bir tane daha. Bu soruların ortak kontrolünü yapıp, hangi soruyu kullanacak, bunları kendi seçmemeli. Birlikte olmanız lazım. Ben bunları çok önemsiyorum. İkinci bir konu da yine çok önemli. Mutlaka bizim eksik gözlemlerimiz olacak, bu eksik gözlemler nasıl doldurulacak, bunu konuşmamız lazım. Bu kişilerin hepsini atacak mıyız, boyu olmayanların hepsini atalım mı, ortalama bir boy mu koyalım, bir regresyon modelinden onlara boy mu input edelim, bunları hep birlikte konuşmamız lazım. Belki kendimizin yapmayacağı ama mutlaka okurken dikkat etmemiz gereken bir şey, eğer bir küme örnekleme varsa mutlaka desen etkisi için kontrol edilmiş olup olmadığını okumamız lazım, çünkü sonuçlar doğru olmaz. İkincisi eğer toplumdaki kişilerin, sayısıyla orantısız bir örnek seçmediyse, diyelim toplumun %10'u kadın %90'ı erkek. Yüz kişi seçtiğiniz zaman 10 tane kadın olacak, 10 kadın bütün kadınları yansıtmaz, ben kadınların %50'sini seçeyim, erkeklerin %10'unu seçeyim, böyle bir örnekle yola çıkayım dediğiniz zaman, ki sıkça yapılıyor. O zaman bunu analizde her kadının kendisi gibi bir diğer kişiyi temsil eder, her erkeğin 10 erkeği temsil ettiği tarzda bir düzenleme yapmak lazım. Buna ağırlıklanma diyoruz. Yoksa ne olur, bu sefer kadınlar lehine ortalama

kayacaktır. Bu ikisine dikkat etmek lazım. Yani bir ağırlıklanma var mı, eğer gerekiyorsa desen etkisi kullanıldı mı? Uygun yöntemler kullanıldı mı? Bir de her yerde vurgulamaya çalışıyorum. Türkiye'de neden bilmiyorum, her türlü prevalans araştırmasında bir prevalans değeri veriliyor ve o iş orada bırakılıyor. Mesela bir toplumda sigara içme sıklığı baktık, %38 çıktı. Ertesi sene başka bir grup tekrar çalışmış, %32 çıkmış. Diyorlar ki %38 idi, %32'ye düştü. Demek ki sigara artık daha az içiliyor. Bu doğru mu? Örneklerden elde edilen değerler en iyi tahminlerdir. Ama bunun bir güven aralığı var. Bu güven aralığı daha önce örneğin 32 ile 44 arasındaysa aslında bir yıl sonrasında anlamlı bir azalma olmamış olabilir. Çünkü amaç ne, bir örnek var ve bu örnekten toplumu temsil ediyoruz. O zaman kural, örneklem üzerinden yapılan her çalışmada araştırmacı tahmini değerlerin yanında mutlaka ilgili güven aralığı değerlerini de vermeli ki, zaman ve gruplar arası karşılaştırmalar doğru şekilde yapılabilir. “Yanında” diye tekrar vurgulamak istiyorum çünkü şu da doğru değil, güven aralıklarını raporun en arkasında veriyorlar. Kimse açıp bakmıyor, önemini bilerseniz, yerini bulursanız bakarsınız, bulmazsanız bakamazsınız. Güven aralığı değerinin her zaman okuduğum yerde olması güzel. Tabii çok değişkenli analizlerde karıştırıcı faktörleri kontrol etmek lazım. Konuyla ilgili kişileri bilgilendirmemiz lazım.

Ben şunun da etik bir sorumluluk olduğunu düşünüyorum. Etik komisyonlarda çalışırken de hep aynı şeyi söylüyorum. Bir şeyi doğru yapmak da etik bir yükümlülüktür. “Etik komisyon IRB değildir, dolayısıyla çalışmanın tasarımına, örnek hacmine, yapılacak istatistiksel analizlere bakmayalım,” görüşü çok doğru değil, çünkü tasarım bozursa sonuç zaten yanlıştır. Etik olarak yanlış bir şey yapıyorsunuz, boşa zaman, insan gücü, para harcıyorsunuz. Bunu da bir etik yükümlülük olarak almak lazım.

Son olarak, bir saha çalışmasını değerlendirenken basamaklara tek tek bakmamak lazım, bir bütün olarak bakmak lazım. Bu bütünde neler var, örnekle ilgili özellikler var. Çerçeve doğru mu, sayı doğru mu, yeterli kişi cevap vermiş mi, sorular uy-

gun sorular mıymış, veri doğru mu toplanmış, doğru mu analiz edilmiş, doğru mu yorumlanmış? Saha çalışması bir ekip işi, bunu unutmamamız lazım ve saha çalışmasında mutlaka doğru bir yöntemi kullanmamız lazım. Bilimsel yöntemi, bilimsel kurallarıyla kullanmamız lazım. Bunun için de 1, bunu kendimiz bileceğiz, bilmezsek yapamayız; 2, araştırmacı bunu mutlaka raporuna, yeterince detaylı olarak yazmak zorunda. Bu bir makaleye kısaca, bir raporsa sayfalarca, ama “Ben nasıl bir yöntem kullandım, analizleri neye göre, nasıl yaptım, hangi ağırlıklı kullandım, nasıl hesapladım, bu neden yapıldı, olası hatalarım neler olabilir?” bunlar araştırmacının sorumluluğunda, çünkü bunları sizden başka hiç kimse bilemez. Yayın çıktıktan sonra herkes sadece sonuçlarını değerlendiriyor. Bu gizli kutunun içinde neler yaptığınızı çok dürüstçe paylaşmanız lazım, tabii kendinize de diğerlerine de. Okuyucular da eğer saha çalışmalarını düzgünce okuyup “İçeriği nedir, nasıl yapılmış, sonuçları geçerli mi, güvenilir mi, inanılır mı?”, bunları tam ve doğru olarak irdeledikleri sürece araştırmacılar da sorumluluklarını daha iyi yerine getirecek. Saha çalışmaları giderek daha iyiye gidecek. Biz de sonunda saha çalışmaları kakadır, tüh, demeyeceğiz, bunun çok yararlarını göreceğiz, paylaşabileceğiz.

Çok teşekkür ediyorum.

Bir Katılımcı: Türkiye’de saha araştırması deyince normal vatandaş için akla en çok reyting ve seçim anketleri geliyor. Bu anketleri yapan ekip gerçekten bilimsel çalışıyor mu?

Bir Katılımcı: Buna bir şey katabilir miyim, geçen gün arabanın radyosunda iki anket firmasının sahibi seçim anketleriyle ilgili birbirlerini yiyorlar. Bir tanesi diyor ki benim kararsızlarım şu kadar, ötekisi dedi ki benim kararsızlarım şu kadar. Sonra biri ötekine sordu, “Sen cevap vermeyenlerle kararsızları ayırdın mı,” dedi. Ayırmamış onları, hepsini aynı kutuya atmış. Yani esas sorulması gereken, bir anket yapıyorsunuz, o anket bittikten sonra keşke şu soruyu da sorsaydık dediğiniz oluyor mu?

Banu Çakır: Çok oluyor. ‘Soruyu sorsaydım’dan ziyade, hemen buraya gelmeden önce yaşadığımız

bir şey, “Bu soruyu keşke şu şekilde sorsaydık”, dedik. Bu çok önemli bir bulgu. İşte onu paylaşmak lazım ki bir sonraki araştırmacı onu sorsun ve sorması gereken tarzda sorsun.

Ama Yaşar Bey’in sorusuna gelince bilmiyorum, içinde olmadığınız hiçbir şeyi değerlendirmek doğru değil. Ama hep şunu söylüyorum. Taraflı olmasanız dahi, kötü niyetli olmasanız dahi, bu tür anketler doğru değil. Tütün kullanım yasası ile ilgili çok sayıda araştırma var. Bütün gazetelerde görüyorsunuz. Falanca kahveye gidildi, falanca alışveriş merkezine gidildi, bir ara biliyorsunuz lokantalarda sigara içiliyordu, kapalı yerlerde içilmiyordu. Kavga çıktı. Toplumun zaten %15’i alışveriş merkezine gitmemiş. Gidenlerin de %30’u alışveriş merkezinde yemek yememiş. Dolayısıyla siz sokaktaki adama alışveriş merkezine gitmenizi etkiledi mi dediğiniz zaman bir şey ifade etmiyor. Ya da alışveriş merkezinin kapısındakine sorduğunuz zaman toplumu temsil etmiyor. Asıl sıkıntı kötü niyetten çok, bilimsel olmaması.

Bir Katılımcı: O zaman içinde olduğum bir şey söyleyeyim, bazen hekimlerin bir hastalıkla ilgili davranışlarını belirlemek için anket çalışmalarını telefonla ya da yüz yüze anketle yapıyorlar, biz bunlara katıyen cevap vermiyoruz ama hepsi dışarıda dolduruyor. Yani bunlar katıyen güvenilmez. Bazıları da bunları alıp bilimsel veriymiş gibi kullanıyor. Bunu da görüyoruz maalesef.

Bir Katılımcı: Soru sorarken konuştuğun kimsenin eğitim durumunu, özellikle evde hanımla konuşurken şikâyetini ne kadar açıklayabilirsin. Yani oradaki ifadenin ne kadar dışına çıkabilirsin?

Banu Çakır: Çok güzel bir soru, aslında bu tamamen çalışmanın tipiyle özel bir şey. Biz hekim olarak, gelen hastaya probe ediyoruz, değil mi? “O da var mıydı, bu da var mıydı, bu oldu mu?” vb. Ankette bunu isteyebiliriz, istemeyebiliriz. Ama her ne yapılacaksa biz her anketin arkasında anketin nasıl yapılacağına dair bir de yazılı metin veriyoruz, mesela “Size bir soru sorulursa bu şekilde açıklayın ya da size şu sorulursa şu kadar açıklayın,” diye. Buna rağmen yoksa anketör yanına yazıyor, sonra ekip içinde durumu paylaşıyorsunuz, çözüm

arıyorsunuz. Saha kılavuzunda (manuelde) ne varsa anketör onu o kadar yapmakla sorumlu. Aksi halde şöyle bir sorun oluyor, anketör soruyor “Herhangi bir sağlık probleminiz var mı?” Beni rahatsız etmediği için “yok” diyorum. Anketör devam ediyor “Arada sırada da olsa başın ağrır mı? Bilgisayar kullanınca ağrıyor mu? Kapalı yerde ağrır mı?”, “Benim başım oldum olası ağrır”, “Evet” diyorum. Anketör kaydediyor: “Baş ağrısı yakınması var.” Diğer anketör bu probingi yapmadı ise (ki yapılması istenmemiş zaten), o “Yakınması yok” yazıyor. O zaman iki farklı kişiden farklı sonuç almamız, durumlarının farklı olduğundan değil de anketörlerin farklı uygulamaları nedeni ile. Hem anket standart olmalı hem de uygulama standart olmalı. Anketörden ne istendiği açık olmalı, kılavuzda yer almalı ve anketör de ne dendiye o kadar yapmalı.

Bir Katılımcı: Peki, bu saha çalışmalarında tekrarlanabilirlik üzerinde durulmayan bir konu oluyor. Mesela onlardan %10'luk bir örnek alıp bir daha bakmak mı iyi, yoksa başka sayılarla bir örnek %10 seçip ona mı bakmak iyi? Yani tekrarlanabilirliğe bakmanın yöntemi nedir?

Banu Çakır: Ankette iki şey kullanıyoruz. Biri kendi sorularımız, bir tanesi ölçekler. Eğer bir şeyi ölçeceksek depresyon envanteri ya da yaşam kalitesi gibi, bunun mutlaka geçerliliğinin, güvenilirliğinin yapılmış olması lazım. Bu toplumda yapılmış olması lazım. Mutlak zorunluluk. Öbür türlü kendi sorularımızın tekrar edilebilirliğini şöyle yapıyoruz. Ön testte soruyu soruyorsunuz, herkes sorudan aynı şeyi mi anlıyor? Eğer aynı şeyi anlıyorsa aynı cevabı verdiğini varsaymanız lazım. Ama ölçtüğünüz şey de önemli. Bilgi sorabilirsiniz, değişmez ama tutum değişebilir, görüş değişebilir. Onda da tekrar aynı kişiye gitmenin çok önemi olmuyor, mesela ötenaziye nasıl bakarsınız, ilk sorduğunda farklı bir cevap 10 gün sonra sorduğunda farklı bir cevap. Çünkü arada onu düşünüyorsunuz. Tutum, görüş. Bunlar tekrar edilebilirliği değişik de olabilir. Ama tekrar edilebilirlik, güvenilirlikten kastımız, ön testte bu soruların doğru cevaplandığını bilmemiz lazım.

Bir Katılımcı: Ben bir deneyimi paylaşmak istiyorum. Ağız hastalıkları prevalansı olarak sahada

araştırma yapmış, yayımlamayı başarmıştık. İlk Türkiye verisiydi. Çok ciddi bir sorun var. Başta anketör eğitimini iyi vermezseniz veriyi iyi alamıyorsunuz. Şablonu iyi hazırlamadan yine kötü dönüyor. Bir sorun daha, dijital ortama taşınmada eğer başında olmazsanız çok hata var.

Bir Katılımcı: Herhangi bir anket bulgusunun %95 güven aralığını hesaplıyoruz. Mesela burada bir yöntem var mı, örneğin gözlenen %95 güven aralığı versus düzeltilmiş %95 güven aralığı. Mesela nonresponse ya da missing data'ya göre %95 güven aralığında düzeltme yapmak mümkün mü? Çok değişkenli analizlerde yapılan sensitivite analizi gibi.

Banu Çakır: Aynen bunda da sensitivite analizi yapıyoruz. Verdiğiniz değer robust mı diye bakıyoruz. Duyarlılık analizleri özellikle missing data olduğunda kullanılıyor.

Bir Katılımcı: Ben kısaca bir şey söyleyeceğim, bizim çalışmamızda örnekleme yapmadık. Türkiye İstatistik Kurumu da bize sadece nüfus aralıklarını yapmış, 0-5, 5-10 yaşlarına göre. Bir ondan faydalanabildik, bir de medyan yaşı öğrenebildik. Bizim çalışmamızda katılım çok yüksek çıktı. Çünkü önce haberdar ettik, lokal gazeteyle, televizyonla, kaymakamla, hoparlörle filan. Bir başka önemli şey söyleyeceğim, şimdi genetik çalışma yapılıyor, genetik çalışmadaki örnekleme büyüklüğü bildiğimiz yöntemlerle tayin edilemiyor. Biz burada bu yöntemle kaç romatoid artrit var, kaç kişi sigara içiyor ama genetik çalışmalardaki örnek büyüklüğünü takip etmek için dünyada henüz herhangi bir şey yok. Alel sıklığı çok değişkendir. Bu alel sıklığının hastalıklar üzerindeki etkisini araştırmak için ben kendim çalışma yaptım, hiçbir yerde bulamadım. Bunu nereden bulacağız. Böyle problem yükselmeye başlıyor. Genetik çalışmalardaki örnek büyüklüğünü nasıl seçeceğiz, bu önemli.

Ayhan Dinç: Bu aralar anket çalışmalarında telefonla anketler gündemde, hatta son siyasi anketlerin bir tanesini cep telefonlarıyla yapmışlar. Bunların doğruluğu ne kadardır? Yüz yüze görüşmenin yanında telefonla görüşmenin değerlendirilmesini nasıl yapabiliriz?

Banu Çakır: Telefonla görüşmelerin birkaç tane sıkıntısı var ama çok iyi yanları da var. Bazen insanlar yüz yüze konuşmaktan çekinebilecekleri konuları telefonda rahat konuşabilir. Bir kere yüz yüze değil telefon edince maliyetten kazanıyor-sunuz, eve gitmeniz gerekmiyor, erişilebilirliğiniz daha yüksek, işin bir bu boyutu var. Ama önemli sıkıntıları bir herkesin telefonu var mı, bu telefonların hepsine ulaşılıyor mu? Örneklem yöntemlerinden biri de aslında bu şekilde ‘random digital dialing’ denilen. Alan koduna göre random numara yaratıyorsunuz ama gittiğinizde o numara çıkmayabiliyor. Şimdi cep telefonu seçmelerinin sebebi şöyle bir şey. Bizim saha araştırmalarında görülen, Türkiye’de artık insanlar cep telefonu kullanıyor. Özellikle öğrencilerde ev telefonu yok. İkincisi cep telefonu ile ulaşılabilirlik ev telefonundan daha kolay. Ama telefonla araştırmalarda birkaç şeye dikkat etmek lazım. Bir tanesi cep telefonu ile her yerde ulaşabiliyorsunuz, şu anda bana cep telefonu gelebilir. Eğer cep telefonu ile anket yapacaksanız bir zaman dilimi belirleyip, randevulaşıp tekrar aramakta fayda var. Çünkü hemen kişinin olduğu yerde (otobüste, yolda, işte vb.) anketi yapmak çevresel olarak doğru değil. Yöntem olarak telefonla çalışma yapmak yanlış bir yöntem değil ama dediğim gibi tüm bu bilimsel doğrulara uymak koşuluyla. Bir, o telefon numaralarından rastgele bir seçim mi yapılıyor? Farklı telefon şirketleri var, bu şirketler nasıl dağılıyor, belki hepsinden yapılması lazım. Evreni nasıl belirleyeceksiniz? Yani buradaki çerçeve, mesela Turkcell ile çalışıyorsanız, Turkcell numaraları oluyor. Bunların diğer numaralardan farklı özellikleri varsa tamamen farklı bir şey olacak. Çerçeve kısıtlayıcı bir özellik. İki, farklı çerçeveli insanların size cevabı. Ama birçok çalışmada da aslında yüz yüze görüşmeler, özellikle sosyal durumlarda kişi size yüz yüze vermek istemediği cevabı, sizi görmediği için telefonla verebiliyor. Sosyal beklenti, cevabı verirken ne olduğunuz, yüz ifadesi vs. Yanlış değil, ne kadar doğru yapılıyor onu bilmek zor, onun için her yapılan çalışmada aslında sormak lazım, nedir çerçeveniz bu yanıt olduğunda. Yalnız politik durumlarda farklı tabii. Onların zaten beklentisi de farklı. Onun için de doğru yüzdelere verebiliyorlar aslında. Çünkü gittikleri gruplar bi-

limselden farklı. Zaten onlar yanlış bir görüşle yola çıktığı için biraz daha farklı bizimkilerden.

Hasan Yazıcı: Anketlemenin çok klasik tarihi yanlışlarından bir tanesi galiba 1930 Hoover anketidir. Bir Amerikan başkanı. Sadece telefonu olanlara sorarlar.

Huri Özdoğan: Anketörler arası ve içi tekrarlanabilirlik ile ilgili bir şey var mı?

Banu Çakır: Mutlak bir yükümlülük aslında. Deminki sorular gibi anketörlerinde biz ön testini yapıyoruz. Anketörlerle çalışmaya başladığımızda ilk önce çalışmayı anlatıyoruz, sonra önce anketörler kendileri çalışırlar. Sonra rol modelleriyle sanki gerçekmiş gibi çalışırız. Sonra sahada yaparız, sonra bu anketlerin, anketörün kendi içindeki tutarlılığı hem de anketörler arasında tutarlılık olması zorunlu. Zaten anketörlerin arasında tutarlılık yoksa ve anketör kendi içinde tutarlı değil, aynı sonucu vermiyorsa çalışmaya baştan hiç başlamıyoruz.

Önder Ergönül: Survey yerine Türkçe’de araştırma diye çevirmektense başka bir kelime, mesela soruşturma gibi kelime düşündünüz mü?

Banu Çakır: Saha araştırması olarak kullanıyorum ben

Önder Ergönül: Araştırma çok genel, sorgulama, soruşturma...

Banu Çakır: Önder sen ve ben bu dönemi kıyısından yakaladık. Bizden biraz daha büyükler sorgulamayı sevmiyorlar. Biraz daha küçüklere anlaşılmıyor. Sorgulama ya da sorma aslında her çalışmada yaptığımız bir şey. Yani kohortta da anket uygulayabilirsin, o zaman da sorgulama, soruşturma yapmış olmuyor musun? Surveyler yerine anket çalışması diyenler de var. Ama anket sadece surveylerde yapılmıyor. Surveyler saha çalışması, yani sahada yapılıyor. Bir grup anket çalışması diyor ama hani deminki örnekte de sahada her çalışma yapılabilir. Onu da karıştırmamak lazım. Vaka kontrolü de sahada yapılabilir. Kohortlar sahada yapılıyor. Bu o değil ama bir kelime bulmamız gerekiyor. Şimdilik saha araştırması diyorum ben.

Günaydın ve bu nazik davet için teşekkürler. Bu konferansa ve Bodrum’a ikinci defa geliyorum. Konuşmalar uzasa da konuşmamı yapmak zorunda kalmam diyordum ama sandalyenin çok sert olduğunu son tecrübemden biliyorum.

Şimdi anlatacağım çalışma Hollanda Rotterdam Erasmus MC’de bulunan benim çalıştığım bölüm ve romatoloji departmanından bazı üyelerin (Karolina Sikorska, Maurits de Rotte, Jolanda Luime ve Mieke Hazes) bir grup çalışması. Ben bir istatistikçiyim ve bunun sıkıcı bir konuşma olacağını tahmin edebiliyorum. Bu yüzden, sonucu baştan anlatmak istedim, eğer konuşmam sizi uyutursa en azından bu sonucu eve götürmüş olursunuz.

Konuşmamın başında sonuçları vereyim: 1. Romatoloji alanında (ve diğer tıp alanlarında) birçok uzunlamasına çalışma uygun bir istatistiksel metodoloji ile analiz edilmiyor. 2. Tekrarlanan anlamlılık testleri ve özet istatistik yöntemlerin çoğu uzunlamasına çalışma için uygun değil. 3. Klasik tekrarlanan ölçüm yöntemleri de geçerli değil. 4. Daha uygun ya da ‘modern yöntemler’ ise karma etki modeli ve genelleştirilmiş tahmin yöntemleri (GEE) yaklaşımıdır. Söz konusu yaklaşımlar tüm zaman dilimlerinde sonuçların kayıtlı olup olmadığına bakılmaksızın eldeki tüm olgulara faydalanılmasını sağlayabilirler. Ne yazık ki, bu son iki modern yaklaşımdan romatolojide sık yararlanılmıyor.

Şimdi bahsedeceğim çalışma bir Juvenil İdiopatik Artrit (JIA) çalışması. İki ayrı merkezde, Utrecht Üniversitesi Tıp Merkezi ve Wilhelmina Çocuk Hastanesi’nde ILAR (International League of Associations for Rheumatology / Uluslararası Romatoloji Dernekleri) kriterlerine göre Juvenil İdiopatik Artrit teşhisi konan 183 çocukla 5 dönemde yürütülen uzunlamasına bir çalışma. Çocuklara 1990 ile 2006 arasında metotreksat başlanmış. Tüm hasta ve ebeveynlerinden etik kurul tarafından hasta onamı alınmış. Çalışma 3 ayda bir standart hastalık aktivitesi raporu kullanılarak sıkı kontrol altında tutulmuş ve sonuçlar kaydedilmiş. Tüm JIA hastalara ABCB1 3435C>T genotiplemesi yapılmış ve 2 gruba ayrılmışlar. İlk grup 58 GC homozigot

ROMATOLOJİDE UZUNLAMASINA ÇALIŞMALARIN ANALİZİ: DOĞRU YAPILIYOR MU?

Emmanuel Lesaffre

*Biyoistatistik Bölümü
Leuven Katolik Üniversitesi
Belçika*

hastadan oluşturulmuş (tüm grubun %28'i). İkinci grup ise T taşıyıcı hastalar (%72). Burada 2 sonuca ulaşılmaya çalışılmış. Bunlardan birincisi süregen bir yanıt, yani ESR (Eritrosit Sedimentasyon Hızı). Yüz seksen üç hastanın 106'sında ESR 5 kez tekrarlanmış, yani olguların %58'i tamamlanmış olgu olarak nitelendirilebilir (tamamlanmış vaka = tüm planlı ziyaretlerde ölçüm yapılmış hasta). Sonuçlara log transformasyonu yapılmış çünkü orijinal skalada çarpık bir dağılım var. İkinci sonuç ise ACR30 ölçütü. Bu sonuç tüm hastalardan tüm zaman dilimlerinde elde edilmiş, o halde bu sonuçta eksik veri yok.

İlk olarak sürekli tekrarlanan ölçümleri içeren çalışmaları değerlendirmek için yapılan bazı klasik analizleri inceleyelim. Tıp literatüründe, her bir zaman diliminde elde edilen grup ortalamalarını karşılaştıran bir hata çubuğu grafiği vardır. Hasta sayısı da belirtildiğinde, zaman içinde eksik verilerin miktarı ve olası bir hatanın olası çeşidi de bulunabilir. İki tip eksik veri oluşumundan bahsedebiliriz. 1. Herhangi bir deneğin çalışmaya başlayıp, zaman içinde bir noktada çalışmadan tamamen ayrılması; 2. Herhangi bir deneğin çalışmadan tamamen çıkmayıp bazı kontrol ziyaretlerine gelmemesi sonucu bazı verilerin eksik kalması.

Böyle bir çalışmayla ilgili eğrilerde genellikle muhtelif zaman noktalarında hesaplanan p değerleri de verilir. Bunlara bakarak ve yanlış olarak, ilacımızın plaseboya üstün olduğuna karar verebiliriz. Örneğin çalışmanın ilk aylarında gruplar arasında anlamlı fark yok da çalışmanın sonunda anlamlı bir p değeri görürsek, "İlacın işlemesi zaman aldı, o nedenle etki ancak çalışmanın sonunda ortaya çıktı," diyebiliriz. Bu da tabii çoğu kez yanıltıcı olur. Önce şimdi verdiğimiz örnekte olduğu gibi zaman içinde uzunlamasına toplanan verilerin istatistik irdelenmesi birkaç klasik yöntemle yapılır:

1. Önceki örnekte gördüğümüz gibi tekrarlanan anlamlılık testleri. Bunlar arasında da en sık eşlenmemiş Student t testi; incelenen yanıt normal dağılım göstermiyorsa Wilcoxon testi ya da yanıt ikili olduğunda (evet veya evet değil gibi) ki kare testi kullanılır.

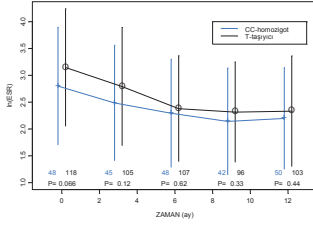
2. Analiz etmek istediğimiz tüm verileri, tek bir değere indirgeyip irdelleyen özet yöntem.
3. 50-60 yıl önce popüler olan iki istatistiksel yöntem, ANOVA ve çok değişkenli ANOVA (MANOVA).

Tekrarlanan anlamlılık testleri hâlâ standart yaklaşımlardan biridir. Her zaman noktasında bir istatistik test uygulanır. Sonuç genellikle anlamlılık derecesine göre 1, 2 veya 3 yıldızla ifade edilir. Ben bunu Meksika restoranlarındaki menülere benzetiyorum: 1 yıldız acılı, 2 aşırı acılı ve NS acısız demektir. Deneysel ve kontrol diyetin karşılaştırıldığı ve serum kolesterolün sonuç olarak kabul edildiği bir diyet örneğinde ise, birçok zaman noktasında anlamsız bir sonuç da alınabilirken sonunda anlamlı bir sonuç alınabilmektedir. BASDAI'nın sonuç kabul edildiği örnekte, bazı "yıldızlar" görülebilir. Her iki örnekte de grup ortalamaları bazı zaman noktalarında anlamlı ölçüde farklı iken diğer zaman noktalarında aynı durumun geçerli olmayabildiğini görüyoruz. Bunun gibi sonuçlara bakıldığında iki tedavi grubu arasında istatistik bir fark olduğu kanaatine varma eğilimi var. Diyet örneğinde, deneysel diyetin serum kolesterol düzeyini düşürdüğü iddia edilebilir. Farkın neden sadece sonunda elde edildiğinin mantıklı bir açıklaması şudur: Diyetin vücutta etkili olabilmesi zaman alır. Peki ya çalışmanın ortasında anlamlı bir sonuç alırsak, ne düşünelim? Aynı şeyi düşünebilir miyiz?

JIA çalışmamıza tekrar dönersek; burada anlamlı sonuca, çalışmanın hem başında hem de sonunda ulaşılmış olduğunu görürüz. Burada tekrarlanan testlerin neden olduğu problemi görüyoruz (1. slayt). Aslında, p değerini belirttiğimiz her noktada bir yorum da yapmış oluyoruz. Yani, karşılık gelen zaman noktalarındaki ortalamaların anlamlı olup olmadığını ifade ediyoruz. Daima %5 oranında bir hata yapmak riskimizin olduğunu hesaba katmazsak kolayca yanlış yönde gidebiliriz. Bu problemin klasik istatistik çözümü ise p değerini düzeltmek. Çoklu testlerin düzeltilmesinde 10 testte anlamlı bir sonuca $p < 0.05$ değil $p < 0.05/10 = 0.005$ ulaşacağımızı unutmamalıyız. Ancak bu yaklaşımın da önemli sorunları vardır.

JIA çalışması

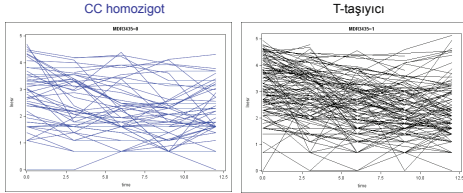
• Klasik analiz (ESR)



1. Slayt

JIA çalışması

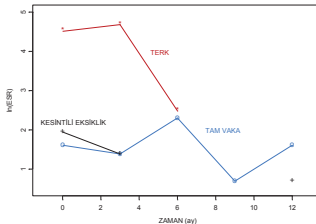
• Tüm bireyler (ESR)



2. Slayt

JIA çalışması

• 3 birey (ESR)



3. Slayt

İlk sorun deneklerde sadece kontrol randevularına düzgün gelmiş olanların bu analizde görünebilmeleridir. Vizitlerine düzgün gelmeyenler analiz dışı kalırlar. İkincisi, bunun yanı sıra bir analiz olma olasılığı vardır. Herhangi bir hastanın o zaman noktasında kontrole gelmemesinin sık bir nedeni hastalığının söz konusu zaman diliminde aktif olmasıdır. Üçüncü ve en önemli sorun ise böyle tekrarlayan analizler, hastalarda, tek tek uzunlamasına yapılan çalışmayı birçok kesitsel çalışma haline getirir. Bütün bu nedenlerden dolayı romatolojide zaman içinde uzunlamasına olan çalışmaların, tekrarlayan zaman noktalarında, tekrarlanan istatistik testlerle analiz edilmesi uygun değildir.

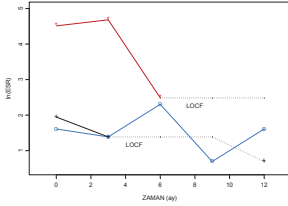
2. slaytta bir spagetti eğri yumağı görülmekte. Bu yumak her bir hastanın zaman içinde kişisel evrimin ayrı ayrı düşünüldüğünde tek tek ayrı eğrilerden oluşur. Yumaktaki eğrilerin her biri başka bir yöne gider ve bütün bu eğrilerden istatistik bir anlam çıkması şaşırtıcı olabilir. Burada yapılması gereken işe her bir eğrinin tek tek basit bir incelemeyle başlamaktır.

3. slaytta bu yumaktan seçtiğim 3 ayrı eğri var. Mavi olan, her vizitine gelmiş, yani çalışmayı eksiksiz tamamlamış bir hastayı simgeliyor. Aynı eğri üzerinde çalışmanın protokolünde saptanmış her zaman noktası için o hastaya ait veri var. Kırmızı olan eğri çalışmayı bir zaman noktasından sonra terk etmiş bir hastaya ait. Siyah olan eğri-deki duruma ise 'kesintili eksiklik' deniyor, yani hasta 3 ve 4. ziyaretlere gelmemiş, sonra devam etmiş.

Zaman içinde uzunlamasına toplanmış verileri analiz yönteminin temelinde çalışmaya katılmış her hastanın eğrisini tek değere indirmek yatıyor. Buna Özet İstatistik Yöntem (summary statistics method, SSM) diyoruz. Bu yaklaşımı JIA çalışması verilerine uyguladık. Her bir profil için eğrinin altındaki alanı (area under the curve, AUC) hesapladık. İlk olarak çalışmada her kontrol vizitine gelmiş olanları değerlendirdik. CC-homozigot ve T-taşıyıcı grupların karşılaştırmasında, eşlenmemiş t testinde anlamlı olmayan bir fark gördük. Veri eksik olduğunda SSM yönteminde sorunlar başlar. Örneğin, çalışmayı tamamlama-

JIA çalışması

- 3 birey (ESR): tamamlanmamış profiller son gözlemin ileriye taşınması (Last-Observation-Carried-Forward (LOCF) yöntemiyle tamamlanmıştır.



4. Slayt

miş olgularınız olduğunda eğrinin altındaki alanı nasıl hesaplıyorsunuz? Alan kesinlikle daha küçük olacaktır, o halde bunu düzeltmeniz gerekir. Bu soruna popüler bir yaklaşım eksik verilerin impütasyonudur. Yani sonucun kayıt edilmediği hallerde beklenebilir bir sonuç tasarlamak ve sonra bu değeri istatistiksel analizde kullanmak. Buna son gözlemi ileriye taşıma da (last observation carried forward, LOCF) denir. Bu konuya daha sonra geri döneceğiz.

JIA çalışmasına eksik veriler için LOCF uyguladığımızda, anlamlı bir sonuç elde edildi. Kısmen eksik verilerin tamamlanması sonucunda istatistiksel olarak anlamlı bir sonuç elde ettik, ki bu durum sonucun açıkça eksik verilerin tamamlanmasına dayalı olduğu anlamına gelir. Önemli noktaya şu ki 'bu tamamlanmış veri doğruyu gerçekten gösteriyor mu?'

Bir profili özetlemenin birçok yolu vardır. Örneğin, her bir profilin ortalama sonlanımını alıp bu değerlere t testi gibi çok basit bir test uygulanabilir. AUC'ye dayalı bir analiz, her hastaya ait veri (o hastanın profili diyebiliriz) ortalamalarının analizi ile aynı değildir. Şimdi 3 kişisel profile ve gerçekleştirilen LOCF'nin (son gözlemin ileriye taşınması) neyi sağladığına bakalım (4. slayt). Bu hasta burada çalışmadan ayrılmış, yani değerleri birbirini takip eden 2 ölçüm dönemine ait ve bunlar da birbiri ile aynı. Ne yapalım şimdi? Bu eksik, burayı tamamladım ve sonra sonuna kadar çizgi çektim. İşte bu 'gerçekleştirilen son gözlemin ileriye taşınması'dır (LOCF). Çok gerçekçi bir impütasyon olduğundan emin değilim. Bu konuya daha sonra tekrar döneceğim.

Özetle, SSM iyi bir yaklaşım olsa da tıp literatüründe çok sık kullanılmıyor, özellikle dengeli veriler için kullanılabiliyor. Tüm hastaların düzenli takip edildiği ve eksik veri olmadığı anlamına geliyor. Düzenli sözcüğünü vurguladım çünkü düzensiz olduğu takdirde bu yaklaşımı kullanamazsınız. LOCF yaklaşımı eksik verilerin tamamlanması için uygun değildir çünkü gerçekçi olmayan verileri tamamlar, gerçek hastalık sürecini hesaba katmaz ve en önemlisi de verilerinizin değişkenliğini göz ardı eder. Bu durumu önceki slaytta görmekteyiz.

Kırmızı profil için ve son gözlemin ileriye taşınması yaklaşımıyla, hastanın sonucu aynı kalmaktadır. Bu sayede, son gözlemin ileriye taşınması yaklaşımının verilerin değişkenliğini göz ardı ettiği sonucuna varmaktayız. Dolayısıyla SSM yaklaşımı bence günümüz romatoloji çalışmaları için uygun değil.

Geri kalan iki ‘standart’ yaklaşım: Tekrarlanan Ölçümler ANOVA ve MANOVA, yaklaşık 50 yıl önce istatistikçiler arasında çok popülerdi ve bu yöntemler tıp ve psikoloji araştırmalarında hâlâ kullanılmakta. Her iki yöntem de dengeli verilere dayanmaktadır ve dolayısıyla bunlar da güncel uzunlamasına romatoloji çalışmaları için uygun değildir. Bunun yanında, ilk yaklaşım korelasyonlarda kısıtlayıcı varsayımlara dayalıdır. Uzunlamasına bir çalışmada hastalarımızı zaman içinde takip ettiğimizden, ilk ve ikinci ölçüm ile ikinci ve üçüncü ölçüm arasında korelasyonlarımızı hesaplayabiliyoruz. Tekrarlanan Ölçümler ANOVA’da time lag (iki dönem arasındaki zaman farkı) dikkate alınmaksızın, bu korelasyonların eşit olduğu kabul edilir. Bizim veri setimiz için de bu durum geçerli değildir. Burada 3 ayrı değişkenin etkisi söz konusu. 1. Tedavinin olası ve bağımsız etkisi (buna grup etkisi de diyebiliriz); 2. Geçen sürenin bağımsız etkisi; 3. Birinci ve ikinci etkilerin birbirleriyle etkileşimi. Bunlar ışığında JIA çalışmasını analiz ettiğimizde, zaman etkisinin (2. etki) ve karşılıklı etkileşimin (3. etki) anlamlı olduğunu gördük. Tedavi etkisi ise (1. etki) ancak, eksik veriler LOCF yaklaşımı ile tamamlandığında anlamlı hale geldi ve etkileşim önemini yitirdi. Ortaya çıkan karmaşayı tahmin edebilirsiniz. Tamamlanmış vakalardan mı yoksa tamamlanmış verilerden mi elde edilen sonuçlara güvenmeliyiz?

Şimdi, demin bahsettiğimiz, Tekrarlanan Ölçümler ANOVA’nın dayalı olduğu veriler arasındaki korelasyonun çeşitli zaman noktalarında birbirine yakın olduğu varsayımına dönelim ve bu varsayımın JIA çalışması için doğru olup olmadığına (5. slayt) bakalım. Tablo 0. ay (bazal hat) ile 3, 6, 9 ve 12. aylarda alınan yanıtlar arasındaki korelasyonu göstermektedir. Açık olarak ortada böyle bir eşitlik yoktur.

MANOVA’nın, Tekrarlanan Ölçümler ANOVA ile aşağı yukarı aynı dezavantajları vardır. MANOVA, ANOVA gibi özel bir korelasyon yapısı varsaymaz. MANOVA’nın güçlü olması için büyük çaplı bir çalışma gerekir. Söz konusu ettiğimiz JIA çalışmasının MANOVA analizi temelde aynı sonucu verir.

Önceki yöntemler uygun değilse, hangileri uygun? Burada uzunlamasına çalışmaların ‘modern’ analizlerine, bu durumda da ESR analizine geliyoruz. Bu nedenle, eksik veri tipleri ve bu verileri ele alma yaklaşımları hakkında konuşacağız. Daha evvel dediğim gibi, eksik veriler analizin en problemlili yönü. Karma etki modelleri ve GEE modelleri hakkında konuşacağız ancak matematik ayrıntılara girmeyeceğiz.

İlk olarak, eksik verilerin tiplerine bakalım. Eksik veri nedir? Açıkçası, bunlar gözlenmemiş verilerdir. Peki bu nasıl oluyor?

- 1– Klinisyenler veya hastalar anket doldurmayı unutmuş olabilirler.
- 2– Hastalar formları doldurmak istememişlerdir.
- 3– Tüp düşüp kırıldığı için kan numunesi yok olmuştur.
- 4– Gelinmeyen doktor ziyaretleri veya randevular vardır.
- 5– Hastalar bu doktora veya kliniğe gelmekten vazgeçmişlerdir.
- 6– Hasta ölmüştür.

Eksik verilerin etkisi nedir? Genellikle etkinin sadece etkinlik kaybı olduğuna inanılmaktadır. Özellikle klinik çalışmalarda böyle düşünülür. Makalelerde çalışma gruplarının büyüklüğü hesaplanırken genellikle şu cümleyi görürüz: 300 hastaya ihtiyaç var ancak %20 oranında bir çalışmadan çıkma beklediğimizden örnek büyüklüğünün %20 artırılması gerek. Eğer sadece örnek büyüklüğü arttırılırsa, eksik veri sorunu göz ardı edilebilir. Aslında, eksik veri olduğunda, doğru elde edilmeyen veriler nedeniyle sonuçlar yanıltıcı olacaktır. Bir örnek vereyim; bir

3. tekrarlayan ölçümlerde ANOVA

JIA çalışması (ESR)

• Eşit korelasyon varsayımı?

ZAMAN (ay)	0	3	6	9	12
0	1	0.71	0.51	0.46	0.42
3	0.71	1	0.61	0.47	0.48
6	0.51	0.61	1	0.70	0.64
9	0.46	0.47	0.70	1	0.69
12	0.42	0.48	0.64	0.69	1

5. Slayt

grupta %10 eksik veriniz ve diğer grupta %10 eksik veriniz varsa, eksik verilerin dengeli olduğunu düşünürsünüz. Yani birinci grupta %10 hasta ilaç işe yaramadığı için ve diğer grupta %10 hasta bu kez tedavi kusursuz bir şekilde işe yaradığı için geri dönmeyecek. Her iki gruptan da elde edilen verinin %90'ı elimizde olsa da geri dönmeme sebepleri farklı. Yani, son analiziniz gerçek tedavi etkisinin yanlış bir tahmini olacak. Her iki durumda da, bir istatistikçi olarak, sorunu ele almanız ve istatistiksel yöntemlerinizi uyarlamanız gerekecektir. Yapmanız gereken, akıllıca, akıllıca sözcüğünün altını çiziyorum, eksik verilerinizi tamamlamaktır. Sorun şu: 'akıllıca' nedir? Asla doğru verileri tamamlayamayacağımızın bilincinde olmalıyız. Bazen insanlar biz istatistikçilere gelip "Gerçekten durum çok karıştı ama bizim sorunumuzu ancak siz çözebilirsiniz, size duacıyız," dediklerini görüyorum. Klinisyenlerle birlikte yapabileceğimiz tek şey aslında birlikte dua etmek.

Bazı terimlerden kısaca bahsetmek istiyorum.

Tekdüze eksiklik: Hasta çalışmadan bir daha dönmek üzere çıkmıştır. Örneğin bir hasta 3. kontrolüne gelmiş ve bir daha geri gelmemiş. Başka bir hasta 2. ziyarette ayrılıyor ve bir daha geri gelmiyor.

Tekdüze olmayan eksiklik: Bazen 'Kesintili eksiklik' de denir. Tahmin edebileceğiniz gibi hastanın kontrole gelmediği, bir ihtimal tatile gittiği veya hasta olduğu, bir daha görünmediği anlamına gelir. Tatilde veya hasta olma yanında belki de sizin ilgilendiğiniz hastalıkla ilgili olarak şehir dışına çıkmış olabilir. Hastanın hasta olması ve bu nedenle hastaneye gelememiş olması da hastalığıyla ilgili olabilir. Diğer bir deyişle eksik veriler istatistik değerlendirilmede yanlışlıklara neden olabilirler.

Rubin eksik veriler için 1975'te, biraz da çapraşık, bir terminoloji önermiştir.

1. Tamamen şansa bağlı eksik veri (Missing completely at random).
2. Şansa bağlı eksik veri (Missing at random).
3. Şansa bağlı olmayan eksik veri (Missing not at random).

İlk anda kulağa aşağı yukarı aynı gibi geliyorlar ama değiller. Şöyle açıklayayım.

İlk eksik veri tipi MCAR'dır (Missing completely at random). Hemen altını çizeyim, burada önemli olan ölçtüğümüz yanıtla ait verinin, örneğin ESR eksikliğidir. MCAR, eksiklik olasılığının tüm yanıtlardan bağımsız olduğu anlamına gelir. Örneğin, dişçilik araştırmalarında rastgele bir diş seçilir, çünkü 32 dişin hepsini çalışmak çok zaman alır. Ya da, bir laboratuvar teknisyeni kan alma tüpünü düşürüyor, değer kaydedilemiyor ancak kaybın yanıtla alakası yok. Veya ESR'ye bakıyorsunuz ama hastayı bir araç kazasında kaybettiniz. Araba kazasının tabii ki ESR ile hiçbir alakası yok ama sürücünde bir ilacın etkisini araştırdığınız başka bir çalışmada durum değişik olabilir. Diyelim çalışma ilacının bir yan etkisi baş dönmesi. Burada tedavi ile araba kazasının bir ilişkisi olabilir. Aynı şekilde hasta bir uçak kazasında öldüğünde hasta pilotun kendisi olabilir. MCAR'ye zararsız eksik veri süreci de denebilir, tepkinin basit ortalamasına 'gerçek ortalamanın yansız tahmini' denir. Kabaca, gözlemlenen verilerin ortalamasını hesaplırsak ve çalışmamızı tekrar edersek, elde ettiğimiz ortalama kabaca MCAR diye değerlendirdiğimiz eksik veriye yakındır. MCAR için çok klasik teknikler uygulayabiliriz, sofistike düzenlemeler yapmamız gerekmez ama bir miktar verim kaybımız olması kaçınılmazdır. Aslında, yukarıda anılan uzunlamasına çalışmalara yapılacak istatistiksel yaklaşımlar dengeli verilere dayalı olmalıdır ve verinin büyük kısmı bu analizde kullanılamayabilir.

Peki ya 'Missing at random' (MAR) nedir? Bu durumda, eksiklik olasılığı gözlemlenen yanıtla bağlıdır. Örneğin, bir anti-hipertansif ilaç çalışmasında kan basıncı kontrol edilemeyen hastalar çalışma dışında bırakılmıştır. Bu durumda eksiklik süreci öyküye dayalıdır. Aslında, hastanın çalışmadan çıkarılacağı önceden tahmin edilebilir. Diğer bir örnek çok aşamalı taramadır. İlk aşamanın sonuçlarından dolayı, bir sonraki aşamalarda veriler eksik kalır.

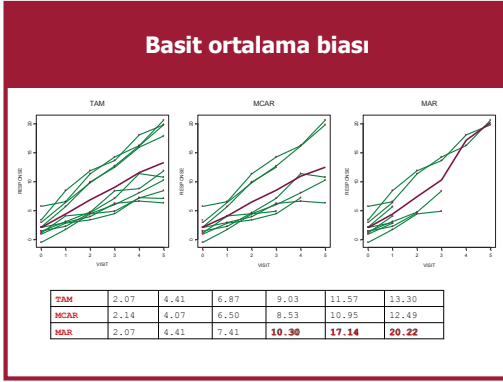
Örneğin elinizdeki iki testten ilki ucuz bir tarama testi, ikinci ise daha ayrıntılı. Diyelim bir hastada ilk testin sonucu negatif. Bu durumda

ikinci teste gerek kalmaz ve bu durumda hastanın ikinci teste ait verisi olmayacaktır. Benzer şekilde, hastaya sigara içip içmediğini, içiyorsa kaç paket içtiğini soruyorsunuz. Hasta sigara içmiyorsa kaç paket içtiğini sorusunu yanıtlamasına gerek yoktur. Bütün bu örnekler MAR olayını tanımlar.

MAR prosesinde verilerinizin basit ortalaması gerçek ortalamanın yanlış bir tahminini verir. Pratikte, hata çubuğu grafikleri tüm hastalar tedaviye devam ettiğinde tedavinin gerçek etkisini tam olarak sergileyemez. Ayrıca, klasik istatistik teknikler (tekrarlanan anlamlılık testleri, özet istatistik yöntemleri, yukarıdaki ANOVA yöntemi) artık kullanılamaz. Ancak, MAR ile MCAR'yi ayırt edebilen istatistik testler vardır. O halde, MAR ile hem etkinlik kaybı vardır hem de doğru yöntemler kullanmazsa sonuçlarda bir yanlışlık ortaya çıkar. Uygun yöntemlerin ne olduğuna daha sonra değineceğim.

'Missing not at random' (MNAR) eksik veri tipinin en zararlı olanıdır. Eksiklik olasılığı hem gözlenen hem de gözlenmeyen olaylara bağlıdır. Örneğin, bir anti-hipertansif çalışmasına dâhil hastanın kan basıncı, çalışma sırasında son derece normal ancak kan basıncında ani bir artış olmuş. Hasta hastaneye veya kendi araştırmacısına gitmemiş, aile hekimine gitmiş. Kan basıncının yeterince kontrol altında olmadığı ve ilacın bu hastaya uygun olmadığı kanaatine varılmış, hasta çalışmadan çıkarılmış. Bu kararın araştırmacıya bildirilmesi gerekirken, bu da yapılmamış. Bu durumda, çalışma kapsamında toplanan verilerle hastanın çalışmadan neden çıktığı öngörülemez. Herhangi bir çalışmada MNAR bulunabileceği hiçbir zaman tam dışlanamaz. MNAR'nin etkisi analizlerin sadece yetersiz ve yanlış olması değildir, sorununun tam ne olduğunu ortaya koyacak bir yol da yoktur. Geriye sadece olası eksik veri sürecini ve sonuçlarını kestirmek ve bunu ana analize eklemek kalır. Bu tür birkaç mekanizma modellediğimizde, buna 'duyarlılık analizi' denir.

Simülasyon egzersizi ise tam yanıtlardan 'missing completely at random' eksik verileri üretme egzersizidir. Bunu yapmak için, her bir ziyarette yazı tura attım ve tura geldiğinde hastayı eledim.



6. Slayt

Ortalama profil hafifçe etkilendi. MAR eksik verisini elde etmek için, yine her bir vizitte yazı tura attım ama bu sefer çıkarılma şansını önceki yanıtı dayandırdım. Yani, önceki yanıt ne kadar zayıfsa çıkarılma şansı o kadar yüksekti. Şimdi ortalama profil orijinal profilinden oldukça farklı. Bunu ortalama değerlere baktığımızda anlayabiliyoruz (6. slayt). MAR'ın altındaki ortalama değerlerin tamamen bozulmuş olduğu açıkça görülüyor. Bu şu demektir, eksik veri prosesi varlığında, makalenizdeki Tablo 1 muhtemelen ciddi bozulmuş verilerle dolu olmalı.

Bu eksik veri sorunlarını nasıl ele aldığımızı bir özetleyelim. İlk ve en önemli konu önleme ve planlamadır. Olmasına izin vermeyin çünkü mutlaka ikinci bir seçenek vardır. Bunu ele almanın tam vaka analizi, mevcut vaka analizi veya olasılık analizi gibi birçok iyi ve kötü yolu var.

'Tam vaka analizi' yaklaşımı sadece tam vakaların analize dâhil edildiği anlamına gelir ve MCAR'de geçerlidir, ancak eksik veri nedeni ile hâlâ verim kaybımız vardır. 'Mevcut vaka analizi' yaklaşımında ise her vizitte sadece orada olan hastalar analize dâhil edilir, yine sadece MCAR'de geçerlidir. Buna örnek, tekrarlanan anlamlılık testi yöntemidir. Diğer bir yaklaşım ise eksik değerlerin imputasyonu ve bunları klasik istatistik tekniikle birleştirmektir. Zaman detaya girmek için yeterli değil. Bu nedenle, kısa bir tanımlama ile yetinmemiz gerekecek. Basit bir değer imputasyon tekniği, basit bir tahminle eksik verinin yerine konulmasını sağlar: Ortalama değer imputasyonda, hastanın eksik yaş verisini ortalama yaş ile tamamlarsınız. Ortalama değer imputasyonunda sorun gerçekçi olmayan imputasyona yol açması ve ölçümler arasındaki korelasyonun göz ardı edilmesidir. Eksik verinin çalışmadaki 'benzer' bir olguya ait değer ile tamamlanması demek olan hot decking de, LOCF gibi önerilmez. Sanırım bu yaklaşımdan kurtulmaya çalışmak benim misyonumun bir parçası. Daha iyi bir yöntem ise, eksik verileri çoklu tamamlamak. Çoklu imputasyon yönteminde, bazı rastgele mekanizmaları birleştiren bir istatistik yöntemle birkaç kez imputasyon yapılır. Farklı tamamlanmış M (sıklıkla 3 – 5) imputasyon veri

setlerinden M ortalamaları ve M varyansları elde edilir. İkinci adımda, M tahminleri genel bir ortalama elde etmek ve gerekli istatistik testler için bir araya getirilir. Bu yaklaşım eksik yanıtlara ve eksik kovaryatlara uygulanır.

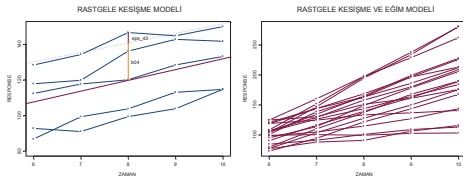
Olasılığa dayalı yaklaşımlar MAR eksik veri prosesleri için de geçerlidir. Bir olasılık yaklaşımında, eksik verilerin impütasyonu gizlidir. İmpütasyon da model temellidir, yani doğrusal yanıtlara yönelik istatistik modeller gizlice boşlukları doldurur. Örneğin doğrusal ve genelleştirilmiş doğrusal regresyon, klasik istatistiksel kesitsel veri modelleri olarak kabul edilebilir. Doğrusal modeller ise lineer karma modeller ve genelleştirilmiş doğrusal karma modellerdir. Doğrusal model, başka deyişle ölçüm modeli, eksik yanıtı otomatik (ve dolayısı ile ekstra model olmadan) tamamlarken, çoklu impütasyon yaklaşımı ek bir impütasyon modeli gerektirir. Alternatif bir yaklaşım ise genelleştirilmiş tahmin denklemleri (GEE) veya ağırlıklı GEE yaklaşımlarıdır. Her ne kadar ilk yaklaşım sadece MCAR'de geçerli ve analize herkes dâhil edilebilirken, ikinci yaklaşım sadece MAR'da geçerlidir. MAR'da kullanılabilen son yaklaşım ise yanlılık yaklaşımıdır, ancak bu yaklaşımda ayrıntıya girmeye gerek yok.

MNAR eksik veri prosesinin en tehlikeli model olduğunu daha önce söylemiştik. Eksik veri prosesinin etkisini anlamak için bir modele ihtiyacımız var. Yani bir öncekinde sadece yanıtla yönelik bir modele ihtiyacımız varken şimdi eksik veri sürecine yönelik de bir modele ihtiyacımız var. Pratikte mantıklı bir eksik veri modeli belirlemek çok karmaşık olup çok nadiren tatmin edicidir. İstatistiksel eksik veri modeli oluşturmaya yönelik birçok çalışma var ama bu modellerin birçoğu aşırı mekanik ve gerçek hastalık süreci ile ilişkisi zayıf.

Karma etki modellerine geri dönelim. 7. slaytta gördüğünüz gibi, (a) bir rastgele kesişme modeli ve (b) bir rastgele kesişme ve eğim modeli var. Rastgele kesişme modelinde profiller temelde genel ortalama profile paraleldir. Rastgele kesişme demek bir birey profilinin genel profile karşılaştırıldığında düşey oynaması demektir. Ancak, bu model daha da karmaşık olabilir, bir rastgele eğim

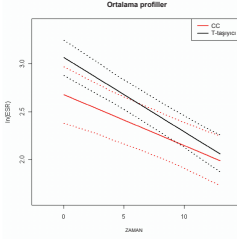
Karma etki modelleri (Mixed effects models)

• İki örnek:



7. Slayt

Karma etki modelleri (Mixed effects models) JIA çalışması (ESR)



8. Slayt

de içerebilir, sadece düşey oynama değil aynı zamanda eğimde düşey oynama görürüz. Yani, bireysel profiller kesişme ve eğimde birbirinden farklılık gösterir (8. slayt). Doğrusal karma modeller uzunlamasına çalışmaların analizinde güçlü birer yardımcıdır, aynı zamanda oldukça sezgiseldirler. Doğrusal karma modellerin ürünü şu tarz sorulara yanıt vermemizi sağlar: Ortalama gelişimde iki grup farklı mı? Ortalama gelişimi belirleyen faktörler nelerdir, rastgele kesilme ve rastgele eğimin değişkenliğini sağlayan faktörler nelerdir? Gaussian yanıtları gibi doğrusal karma modellerinde dağılım varsayımları yer alır, aynı zamanda (eski yanıtlarına göre) ayrılan olguların profilleri çalışmada kalanlarla benzerlik göstermektedir. Bu varsayım uygulamada her zaman geçerli olan bir varsayım değildir ama gerçekleştirilen LOCF'den daha mantıklı ve test edilebilen bir varsayımdır. Bunun yanında, doğrusal karma model dengesiz verilere izin verir, hiçbir hastanın veri eksikliğinden dolayı çalışmadan silinmesi gerekmez. Düzensiz dönemlere de izin verir, profillerin gelişimi ve korelasyon yapısı da dâhil olmak üzere her şey esnek bir şekilde modellenenebilir. Dolayısıyla güncel uzunlamasına romatoloji çalışmalarında bu yaklaşımı öneririm. Ancak bir istatistikçiye ihtiyacınız olacak.

Doğrusal karma modellerde Gaussian yanıtları kullanılır. Genel doğrusal karma modeller, örneğin ikili veya sayılı yanıtlarda veya yanıt sürekli ancak yanlı dağılım gösterdiğinde kullanılabilir.

Tüm hastaların kullanıldığı rastgele kesişim ve eğim modelini JIA çalışmamıza uyguladığımızda açık hiçbir impütasyon olmadığı görülüyor (9. slayt).

Sonraki “modern” yaklaşım GEE yaklaşımı. Bu yaklaşım genelde ikili veya sayısal sonuçlarda kullanılır. GEE'nin avantajı karma modelden daha az varsayıma dayalı olmasıdır. Dezavantajı ise sadece MCAR eksik veri prosesinde güçlü olması ve daha büyük örneğe ihtiyaç duymasındır. Yalnız, olgular ağırlıklandırılarak MAR prosesinde de güçlü hale getirilebilir. GEE yaklaşımını JIA çalışmasında ACR30 yanıtına uyguladık. Hiçbir etki 0.05 seviyesinde anlamlı hale gelmedi.

Karma etki modelleri JIA çalışması (ESR)

- TUM olguları katarak:
 - CC-Homozigot – T-taşıyıcı=
 - Zaman etkisi: $P < 0.001$
 - Genotip etkisi: $P=0.03$
 - Zaman * genotip etkisi: $P=0.09$
 - + TAHMİNLER ve %95 güven aralığı

9. Slayt

Burada, son olarak bazı literatür incelemelerine değineceğim, çünkü daha önce de söylediğim gibi bu modern yaklaşımlar literatürde çok sık kullanılmıyor. 2000 yılına kadar en az üç dönem artritli hastaların takip edildiği ve seçilmiş iki dergide yayımlanmış çalışmaları PubMed'den taradık. Zaman içinde değişebilen yanıtlara baktık. Araştırmada; ilk sonuçlar, ikincil sonuçlar, eksik verileri ele alma yöntemleri, istatistik analiz tekniği ve uzunlamasına araştırma konusunu analiz etmedeki başarılarına göre yazıları sıraya koyduk. 203 çalışmadan 156 tanesi artrit ile ilgiliydi, 11 çalışma JIA hastalarında yapılmıştı ve 58 tane randomize kontrollü çalışma bulunuyordu. Nihayet 98 çalışma ise ileriye dönük olarak izlenmiş kohortlarla ilgiliydi. 110 çalışma seçtik, bunları istatistik metodoloji kullanım uygunluklarına göre sıraladık ve 110 çalışmadan sadece 17'sinin uzunlamasına davranışı en iyi şekilde kullandığını tespit ettik. Ne yazık ki, uzunlamasına bir çalışmayı yürütmek için çok fazla çaba, zaman, enerji ve para harcıyor ama tüm bu çabalardan yeterli veri elde edilemeyebiliyor.

Bu toplantıların harika olduğunu düşünüyorum ama bunları uluslararası topluma daha açık hale getirmemiz gerektiğine inanıyorum. Beni dinlediğiniz için çok teşekkür ederim.

Bir Katılımcı: Belki çok naif bir soru olacak ama sadece başlangıç ve son örnekleri alsanız?

Emmanuel Lesaffre: Bu güzel bir soru, sadece ilk ve son ölçümü alarak başlangıçtan bu yana bir değişiklik olup olmadığı sorusu çoğu araştırmacının cevabını merak ettiği bir soru. Uzunlamasına çalışmanın avantajı, eğer hastalarda ölçüm yapıyorsanız ve çalışmanın sonlanmasından önce hasta çalışmadan ayrılırsa, değerin ne olabileceğine dair duyarlı ve akıllıca bir impütasyonda bulunulabilir. Ama çalışmanın başlangıcında ayrılan hasta varsa ve siz ilk ve son sonuçlara bakarsanız, çalışmanın sonunda sonucun ne olabileceğine dair gerçekten duyarlı ve akıllıca bir impütasyonda bulunamazsınız. Yani uzunlamasına çalışmalar, tekrarlanan ölçümler, eksik verileri açıkça veya gizlice tamamlamanıza olanak sağlıyor. Uzunlamasına bir çalışmanız olsa bile tüm dönemlere bakmanız gerekmediğini

söyleyebilirsiniz, belki sadece başlangıç ve sona bakabilirsiniz, ancak böyle bir yaklaşım, bahsettiğim karma modelde olduğu gibi, çok sayıda veriyi analiz dışı bırakır. Her şeyden önce, tüm veri noktası aralıklarını kullanırsanız iki avantajınız olur; 1) daha verimli olursunuz, çalışmanın gücü artar; 2) hastanın tam profilini çıkarabilirsiniz. Bazen kolesterolü çok yavaş değil de çok hızlı düşen bir hasta da kolesterol hareketi ile çok ilgilenilmektedir. Yani bu tekrarlanan ölçümler, uzunlamasına çalışmalar ve istatistik metodolojilerin her şeyden önce primer soruyla ilgili olanlara göre ek avantajları var. Benim tavsiyem şöyle olacak: Zaman içinde tüm profili modelleyin.

Nathan Vastesaeger: Birçok örneklem boyutu hesaplamalarından bahsettiniz. Hastaların %15-20'sinin 6-12 ay içinde çalışmadan çıkacağını öngörebiliriz, bu nedenle bunu kontrol etmek için örneklem boyutunu artırırız. Siz bunun doğru bir yol olmadığını iddia ediyorsunuz. Bir protokol hazırlanıp son örneklem boyutumuz hesaplamamızda kullanılacaksa, bunu yine de ekleyip, daha sonra eksikliği nasıl tanımlayacağımıza dair tavsiyeniz ne olurdu? Bu soruyla ilgili çok şeffaf olmak için, bunu tanımak ama aynı zamanda sonunda yine de uygun örneklemeye sahip olduğunuzdan emin olmak adına ne tavsiye ederdiniz?

Emmanuel Lesaffre: Sanırım yine önceki sorunun cevabına geliyoruz. Bence hastanızı tekrar tekrar muayene etmelisiniz ve daha sonra da uygun istatistiksel metodoloji kullanarak analiz etmelisiniz. Uygun istatistiksel metodoloji yanlılığın büyük bir kısmını ortadan kaldırabilir çünkü daha önce söylediğim gibi 'missing at random' karma model yaklaşımıyla ele alınabilir ve bu karma model yaklaşımı klasik metodolojilerle ele alınmayan birçok yanlılığı ortadan kaldırabilir. Kısacası, benim tavsiyem şu olacaktır: Evet, örneklem boyutunu arttırın ama aynı zamanda istatistik metodolojinizi de değiştirin.

Nathan Vastesaeger: Ve protokolünüzde eksiklik olacağını belirtin ve bunu neden yaptığınızı açıklayın.

Emmanuel Lesaffre: Yetkililer bunun gibi karmaşık bir analiz yapmanıza izin vermiyorsa, en azın-

dan daha modern yaklaşımlarla ek bir analiz, bir duyarlılık analizi yapın.

Bir Katılımcı: Doğru anladıysam, karma model analizinde mevcut değişkenlerin biçimine bakarak bazı değişkenleri tamamlıyorsunuz.

Emmanuel Lesaffre: Örtülü olarak, açıkça olan bir şey yok.

Bir Katılımcı: Bir şeyleri tamamlamak yerine, örneğin eksik verileri analiz etmenin bir yolu, örneğin bir hayatta kalma analizinde, sansürlü gözlemler kullanmaktır ama farklı olarak gözlemlerinizin bir kısmını kullanırsınız.

Emmanuel Lesaffre: Esasen karma modellerde yaptığımız da bu.

ERKEN ROMATOİD ARTRİTTE UZUN DÖNEM PROGNOZU ÖNGÖREBİLEN ÖGELER VAR MI?

Nathan Vastesaeger

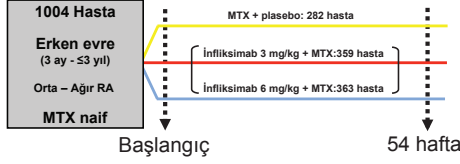
*MSD, Exec. Dir, Medical Affairs
GMA Nordics
İsveç*

Klinik deneylerden elde ettiğiniz bilgiler ve bu bilgilerin klinik uygulamada nasıl kullanacağınız- la ilgili iki konuşma yapacağım. İlk konuşmada, Romatoid Artrit (RA) neleri bildiğimize ve klinik çalışmalardan elde edilen belirli bilgilerin tek tek, hasta bazında nasıl kullanılabileceğine değineceğim. İkinci konuşmamda ise Ankilozan Spondilite (AS) öngörü modellerimizi oluşturmak için kullandığımız metodolojiden bahsedeceğim. Amacım RA ve AS'deki bu verilerin hekime hasta tedavisi ve prognoz tahmini açılarından nasıl bir kılavuz oluşturabileceği. RA ile ilgili ilk konuşmada özet olarak dört konuyu ele alacağım.

1. Prognostik faktörleri kullanmak ne açıdan önemlidir?
2. Prognoz öngörücüleri nelerdir ve hangileri kullanılmalıdır?
3. Prognoz öngörücüleri uygulamada nasıl kullanılır?
4. Uzun vadeli prognozu öngörürler mi?

Burada iki çok önemli çalışmayı ele alacağım, birincisi kendi ASPIRE çalışmamız. Birçoğunuzun aşına olduğu diğer çok önemli çalışma ise, Hollanda'da yapılmış BeSt çalışması. Bu da hedefe tedavi, yani tedavi stratejileri hakkında. Odaklanacağımız konu ise uzun vadede radyografik hastalık ilerlemesi. Radyografik ilerlemenin uzun vadede engelliliğe neden olduğunu göreceksiniz. 'Son Gözlemin İleri Taşınması' (Last Observation Carried Forward - LOCF) hakkındaki bazı tartışmaları da unutmadan, radyografik ilerlemenin tek yönlü olduğunu söyleyebiliriz. Yani hastalar hemen hep kötüye gidiyor. Engellilik ve hastalık aktivitesi zaman içinde düzelebilir ise de radyografik kötüleşme durdurulabilir ama tersine çevrilemez. Prognostik faktörlerin hangi hastanın daha yoğun bir tedaviye ihtiyaç duyduğuna ve bu tedavinin ne kadar erken başlatılması gerektiğine karar verebilmemize gerçekten yardımcı olabileceğini düşünüyorum. İsviçre'den Alex Fink'in yaptığı bir metaanaliz; gecikmiş tedavinin, DMARDS veya biyolojik, sadece ilaçların kullanılmadığı süreyle sınırlı olmadığını gösteriyor. Bu beklenen bir şey. Daha da önemlisi söz konusu kötü etki, gerekli ilaçlar başladıktan sonra da devam ediyor. On iki çalışmayı birleştiren metaanalizde radyolojik

ASPIRE: MTX naif hastalarda MTX-PLB'ye karşı MTX+IFX ile randomize kör aktif kontrollü çalışma



St. Clair W, et al. Arthritis Rheum. 2004;50(11):3432-43.

1. Slayt

hastalık ilerleme hızının, etkili tedavinin gecikmesiyle, Sharp/van der Heijde skorunda yılda 1.4 birim artış olduğu anlaşıyor. Bu da, gerçekten ihtiyacı olan bir hastada tedaviyi geciktirmenin uzun vadeli hastalık sürecine olumsuz etkisi olacağı anlamına gelmekte. Elbette bunlar ortalama sonuçlar ve biliyoruz ki gerçek hayatta bazı hastalarda hiç radyolojik kötüleşme görülmediği gibi bazılarında ise bu çok fazla olabiliyor.

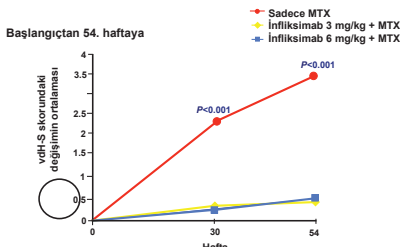
İkinci önemli konu ise sadece tedaviye başlama zamanının yanında hangi tedaviye erken başladığınızın da önemli olduğudur. Birden fazla DMARD + kortikosteroidlere (kombine tedavi) karşı sadece tek bir DMARD tedavisinin etkilerini karşılaştıran ilk çalışmalardan biri olan COBRA çalışmasını hepimizin bildiğini düşünüyorum. Burada, kombine tedaviye alan hastalarda radyografik ilerleme daha başarıyla durduruluyordu.

Bunun yanında aynı hastalar çalışmadan sonra da takip edilmiş ve radyolojik kötüleşme hızı başlangıçta monoterapi alan hastalarda daha yüksek olmaya devam etmişti.

Üçüncü çalışma ise daha detaylı olarak bahsedeceğim ASPIRE çalışması (1. slayt). ASPIRE randomize, aktif, kontrollü bir çalışmaydı. 'Aktif çalışma' demek çok önemli çünkü tüm hastalar metotreksat (MTX) kullanımı açısından naifti, yani daha evvel hiç MTX kullanmamışlardı. Sanırım, hepimiz erken evre bir RA hastası gördüğümüzde tek başına veya başka tedavilerle kombine MTX başlarız. ASPIRE'de karşılaştırılan MTX + plasebo veya MTX + infliximab idi. 1000 hasta, randomize, paralel kontrol gruplarıyla birlikte çalışılmıştı. Esasında infliximab 3 mg/kg ve 6 mg/kg olmak üzere ayrı kolları vardı, ancak bu kollar analizlerde birleştirildi ve salt MTX alan hastalarla karşılaştırıldı.

İkinci slaytta COBRA çalışmasındakine benzer şekilde radyolojik kötüleşme skorlarının ortalaması görülüyor. Kombine tedavi kullanıldığında radyografik kötüleşme belirgin olarak duraklıyor. Bu çok önemli, ortalamalardan bahsettiğimizde ve tek tek hastalara gerçekten ne olduğuna baktığımızda, beyaz çizginin üstündeki hastalar ilerliyor,

IFX+MTX radyografik skorun progresyon ortalamasında anlamlı derecede azalma sağlamakta



St. Clair W, et al. Arthritis Rheum. 2004;50(11):3432-43.

2. Slayt

ama dürüst olmak gerekirse farklar gerçekten küçük, hatta bu açıdan baktığınızda radyolojik kötüleşme o kadar dramatik değil.

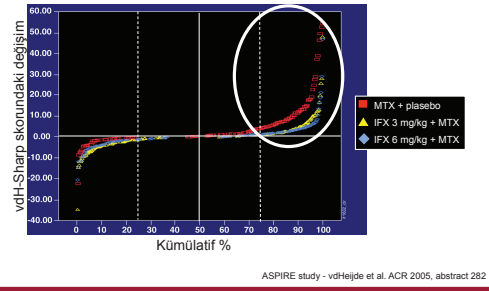
Burada önemli soru şu: Elimizde hangi hastaların kötü, hangilerinin iyi bir prognoza sahip olduğunu bize belirtecek prognoz göstergeleri var mı? (3. slayt). Bunları bilebilmek bize hangi hastaları yoğun bir şekilde, örneğin biyolojikleri erkenden başlayarak, kullanmamızı, buna karşın iyi prognozlu kimi hastaları sadece metotreksat ile izlemede yardımcı olacak. Burada bilmemiz gereken, ilerleme gösterecek hastaları öngörebilmek için prognostik faktörler kullanabilir miyiz? Öyle ki bu hastalarda daha yoğun tedavi, biyolojik tedaviye çok erken ihtiyaç duymayabilir ve hatta sadece metotreksat ile idare edebiliriz.

Prognostik faktörleri saptadığınızda bu hastaları daha iyi tanımlarsınız. Yine çok önemli olan bir nokta, bunlar sadece metotreksat aldıklarında ilerleme göstermeyecek hastalardır. Diğer bir deyişle bu hastalara gereksiz, ekstra tedavi uygulamak için bir sebep yoktur. Birden fazla DMARD, potansiyel olarak biyolojik ama belki sadece daha yoğun tedavi yeterli.

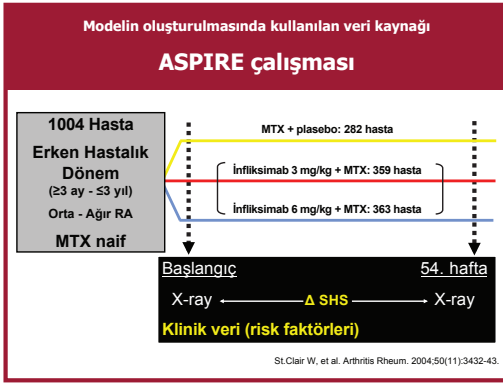
Burada bir 'eşik'(cut off) tanımlı yapmak önemli. Yani bir hastanın kötüleşmesinden ne kastediyoruz? Belki de sadece metotreksat yeterli olmayacak. Bu hastaları nasıl bulabiliriz? Bununla ilgili çok fazla literatür var ve yıllar boyunca tanımlanmış çok sayıda prognostik faktör olduğunu düşünüyorum.

Bu bağlamda hangi hastanın prognozunun iyi ve hangisinin kötü olacağını gösterecek çok yaygın kullanılan CRP, ESH (eritrosit sedimentasyon hızı) gibi testler var. Romatoid faktör ve ACCP'yi çok iyi biliyoruz. Bunlar yanında klinik uygulamada kullandığınız klinik değerlendirmeler de, çeşitli yayınlarda belirtildiği gibi, çok faydalı. Radyografilere gelince; biliyoruz ki başlangıçta radyolojik hasar gösteren hastaların prognozu iyi değil. Bunun yanında özellikle MRI'da kemik iliği ödemi kötü bir prognozu gösteriyor. Yine kötü prognoz göstergeleri arasında kemik metabolizması göstergeleri var ve bunlar arasında da en önemlileri

Prognostik faktörler, ortalamayı değiştiren (kötü prognoz) aykırı (outlier) değerlerin belirlenmesine yardımcı olur.



3. Slayt



4. Slayt

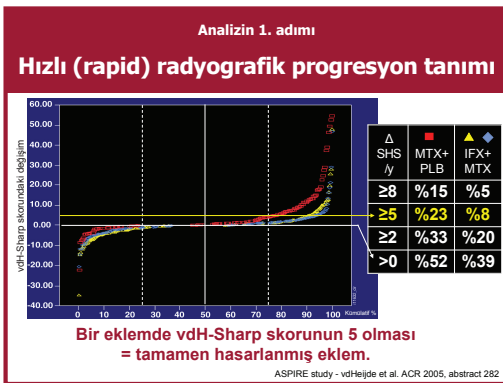
CTX-2 ve CTX-1. Peki, klinik uygulamada bu göstergelerden hangisini kullanalım?

Yeni tanımlanmış prognoz göstergelerini -ESH, şiş eklem sayısı (swollen joint count / SJC) örneği- eskilerle kıyaslayan birçok çalışma vardır. Ancak bu yeni göstergeler geleneksel göstergelerle karşılaştırıldığında iyi performans gösteriyorsa bu, geleneksel göstergelerin de iyi performans gösterdiğini ve uygulamada hazır bulunup hem de kullanılmalarının ucuz olduğunu söyler.

Şimdi size bir kaç slaytta prognoz değerlendirmesi için bir model sunacağım. Tekrar ASPIRE çalışmasına dönelim ve hatırlayalım (4. slayt). Bu bir paralel tasarım, metotreksat monoterapi kolu ile infliksımab + metotreksat kollarını karşılaştırılması çalışması. Burada amacımız Sharp/van der Heijde radyolojik skorunda belirgin bir kötüleşme gösterecek olan hastaları çalışmanın başında nasıl saptanabileceği.

İlk önce, neye hasar diyeceğimize karar vermemiz gerekti. Bunu saptarken okuyucu içi ve arası farkları da göz önüne alan bir radyolojik kötüleşme eşik değeri aldık. Bu eşik değeri önemli. Şöyle ki; diyelim bunu Sharp/van der Heijde skorunda 1 ünite olarak aldık. O zaman görürüz ki 1 yıl sonunda hastaların %52'sinde radyolojik skorda bir değişiklik olmuş. Bu tabii anlamsız. Bu kadar küçük değişiklik radyografileri okumanın hatası içine girer. O nedenle biz eşikimizi 5 ünite olarak aldık. Bunun sonucunda da gördük ki 1 yıl sonunda salt metotreksat alan grupta %23 bir kötüleşme olurken metotreksat + infliksımab kolunda sadece %8'lik bir ilerleme olmuş (5. slayt). Burada aldığımız 5 ünite tabii tamamen isteğe bağlı, ancak biz şöyle bir sebeplendirme yaptık. Eğer bir hastanın tek bir eklemi hastaysa ve 1 yıl sonunda o eklem tamamen harap olursa, söz konusu hastada 1 yıl içinde 5 ünite kötüleşme olacak demektir. Tabii gerçek hayatta böyle olmuyor. Bazı hasta eklemlerde bu kötüleşme 1 ünite, bazılarında da 3 oluyor.

Özetle kanımızca 5 ünite eşikimiz tabii bir yerde gelişigüzel ancak bize okuma dalgalanmalarından ziyade hastalardaki anlamlı değişiklikleri gösteriyor.



5. Slayt

Analizin 2. adımı

ASPIRE'de progresyonun öngörücülerini belirlemekBaşlangıç değişkenlerinin Δ SHS ile korelasyonu

Sürekli başlangıç değişkenleri	MTX + plasebo	IFX + MTX (tek)
CRP*	P<0.0001	NS
ESR*	P<0.0001	NS
SJC	P<0.001	NS
RF-düzeyi	P<0.05	NS

*CRP ile ESR korelasyon göstermektedir (katsayı >0.3)

Başlangıçta X-ışınına bağlı hasar öngörücü olarak kullanılamaz çünkü hastaların >%80'inde mevcut.

Vastesaeger et al. Rheumatology (Oxford) - 2009

6. Slayt

Böylece eşik değerimizi tanımladıktan sonra sıra prognoz göstergelerini saptamaya geldi. Bu amaçla 6. slaytta gördüğümüz gibi önce çeşitli göstergelerin baz (çalışma başlangıcında) olan değerleriyle çalışma sonundaki değerleri arasındaki farkı irdeleyen tek yönlü bir irdeleme yaptık. ESH veya CRP değerlerinden bir tanesini seçip (bir tanesini seçmek yeterli çünkü aralarında anlamlı ilişki var) SJP ve RF titreleriyle ayrı ayrı radyolojik kötüleşmenin istatistik ilişkisine baktık. Ayrıca bu 3 değişkenin birlikte kullanımını irdeleyen modeller de oluşturduk.

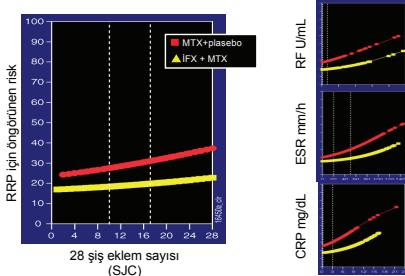
6. slaytta açık olarak görüldüğü gibi bu faktörlerin Sharp/van der Heijde skoruyla ilişkisi esas olarak metotreksat + plasebo grubunda belirli.

7. slayt çok önemli. Burada gördüğünüz gibi değişkenleri hızlı ilerleme olasılığıyla ilişkilendirmekte. Şiş Eklem Sayısı 0-28 arasında değişiklik gösteriyor, Y-ekseni başlangıçta bu özelliklere sahip olup hızla ilerleyecek olan hasta oranını temsil ediyor. Gördüğünüz gibi iki önemli olay var. Hem kırmızı olarak gösterilen metotreksat + plasebo grubunda hem de infliksimab + metotreksat grubunda, SJC arttığında radyolojik ilerleme hızında hafif bir artış var. Aynı ilişki RF, ESR ve CRP için de geçerli. Çok önemli olan ikinci şey ise gördüğünüz gibi eğriler birbirinden uzaklaşıyor. Yani bu göstergeler başlangıçta ne kadar yüksek olursa kombine tedavinin etkisi de daha fazla oluyor.

Şimdi cevap vermemiz gereken bütün bu bulguları hasta bakımından nasıl bir araya getirmemiz gerektiği. Ne zaman metotreksatı tek başına, ne zaman da kombinasyon tedavisini kullanacağız? Burada kardiyologların yaptıkları bir şey bize yol gösterdi. Biliyorsunuz hastalarında kardiyak riski saptamak için cinsiyet, sigara içme alışkanlıkları, yaş, kolesterol düzeyi ve sistolik kan basıncını bir arada kullanıp 10 yıllık bir zaman diliminde ölümcül kardiyovasküler olay riskini öngörebilirler. Sigara içenler, erkekler, yaşlılar, kolesterol düzeyleri yüksek olanlar ve sistolik kan basıncı yüksek olanlar artmış risk grubuna girerler.

Hasan Yazıcı: Dediklerinizi daha iyi değerlendirebilmek için ana varsayıma bakalım. Özetle diyorsunuz ki sadece metotreksat alıyorsanız CRP,

Analizin 3. adımı

Lojistik regresyon ile hızlı (rapid) radyografik progresyonu (RRP) öngörme**7. Slayt**

ESH ve diğer endeksleriniz çalışma sonundaki erozyonlarla iyi ilişki gösterir. Ama metotreksat + infliksimab grubundaysanız durum böyle değildir. Bunun biyolojik açıklaması ne olabilir?

Nathan Vastesaeager: İki husus var; sanırım genel anlamda, daha yoğun olan tedavi hastalık aktivitesini daha çok bastırarak ve dolayısıyla hastalar daha az ilerleme gösterecek, ikincisi ise insanların bahsettiği bu 'ayrıklık' (uncoupling), bunun gerçek olup olmadığını bilmiyorum, belki başka bir sebebi de ayrı bir süreçte bazı hastalar için anti-TNF'nin azalması.

Hasan Yazıcı: Söylemeye çalıştığım şu: Verileriniz iki grup arasındaki farkın büyüklüğü hakkında bize bir şey anlatmıyor. Çok iyi bildiğiniz üzere, bunu Sharp/van der Heijde skoru için yayımladık. Bu skorlamada bir hastanın alabileceği en yüksek skor 480'dir. Halbuki biz bu verilerde 2-3 puanlık değişikliklerden bahsediyoruz. Birçok kez belirttiğimiz gibi bu elbette gerçekçi değil. Klinisyenler olarak bilmemiz gereken diğer bir nokta şu: Verilerinizde ESH ile mükemmel korelasyon gösteren ve anlamlı değişikliğin olduğu erozyon sayısında acaba DAS-28 veya RAPID 3 gibi klinik göstergelerle ilişki nasıl? Bunu bilmemiz önemli.

Nathan Vastesaeager: Haklısınız. Ben de radyografik kötüleşmeye fazla odaklandığımızı düşünüyorum. Ancak burada bir model oluşturmaya çalışıyoruz. Burada anahtar kimi gerçekten tedavi ettiğinizdir.

Hasan Yazıcı: Bu konuda sizinle hemfikirim ama o zaman bana iyileştirebileceğim, dişe dokunur başka bir ölçüt lazım. Bu, örneğin hastanın işine devam edebilmesi olabilir. Öyle ki bu yeni ilaçları hastalarımıza önerirken daha sağlam gerekçeler istiyorum.

Nathan Vastesaeager: Henüz oraya gelmediğimizi söyleyebilirim, hâlâ bunun üzerinde çalışıyoruz. Modelde, size daha sonra göstereceğim, maliyet değeri de hesaplıyoruz. Bu ne anlama geliyor? Bu hastaları sadece metotreksat ile daha yoğun bir şekilde tedavi ederseniz topluma maliyeti ne olur? Bu devam etmekte olan bir çalışmadır.

Şimdi kardiyologlardan uyarladığımız modele dönmek istiyorum. Gösterdiğim slaytta (8. slayt) çok yüksek değerlerden daha aza veya normale doğru değişiklik gösteren bir CRP'miz, düşükten 20 üzerine çıkan bir RF'miz, 17'den fazla sayıdaki eklemde ≤ 10 eklemle inebilen SJC değerleri var. Kutulardaki her sayı Sharp/van der Heijde skorusunda en az 5 puan radyolojik kötüleşme gösteren hastaların oranını temsil ediyor. Örneğin, buradan başlayalım; RF'si düşük, CRP'si yükselmemiş ve diyelim ki 10'dan az şiş eklemi olan bir hasta var. Bu hastanın sadece metotreksat ile tedavi edildiğinde hızla ilerleme gösterme riski %7, kombinasyon tedavisiyle tedavi edildiğinde %4'tür. Bu da bize şunu anlatıyor; monoterapi ile kombinasyon tedavisi arasındaki fark o kadar küçüktür. Bu hastayı başlangıçta bir biyolojik ajan ile tedavi etmeyin derim. Belki de bu hasta birden fazla DMARD'lı kombinasyon terapisine, belki de tek başına metotreksata başka bir ilaç eklenmesine 'step-up' a bile ihtiyaç duymuyordur. Ancak slaydımızda bir diğer kutucukta durum değişmektedir. Burada radyolojik skorun ilerleme olasılığı %47'dir. Bu bir ortalama değil. Bu, en az 5 puan ilerleme gösteren hastaların oranıdır, yani bu ortalamanın çok daha yüksek olduğunu tahmin edebilirsiniz. Burada yaklaşık %33'lük fark ise, CRP'si yüksek, RF'si yüksek ve SJC'si çok yüksek olan bir hastanın riskinin, sadece metotreksat ile tedavi edilenle karşılaştırıldığında, hızlı ilerleme göstermesi olasılığının çok yüksek olduğunu gösterir. Ancak, yoğunlaştırılmış bir tedavi ile, örneğin infliksimab + metotreksat ile bunu bastırma olasılığınız yüksektir.

Özetle bu matris bize, AS için de göstereceğim gibi hikâyenin her iki yönünü, tedavi açısından nerede daha yoğun olmamız gerektiğini veya nerede başlangıçta çok yoğun olmamız gerekmediğini anlatıyor. Buraya olayın toplum ve maliyet boyutlarını da eklememiz gerekirse da tedavi masraflarını karşılayan kurumlar daha çok NNT'ye (number needed to treat = tedavi için gerekli sayı) bel bağlıyorlar.

9. slaytta göstergeleri belirtilen bir hastayı ele aldığımızda radyolojik kötüleşme olasılığı, hasta sadece metotreksat kullandığında %7, metotreksata infliksimab eklendiğinde ise %4. Buradan gi-

MTX-mono ya da kombine tedavi ile ilişkili hızlı radyografik ilerlemeyi öngörme

Progresyon gösteren hastaların oranı ≥ 5 gösterilmektedir

		IFX + MTX			MTX mono			CRP (mg/dL)	RRP risk (%)
		>17	10-17	<10	>17	10-17	<10		
28 Şiş Eklemleri Sayısı (SJC)	>17	8 (5.16)	11 (7.16)	14 (9.20)	33 (22.47)	40 (26.31)	47 (30.59)	≥ 3	
	10-17	8 (5.12)	10 (6.16)	13 (9.15)	31 (21.44)	38 (24.47)	45 (29.56)	0.6-3	
	<10	7 (4.12)	9 (6.15)	12 (7.19)	29 (18.44)	35 (24.47)	42 (27.19)	<0.6	
	>17	6 (4.15)	8 (6.11)	10 (7.15)	17 (11.26)	22 (16.30)	27 (19.37)		
10-17	>17	6 (4.15)	7 (6.10)	10 (7.15)	16 (11.26)	20 (16.30)	25 (19.37)		
	10-17	4 (2.7)	5 (6.11)	8 (4.13)	8 (4.13)	11 (6.15)	14 (7.16)		
	<10	4 (2.7)	5 (6.11)	7 (4.13)	7 (4.13)	10 (6.16)	12 (7.16)		
	<10	4 (2.7)	5 (6.11)	6 (4.13)	7 (4.13)	9 (6.15)	11 (6.20)		
		RF (U/mL)			RF (U/mL)				
		<80 80-200 >200			<80 80-200 >200				

CRP yerine ESR koyulduğunda da benzer bir model oluşturmaktadır.

Vastesaeger et al. Rheumatology (Oxford) - 2009

8. Slayt

Örnek
47 yaş, kadın; SJC=10. CRP=0.55

		IFX + MTX			MTX mono				
28 Şiş Eklemleri Sayısı (SJC)	>17	8 (5.16)	11 (7.16)	14 (9.20)	33 (22.47)	40 (26.31)	47 (30.59)	≥3	
	10-17	8 (5.12)	10 (7.14)	13 (9.15)	31 (21.44)	38 (24.46)	45 (29.56)		
	<10	7 (4.12)	9 (6.15)	12 (7.19)	29 (18.44)	35 (24.49)	42 (27.19)		
	>17	6 (4.15)	8 (6.11)	10 (7.13)	17 (11.26)	22 (16.30)	27 (19.37)	0.6-3	
	10-17	6 (4.05)	7 (6.10)	10 (7.13)	16 (11.23)	20 (16.30)	25 (19.37)		
	<10	5 (3.03)	7 (4.10)	9 (6.11)	15 (9.23)	19 (13.37)	23 (16.32)		
	>17	4 (2.03)	6 (3.10)	8 (4.13)	8 (4.13)	11 (6.13)	14 (7.24)	<0.6	
	<10	4 (2.7)	5 (6.11)	7 (4.14)	7 (4.13)	10 (6.16)	12 (7.26)		
		RF (U/mL)			RF (U/mL)				
		<80	80-200	>200	<80	80-200	>200		

Düşük risk profilili hasta
 $TGS = 1 / (0.07-0.04) = 33.3$

9. Slayt

Örnek
47 yaş, kadın; SJC=18, CRP=3.3 mg/dL, RF=235 U/mL

		IFX + MTX			MTX mono			
28 Şiş Eklemleri Sayısı (SJC)	>17	8 (5.16)	11 (7.16)	14 (9.20)	33 (22.47)	40 (26.31)	47 (30.59)	≥3
	10-17	8 (5.12)	10 (6.16)	13 (9.15)	31 (21.44)	38 (24.47)	45 (29.56)	
	<10	7 (4.12)	9 (6.15)	12 (7.19)	29 (18.44)	35 (24.47)	42 (27.19)	
	>17	6 (4.15)	8 (6.11)	10 (7.15)	17 (11.26)	22 (16.30)	27 (19.37)	
	10-17	6 (4.8)	7 (6.8)	10 (7.1)	16 (11.2)	20 (16.3)	25 (19.3)	0.6-3
	<10	5 (3.8)	7 (6.8)	9 (6.1)	15 (9.8)	19 (13.1)	23 (16.3)	
	>17	4 (2.8)	5 (3.8)	8 (6.1)	8 (4.1)	11 (6.1)	14 (7.1)	
	10-17	4 (2.7)	5 (3.8)	7 (4.1)	7 (4.1)	10 (6.1)	12 (7.1)	<0.6
<10	4 (2.7)	5 (3.8)	6 (4.1)	7 (4.1)	9 (6.1)	11 (6.3)		
		<80	80-200	>200	<80	80-200	>200	
		RF (U/mL)			RF (U/mL)			

Düşük risk profilili hasta
 $TGS = 1 / (0.07-0.04) = 33.3$

10. Slayt

derek böyle bir hastada kombine tedavinin NNT'si 33. Yani bu özelliklere sahip hastalar arasında kombine tedavinin metotreksat mono tedavisine olan üstünlüğü ancak 33 hasta tedavi edilince, bu hastaların birinde ortaya çıkıyor. Bu da söz konusu hastada bir biyolojik ajan kullanmanın anlamsızlığına işaret ediyor. Belki başka kombine tedaviler de bir anlam ifade etmiyor.

Halbuki bir sonraki slaytta anlatılan daha ağır hastada (10. slayt) tedavi edilmesi gereken sayı 2.9. Kardiyoloji ve diğer bilim dallarında kullanılan birçok ilacın NNT'sinin 10-15 veya daha yüksek olduğunu biliyoruz. Bunlar ışığında 2.9 NNT için çok akılcı bir sayıdır ve bize "Evet, böyle hastaları daha yoğun bir şekilde tedavi etmek anlamlı olacaktır" der.

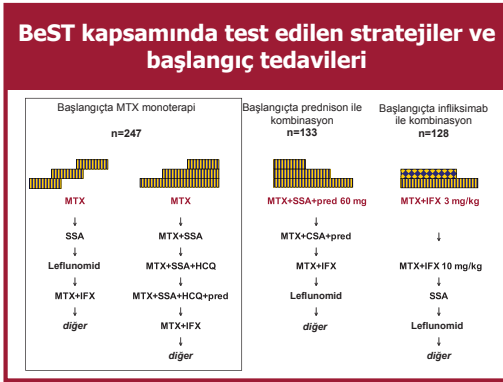
Size ASPIRE çalışmasının tasarımını gösterdim. Daha baştan hastaları bir tedavi grubuna randomize ediyorsunuz, daha sora da bir bütün yıl aynı tedavide tutuyorsunuz. Şimdi siz bana bunun klinik uygulamayla hiç ilişkisi olmadığını söylerseniz tamamen haklı olursunuz. Günlük hayatta bir hastaya ilk 3 ay metotreksat verdikten sonra bir iyileşme görmezseniz tabii başka bir şey yaparsınız. İşte bu nedenle şimdi BeSt çalışmasını geçiyor ve şu soruyu soruyoruz: "Aynı analizleri BeSt çalışmasında tekrarlayabilir miyiz?" BeSt çalışmasının avantajı ilaç tedavisini daha yoğunlaştırma (step-up) rejimlerini ilaç azaltma (top-down) rejimleriyle karşılaştırmasıdır.

11. ve 12. slaytlarda gördüğünüz gibi bir grup hastaya önce metotreksat veya metotreksat + sulfasalazin başlanmış, daha sonra da sizin klinik uygulamada yapacağınız gibi ek ilaçların eklendiği step-up rejimler uygulanmış. Buna karşın başka gruplara daha önce konuştuğumuz infliksimab veya COBRA kombine tedavileri verilip step-down uygulama yapılmış. Burada BeSt çalışmasının genel klinik uygulamadan, doğal olarak, daha ideal koşullarda yapıldığını belirtmem gerekiyor. Görüldüğü gibi hastalar her 3 ayda bir, 'kullanılan tedavi uygun mu' diye titizlikle değerlendirilmiş.

Şimdi sormamız gereken şu: Biraz evvel ASPIRE çalışmasında tanımladığımız ve radyolojik



11. Slayt



12. Slayt

BeST'de RRP'nin öngörücüleri

Öngörücü	OR	%95 GA
RF/ACPA her ikisi de negatif	ref	
1 pozitif	2.5	1.01-6.1
her ikisi de pozitif	4.0	1.9-8.5
Erozyon 0	ref	
1-4	1.3	0.6-3.1
≥4	3.8	1.6-8.9
CRP mg/L <10	ref	
10-35	1.5	0.7-3.2
≥35	4.8	2.3-9.7
Tedavi mono	ref	
kombi prednison	0.2	0.1-0.4
kombi IFX	0.1	0.1-0.3

Visser, ACR 2009, submitted

13. Slayt

ilerlemeyi öngörebildiğini savunduğumuz göstergeler BeSt gruplarında da işliyor mu? Bu arada unutmamamız gereken BeSt çalışmasının çeşitli gruplarından olan hastaların %25 gibi önemli bir yüzdesi bir yıl gibi kısa bir zaman süresi sonunda infliksimab kullanıyorlardı.

Şimdi 13. slayta bakalım. BeSt çalışmasında çeşitli hastalık ağırlığı göstergeleriyle hızlı radyolojik ilerlemenin (Rapid Radiologic Progression / RRP) ilişkisini gösteriyor. Gördüğünüz gibi BeSt çalışmasında kullanılan göstergeler biraz farklı. Bunda da bir sorun yok. Yaptığımız egzersizde, klinik uygulamada gerçekten kullanılan ağırlık göstergelerini kullanmak önemli. Örneğin burada beni dinleyenlerin kendileri hasta bakarken bu amaçla RF, ACCP, SJC, ESH ve CRP kullandıklarını düşünüyorum. ASPIRE'de, anti-CCP değerleri yoktu ve bu yüzden ilgili analizi gerçekleştiremedik. BeSt çalışmasında, anti-CCP ve RF vardı ama düzeyleri yoktu ve bu nedenle analizlerde her ikisi de pozitif diye değerlendirme yapılmış. Bu arada klinik uygulamada gerçekten zor olduğunu düşündüğüm erozyon skoru da kullanılmış. Ayrıca, ASPIRE'de olduğu gibi CRP burada da var. Bundan sonra başlangıç özelliklerine göre hızlı ilerleme gösteren hastaları öngörebilen matris modelleri oluşturuluyor. Başlangıçtaki 4 grup, bu analizler için 2 grup halinde kümelendi. Başlangıçta metotreksat monoterapisiyle başlayan iki grup birleştirilip, analizler için metotreksat monoterapi grubu şeklinde tek grup halinde; başlangıçta metotreksat + prednison veya metotreksat+ infliksimab alan 2 grup ise yine tek grup halinde kümelendi. Göreceğiniz gibi başlangıçta metotreksat monoterapi kullanmış grupta 1 yıl içinde söz konusu gösterge değerlerinde önemli farklar var. Hâlbuki aynı değerler kombine tedavi grubunda böyle farklar göstermiyor. Yine aynı gruplar arasında infliksimab açısından NNT'ye bakarsak (14. slayt) bu sayı metotreksat grubunda 25 iken kombine tedavi grubunda 2.3 ve bu değerler yukarıda ASPIRE çalışması analizlerinde bulduğularımıza çok benziyor.

BeSt çalışması için pozitif prediktif değerler de hesaplanmış ve bu değer söz konusu göster-

Başlangıçta IFX için tedavi edilmesi gereken sayı

Başlangıçta monoterapi					Başlangıçta IFX kombinasyonu				
CRP (mg/L)	≥35			Erozyon (mm)	≥35			Erozyon (mm)	
	47	69	78		11	24	34		
	24	48	58		4	10	15		
	19	37	49		3	8	12		
10-35	22	42	54	≥4	4	9	14	≥4	
	9	20	29	1-4	1	3	6	1-4	
<10	7	16	23	≥4	1	3	4	≥4	
	16	32	43	1-4	3	6	10	1-4	
	8	18	23	≥4	2	3	3	≥4	
	5	11	17	1-4	1	2	3	1-4	
RF ve ACPA					RF ve ACPA				
-/- or +/-					-/- or +/-				

Yüksek riskli hasta profili: TGS=1 / (0.78-0.34) = 2.3

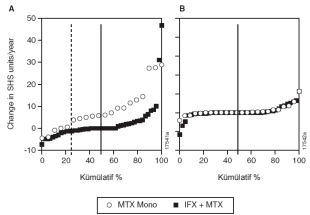
Yüksek riskli hasta profili: TGS=1 / (0.05-0.01) = 25

Visser, ACR 2009, submitted

14. Slayt

Kötü ve iyi prognoz alt gruplarında tedaviye göre progresyon durumu

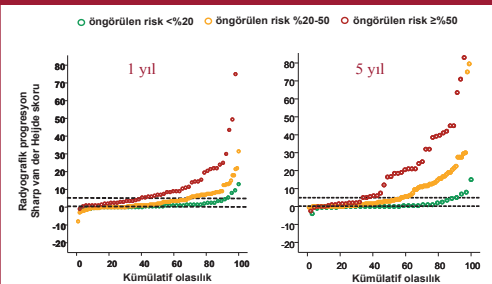
Kötü prognoz grubu (N=65) İyî prognoz grubu (N=68)
CRP>3 mg/dL, RF>200 U, SJC > 17 CRP<0.6 mg/dL, RF<80 U/mL, SJC 10-17



Vastesaeger et al. Rheumatology (Oxford) - 2009

15. Slayt

Risk profiline göre 1 ve 5 yıllık progresyon



16. Slayt

gelerimize göre yüksek risk grubunda %62 olarak bulunmuş. Bu matrisler kimlerin metotreksat ile tedavi edilmesi gerektiğini gösterme konusunda daha iyiler. Aynı zamanda kimlerin ilk kombinasyon terapisiyle tedavi edilmesi gerektiğini de söylüyorlar. Sorun şu ki bu hastaların %38'i ilk kombine tedaviyi aldıklarında aşırı mı tedavi edildiler? Hızlı ilerleme göstermesi öngörülen hastaların profiline bakıldığında, kötü prognozlarının olacağı kesindi, ama aynı zamanda çok yüksek CRP, yüksek SJC gösterdiler, erozyonları başlangıçta, yani prognoz dışında, daha yoğun bir rejimle tedavi edilmek için gerekli nedenler olduğunu söyleyebilirim.

Şimdi yine radyografilere geri dönüp bu modelleri biraz daha irdeleyelim.

15. slaytta ASPIRE'den gelen iki alt grubu karşılaştırıyorum. Randomize bir çalışma olarak ASPIRE'de bir hastanın nasıl tedavi edileceğinin yazı turayla belirlendiğini unutmayın. Ama kötü prognoza sahip 65 hastadan oluşan grupta, erozyon skolası değerlerinin hastaların tedavi yöntemine göre birbirlerinden nasıl ayrıldıklarına dikkat edin. Buna karşın başlangıçta göstergeleri iyi olan hastalarda bir yıl içinde erozyon sayısında tedavi yönteminin bir farka neden olmadığını görüyorsunuz. Evet, bu slayt size aynı zamanda bir grup hastanın da gereksiz yere kombine tedavi aldığını gösteriyor. Ama bilseydiniz, öngörebilseydiniz, başlangıçta bu hastaları yoğun bir rejimle tedavi etmez miydiniz?

BeSt çalışmasında buna ışık tutabilecek 5 yıllık veri var. BeSt çalışmasından hatırlayın. Bir yıl sonunda tüm hastalar ortalama olarak aynı HAQ skoruna ve tüm hastalar ortalama olarak neredeyse benzer DAS-28 hastalık aktivitesine sahip oldular. Soru şu ki; gerçekten yoğun yönetime ihtiyacı olan hastaları tanımlayıp aralarından seçebilir miydik?

16. slaytta BeSt çalışmasından 2 grubu karşılaştırıyoruz. Burada başlangıçta, %20'den daha az ve %50'den fazla risk öngörülen hastaların 1. ve 5. yıl sonunda erozyonları karşılaştırılıyor. Size daha önce gösterdiğim gibi, başlangıç olarak kombine tedavi alan hastaların hiçbirinde %50'den daha fazla risk öngörülüyordu. Diğer bir deyişle bu-

rada kırmızı noktalarla temsil edilen tüm hastalar başlangıçta monoterapi grubundaydı. Bu hastalar ilk baştan COBRA ya da biyolojik ajanlı bir rejimle tedavi edilmiş olsalardı slaytta 1 ve 5 yılda, çok çarpıcı bir şekilde görülen farklar ortaya çıkmayacaktı.

Sonuç olarak diyeceğim söz ettiğimiz göstergeleri siz de hastalarınıza tedavi planı ayarlarken kullanıyorsunuz. Günlük uygulamanızda bunları kullandığınızı, kimleri yoğun rejimle tedavi edeceğinize ve kimleri etmeyeceğinize karar vermek için de bunlardan yararlandığınızı düşünüyorum. Ama bu göstergeleri BeSt'te olduğu gibi modeller halinde kullanmanızın da size yardımcı olacağını düşünüyorum. Nihayet Hasan'ın dediği gibi bu tedavileri karşılayan kurumların bu modellere nerede katılıp katılmayacaklarına da karar vermemiz gerek. Bu konuda devam eden bir çalışmamız var ve gelecekte fırsat çıkarsa, bunlardan da bahsedeceğim.

Hasan Yazıcı: Verilere geri dönüp çeşitli başka ölçütlere göre de analizler yaptınız mı? Örneğin erkek veya kadın, eğitim düzeyi 12 yıl ve üstü ya da 6 yıl ve altı, genellikle ya farklı bir prognoz ya da daha iyi bir prognoz ile ilişkili olan bazı faktörler gibi.

Nathan Vastesaeager: Bu RA çalışmalarında biyolojik parametreler bazında yansız analizler yapmaya çalıştık. HAQ vb gibi şeyler dışarıda bırakıldı. Cinsiyeti dâhil ettik ama verilerimizde ilerleme gösterenlerle göstermeyenler arasında bir fark göremedik. Eğitim düzeyi ise araştırılmadı.

Hasan Yazıcı: Bunu belki yarın tartışabiliriz. Birçok çalışmada, en istikrarlı faktörlerden biri eğitim düzeyi yüksek olan hastaların, herhangi bir hastalıkta, prognozlarının da daha iyi olduğudur.

Nathan Vastesaeager: Sigaranın da çok önemli bir gösterge olduğunu düşünüyorum.

Bir Katılımcı: CRP, SJC ve RF katarak hazırladığınız model bana oldukça mantıklı geldi. Ancak, son derece zıt sonuçlar veren çalışmalar var. Daha bu hafta, SWEFOT araştırması raporunu okuduk. Altı ay metotreksat tedavisi ve ardından iki kol, biri kombine tedavi ve diğeri ise biyolojik tedavinin ardından grubun yaklaşık %60'ının metotreksa-

ta yanıt vermediğini gözlemlemişler. Yine aynı araştırmada yegâne değişkenler HAQ, cinsiyet ve sigara kullanma alışkanlığı olup, CRP, RF ve diğer değişkenlere bu araştırmanın hiçbir yerinde yer verilmemiş. Bu, iki araştırmada uygulanan yöntemlerin birbirlerinden çok farklı olmasından kaynaklanıyor olabilir mi?

Nathan Vastesaeager: ASPIRE modeli, klinik uygulamalarda kullanılamaması nedeniyle eleştirildiğinden bu analizi yapmak için BeSt çalışmasını inceledik. Bu arada BeSt çalışmasını yapanlardan öğreniyoruz ki artık mono metotreksattan kısa bir zaman sonra hastalara ya hemen üçlü DMARD veya bir biyolojik ekliyorlar. Bu da SWEFOT'ta yapıla oldukça benziyor.

SWEFOT'un analizleri devam ediyor ancak ben ayrıntılarını bilmiyorum. SWEFOT'tan elde edilen bu verileri görmek istememin önemli nedeni, tüm hastaların 3 ay süreyle metotreksat almaları ve hastanın hastalık aktivite düzeyi düşükse, örneğin DAS 28 değerinin 3.2'den düşük olması halinde, hastaya metotreksat verilmeye devam edilmesi. Düşük hastalık aktivitesi görülmeyen hastalar da 2 yoğun tedavi rejiminin bir tanesine randomize ediliyorlar. SWEFOT'da gösterilen şuydu: Metotreksat kullanmaya devam eden hasta grubunda hâlâ ciddi ilerleme vardı ve bunun yanında bazı hastaların hastalık aktiviteleri düşük olmasına rağmen, ki bu nedenle bu hastalara çalışmada yoğun tedavi gruplarına alınmıyorlardı, klinik seyirleri kötü gidiyordu. Benim merak ettiğim -ki modelimizde de bu veriler yok- hastalık aktivite düzeyi düşük, ancak hâlâ ilerleme gösteren hastaları nasıl tanımlayacağız? İşte klinikte asıl dikkat etmemiz gereken hastalar bu hastalar. Hikâyemizin 'devam edecek' (to be continued) olmasının nedeni de bu. Hâlâ farklı hasta kohortlarında çalışmaya devam ediyoruz. Çok önemli bir diğer husus da ödeme yapanları ilgilendiren epidemiyoloji meselesi; bu grup, daha işin başında oldukça yoğun bir tedavi uygulamak isteyebileceğiniz kadar büyük mü? Çalışmalar devam ediyor.

Bir Katılımcı: X değerinden yüksek bir Van der Heijde ilerleme hızına sahip hastaların yüzde kaçının 5 yıl zarfında işlerini kaybedeceklerini bilmeden

bu tartışmayı sonlandırabileceğimizi düşünüyor musunuz?

Nathan Vastesaeager: Bu, devam etmekte olan bir çalışma ve Hasan'ın sorusuna benzer bir soru sanırım. Uzun zamandır romatoloji uzmanlarıyla konuşuyorum ve hepinizin farklı bir şeyler beklediğinizi söyleyebilirim; belli bir grup için önemli bir şey bulduğunuzda bir diğer grubun eleştirilerine maruz kalıyorsunuz. Sorun da bu. BeSt gibi bu araştırmaları hepimiz biliyoruz ve burada söylediklerimizin yeni şeyler olmadığını sanıyorum. Yani sorun, bu bildiklerimizi uygulamaya geçirmekte pek ilerleme kaydetmememiz. Her şeye rağmen esas sorun hastanın işine devam edip edemeyeceği ise o analiz de mutlaka yapılmalı. Ancak önemli olan bahsettiklerimizi uygulamaya geçirmek.

Bir Katılımcı: Benim sorum, kanıta dayalı ilaç uygulamaları hakkında. Hastalığın seyri ile ilgili araştırmalarda 1. derece kanıt randomize kontrollü deneylerden değil de prospektif kohort çalışmalarından oluşur. İkinci derece kanıt ise bazı durumlarda randomize çalışmaların plasebo kollarından gelir. Bu durumda sizin araştırmanız, kanıta dayalı ilaç grafiğinde nereye koyacağız?

Nathan Vastesaeager: Söyleyebileceğim, bunun, bir randomize kontrollü çalışmanın HAQ sonuçlarını elde ettikten sonra yapılmış analizidir. Bu konuda başka bir şey diyemeyeceğim. Bana diğer sorulan, sonuçlarımızın neden klinik deneylerle doğrulanmadığıdır. Tedavi için gereken sayıya ulaşmak istiyorsanız (NNT), bir şekilde randomizasyon yapmak zorundasınız. Bu durumda randomizasyona girecek hastalar arasında, kötü göstergeleri olmasına rağmen, hâlâ sadece metotreksata devam etmek zorunluluğu olan hastalar bulunur. Seçtikleriniz arasında hâlâ metotreksat ile tedavi gören kötü durumda hastalar da bulunur. Bu nedenle girişte de değindim; elinizde hatırı sayılır miktarda iyi bilgi var, ancak bunları klinik ortamda nasıl uygulamaya geçireceksiniz? Bana sorarsanız, bu tamamen size bağlı.

Bir Katılımcı: Sanırım, ileriye dönük kohortlarda veri toplayabiliriz, ancak istatistiksel anlamda analitik yöntemler, matris modeller ve çok değiş-

kenli analizler olacak. Bunun bir 'çoklu durum' (multistate) modeli olup olmadığını sormak isterim; gidişatı çok değişkenli analizden oldukça farklı olan enfeksiyöz hastalıklarda bu tabiri kullanıyoruz. Anladığım kadarıyla bu matris modellerdeki hesaplamalar az veya çok birbirinin aynı olduğundan, önce matematiksel olarak bu matrisleri elde ettikten sonra bu modeller arasında bir takım hesaplamalardan yararlanıyorsunuz.

Nathan Vastesaeager: Matrisi oluşturmak için uygunluğu ayarlanmış bir lojistik regresyon kullanıyoruz. Yani, elimizdeki veri setine dayalı lojistik regresyondan yararlanıyoruz. ASPIRE araştırmalarında ASPIRE; BeSt araştırmalarında da BeSt araştırmaları esas alınmıştır.

Bir Katılımcı: Varyans analizinden farkı bu mu, örneğin lojistik regresyon?

Nathan Vastesaeager: Benzer, ancak veriler kutucuk içerisinde elde edilir. Elinizde ana hatlarıyla belirlenmiş kutucuklar vardır ve kutucukların her birisi için bir ortalama elde ettiğiniz söylenebilir; ancak AS modelinde alt gruplardan her birisinde yeterli sayıda örnek elde edebilmek için 72 alt grup oluşturduk. Bu nedenle ek değerlendirmeler yapmanız gerekmektedir ve kutucukların her birisinde bir cevap oranı veya başka tabirle bir ilerleme hızı elde etmek için modelleme yöntemine başvurmanızın nedeni de budur.

Bir Katılımcı: Toplam örneklem büyüklüğü neydi?

Nathan Vastesaeager: 1,004

Bir Katılımcı: Elinizde kaç hücre vardı, 72 mi?

Nathan Vastesaeager: 3x3x3x2. Tedavi için 2, CRP için 3, RF için 3 ve Ortak Düzey için 3.

Bir Katılımcı: Öyleyse bu hücreleri, etkileşim süreçleri gibi bir tür etki değişimiyle mi elde ettiniz; bunlar etkileşim süreçlerinden mi geliyorlar?

Nathan Vastesaeager: Çoklu regresyon analizlerinden elde edilen değişkenleri seçtik; dolayısıyla bunlar bağımsız olarak modele katkı yapıyorlar.

KLİNİK ÇALIŞMALAR VE KLİNİK PRATİKTE ANTİ-TNF TEDAVİSİ İÇİN AS HASTALARININ SEÇİMİ

Nathan Vastesaeger

*MSD, Exec. Dir, Medical Affairs
GMA Nordics
İsveç*

Ankilozan spondilitte de benzer durum var. Burada da doktorlar hangi hedeflerle bir tedaviye başlıyorlar? Doktorlar hastaların kendilerini iyi hissetmelerini, BASDAI, BASFI vb. gibi değerlerin hastalarında düzeldiğini görmek için tedaviye başlıyorlar. Ankiloz oluşumunu önlemek için elimizde yararlılığı kanıtlanmış bir ilaç henüz yok. Dolayısıyla AS'de tedaviye başlamanızın nedeninin bu olduğunu sanmıyorum; hastaları kendilerini kötü hissettikleri için, işe gidemedikleri için, ayakkabı bağlarını bağlayamadıkları, raftan bir şey alamadıkları için tedaviye başlarsınız ve onlarda iyileşme veya düzelme istersiniz. Tedavi kılavuzlarına göre tedavinin başlaması için en azından 4'lük bir BASDAI'ya ve Non-Steroid Antiinflatuar ilaçlardan (NSAİİ) başarısız sonuç alınmasına gerek vardır. Bunlar yanında bir de uzman görüşüne gerek duyulur. Elimizdeki AS tedavi kılavuzları oldukça önemli öneriler içerirler. Hazırlanmalarında öncelikle hastalık aktivitesi ve uzman görüşüne ağırlık verilmiştir. Ancak bu kılavuzlar bize hangi hastaların tedaviye yanıt verip hangilerinin vermeyeceğini belirtmez.

Dolayısıyla biz AS'de de, aynen romatoid artitte yaptığımız gibi önce ağır hastalık öngörücü faktörleri belirledik. Ondan sonra da hangi hastaların, kötü prognozdan çok iyi prognozları olduğunu saptamaya çalıştık. Modeli açıklarken ASAS, BASDAI50 yanıtları vb. hakkında konuşacağım.

Konuşmamın 4 bölümü var. Birinci bölüm modeli nasıl oluşturduğumuz hakkında. Bu, kafanızdaki soruların bazılarına cevap olabilir. Yaptığımız şey, AS'de cevap ve remisyonu doğru öngörmeye yarayacak 6 adet bağımsız öngörücü değişkeni seçerek bir model oluşturmaktır. Öngörme oranını tedaviye herhangi bir yanıt için 1/3, remisyon için ise 1/6 olarak hesapladık; dolayısıyla AS için 6 değişkenden oluşan veri seti herhangi bir tedavi yanıtını 1/3, remisyonu da 1/6 oranında öngörebilecekti. Kullandığımız ASSERT ve GO-RAISE araştırmalarının her ikisi de infliksimab veya golimumabın hastaların geleneksel tedaviye devamını karşılaştıran randomize kontrollü çalışmalardır. Tüm hastalar NSAİİ, bazıları da ek olarak DMARD almaktadırlar. Prognoz göstergeleri olarak çalış-

Başlangıça göre popülasyon	ASSERT	GO-RAISE	Birleşik
N	279	356	635
Erkek (%)	80,6	71,6	75,6
Yaş	39,8±10,2	39,3±12,1	39,5±11,3
BASDAI skoru	6,4±1,6	6,7±1,5	6,6±1,5
BASFI skoru	5,8±2,0	5,1±2,4	5,4±2,2
BASMI skoru	4,0±2,1	3,5±2,2	3,7±2,1
Servikal rotasyon (°)	45,8±21,9	48,8±20,3	47,5±21,0
İnter-malleolar mesafe (cm)	95,3±30,6	100,8±24,8	98,3±27,7
Lateral fleksiyon (cm)	11,2±11,0	11,0±5,8	11,1±8,5
Tragus-Duvar mesafesi (cm)	17,0±6,3	14,2±5,7	15,4±6,1
CRP (mg/dL)	2,4±2,7	1,8±2,0	2,1±2,4
Entezit (%)	63,3	63,7	63,5
HLA-B27 pozitif (%)	87,1	83,4	85,0
Global hasta değerlendirilmesi (cm)	6,8±1,8	7,0±1,8	6,9±1,8

1. Tablo

maya girişteki demografik ve klinik değişkenleri kullandık. Bunların 12 haftadaki BASDAI skorlarıyla 24. haftada ASAS kriterlerine göre kısmi remisyona girme oranlarına etkilerini, iki çalışma için hem ayrı ayrı hem de birleştirip irdeledik. AS'de BeSt'e benzer bir çalışma yok; aslında olsaydı amacımız için iyi olurdu. Ancak AS'de o kadar fazla tedavi seçeneğimiz yok ve dolayısıyla da burada anlatacağım RA'dakinden biraz daha basit olabilir.

1. tabloda veriler özet olarak veriliyor. Gördüğümüz gibi burada genelde AS araştırmaları için alışık olduğumuzdan daha fazla hasta çalışılmış. Özellikle genç hastalar var ve bekleneneği gibi hastaların büyük bölümü HLA B27 pozitif. Hastaların görece genç yaşları modelimizin daha ziyade sağlık ve ekonomi/topluma maliyet ilişkilerinde önemli olabileceğini düşündürüyor.

2. tabloda çalışmanın başında prognoz öngörücüsü olarak aldığımız 20 değişkenin listesi var.

Temelde öngörücü olarak seçerek elediğimiz ve burada liste halinde verilen 20 değişkenle yola çıktık. Tek değişkenli analiz uyguladık. Burada amacımız sadece tedaviye yanıt verenlerle vermeyenlerin başlangıç verilerini karşılaştırmak ve her değişken bazında bunun farklı olup olmadığına bakmaktı. Değişkenler açısından eleme yapılması söz konusu olduğunda P değerlerini, bu egzersiz sadece bir arayacı çalışma olduğundan, 0.1 düzeyinde tuttuk.

Örneğin yaşı ele alalım. ASSERT araştırması, GO-RAISE araştırması ve karma araştırmalarda örnek olarak 12. haftada BASDAI50'ye baktığımızda tedaviye yanıt verenler ile vermeyenler arasında önemli fark olduğunu ve keza ASAS'a göre kısmi remisyona gösteren ve göstermeyen hastalarda, irdelediğimiz öngörücülerin bulunup bulunmaması açısından anlamlı farklar olduğunu görüyorsunuz. Analizlerimizde bundan sonra böyle tedaviye yanıt vermeyen gruplar arasında farklılık gösteren 12 değişken aldık ve ilk aşamada tek değişkenli olarak yaptığımız irdelemeyi bu kez çoklu regresyon analiziyle inceledik.

Başlangıçta öngörücü olarak taranan 20 değişken

Hastalık süresi	Göğüs enfeksiyonu
Yaş	BASDAI
Cinsiyet	BASFI
Vücut kitle indeksi	İnflamasyon (BASDAI Q5-6)
BASMI	C-reaktif protein
Servikal rotasyon	Entezit skoru
Tragus-Duvar mesafesi	Tedavi
Schober indeksi	HLA-B27 genotipi
Omurganın lateral fleksiyonu	Global hasta değerlendirilmesi
İnter-malleolar mesafe	Hasta ağrı değerlendirilmesi

* Görüntüleme verisi ve eklem sayısı her hasta için hesaplanmadı ve dolayısıyla inceleme dışı bırakıldı.

2. Tablo

Çok değişkenli regresyon ve "stepwise" seçim (24. hafta seçim)

	ASSERT		GO-RAISE		Birleşik	
	Tüm 12	Korunan	Tüm 12	Korunan	Tüm 12	Korunan
Yaş	X	X	X		X	
BASDAI	X		X		X	
BASFI	X		X	X	X	X
BASMI	X	X	X		X	
Servikal rotasyon	X		X		X	X
Intermalleolar mesafe	X		X		X	
Tragus-duvar mesafesi	X		X		X	
Entezit	X		X	X	X	X
CRP	X	X	X		X	X
Global hasta	X		X		X	
HLA-B27	X		X		X	
Tedavi	X	X	X	X	X	X
ROC-AUC	%84	%80	%78	%74	%79	%79

*Tüm = Tek değişkenli taramada p<0.1 değerli çok değişkenli regresyon modelleri.

*Seçim = 24. haftadaki ASAS kısmı remisyon için öngörü değişkenlerinin "Stepwise" seçilmesi.

3. Tablo

3. tablonun birinci sütununda çalışmanın 24. haftasındaki ASSERT ve BASDAI50'yi ne kadar öngörebildiklerini anlayabilmek için seçtiğimiz 12 değişkene ait veriler var. Burada yapılan bir hassasiyet analizinde ROC eğrisinin altında kalan alanlara baktık. Eğri altındaki alan bir modelin tahmin hassasiyetinin ölçümünde kullanılır. %70-80 aralığında ve özellikle de %80'in üzerinde oranlara ulaştığımızda oldukça iyi bir tahmin modeliniz var demektir. Tüm 12 değişkeni dâhil ettiğiniz zaman BASDAI50 için bu 12 değişkenden sadece 4 tanesi ROC eğrisinde %80 üzerinde bir değer gösteriyordu. Böylelikle ASSERT'den de sadece 4 değişkeni eklediğimizde 12 toplam değişkenle ROC'ta %84'e yakın bir hassasiyet elde edebiliyoruz. GO-RAISE'de de benzer sonuçlar elde edildi. Bundan sonra, 3. tabloda görüldüğü gibi her iki çalışmayı birleştirdik ve bir birleşik veri tabanı için 5 gösterge ortaya çıktı. Ancak analizlerimize devam ettik ve sonunda 8 göstergeli bir modele ulaştık (4. tablo).

Baktığımız değişkenler arasında örneğin yaş faktörünün sadece 12. haftada anlamlılığını koruduğunu, BASMI'nin bir kere anlamlılığını koruduğunu, buna karşın CRP'nin, tedaviye yanıtı öngörmekte her zaman anlamlılığını koruduğunu görüyoruz.

Örneklem büyüklüğümüzün fazla sayıda öngörücü elde etmemize izin vermediğini düşündüğümüz için öngörücü sayısını 8'den de 6'ya indirelim dedik. Bir biyo-istatistikçiden görüş aldık ve birleşik veri dizindeki toplam 635 hastalık bir örneklemde 100 hasta başına 1 öngörücü almanın daha anlamlı olacağını düşündük. 8'den 6'ya inerken değişkenler arasındaki ilişkiyi daha derinlemesine araştırdık. İşlevsellik ile ölçüm arasında önemli ilişki vardır. Bu nedenle hastaların servikal rotasyon ve BASMI değişkenlerini değerlendirme dışı bırakmaya karar verip, sonunda 6 değişkene karar verdik. Bundan sonra da bu 6 değişkenli modeli bir çoklu regresyon modeline uyarladık. Gördük ki bu model hem tedaviye yanıt hem cevap hem de remisyon için oldukça iyi öngörücü oluyor.

Aşamalı seçim sonrası korunan 8 değişkenin özeti

	ASSERT		GO-RAISE		Birleşik	
	12. hafta cevap	24. hafta remisyon	12. hafta cevap	24. hafta remisyon	12. hafta cevap	24. hafta remisyon
Yaş	X	X				
BASFI	X		X	X	X	X
BASMI		X				
Serv. rotasyonu						X
Entezit skoru				X	X	X
CRP	X	X	X		X	X
HLA-B27			X		X	X
Tedavi	X	X	X	X	X	X
ROC-AUC	%81.2	%80.1	%74.7	%74.0	%76.9	%78.5

4. Tablo

Daha sonra bu değişkenleri sınıflandırarak bir matris modeli elde ettik ve bunu AS'ye uyarladık. Matrisin bu son şeklinde 1. Entezit bulunup bulunmadığı; 2. HLA B27'nin bulunup bulunmadığı; 3. BASFI; 4. CRP; 5. Yaş ve çok önemli olarak 6. hastanın ne tedavi aldığı vardı. İşte bu modelle incelediğimizde gördük ki hazırladığımız matris bize anti-TNF tedavisinde BASDAI50 yanıtını %81 duyarlılıkla veriyor. Ama bunun yanında da görüyorsunuz %5 hastada anti-TNF kullanılmış ancak sonuçta bir değişiklik olmamış. Geriye dönüp bakınca bu hastaların sadece NSAİİ ile de tedavi edilebileceklerini düşünüyoruz. Hazırladığımız matris remisyona girebilecek hastaları da öngörmekte başarılı ancak tam remisyonu öngörmek daha zor. Özetle 9 yıl hastalık süresi olan bir AS kohortunda hazırladığımız matrisin duyarlılığı %55 bulundu. Bizce bu hiç de fena değil.

Bunun yanında hiç enteziti olmayıp, HLA B27 pozitif, BASFI'si görece düşük ancak genç olan hastalarda, özellikle inflamasyon göstergeleri varsa, yine anti-TNF kullanmak akılcı olur ve bu onları remisyona dahi sokabilir. Bir de şu durumu düşünelim. Diyelim bir hasta var. HLA B27 taşıyor, CRP'si normal, BASFI'si çok yüksek- yani birçok yakınması var ve neresine bassanız ağrı duyuyor, yani entezit skoru yüksek. Böyle bir durumda düşünmeniz gereken belki hastanın esas sorununun AS değil de fibromiyalji olabileceği. Böyle hastaları herhalde anti-TNF ile tedavi etmemek gerek.

Bir hastanın beli eğrilmiş ve omurgası tamamen sabitlemiş ise bu hastanın düzelebilmesi son derece güçtür. Yani hastayı remisyona sokamazsınız. Yüksek CRP, pozitif HLA B27, basınç testleri sırasında entezit skorunun alınamaması ve anti-TNF ajanı ile tedavi BASDAI50 ile ilişkilidir. Hem ASSERT ve hem de GO-RAISE araştırmalarında birleşik veri seti, oldukça iyi bir öngörü hassasiyeti sağladığı gibi, tedavi edilebilen ve tedaviye gerek duymayabilecek olan RA modeline benzer sonuçlar elde edilmektedir.

Emmanuel Lesaffre: Ana hedefiniz sadece öngörmek mi, yoksa neden bu öngörüü yaptığınızı anlamak mı? Eğer ana hedef sadece öngörüye böyle aşamalı eleme usulüyle bu kapsamlı çalış-

mayı yapmanıza gerek yok. Bunun yerine daha sabit regresyon katsayıları elde etmenizi sağlayacak başka bir hesaplama yaklaşımı benimseyebilirsiniz. Çok sayıda regresyon değişkeni ve bir o kadar az veriniz varsa buradaki güçlük sonuçlarınızın pek stabil olamayacağı. Yaptığınız analizleri gerçekleştirmek için başka yöntemler de bulunuyor.

İkinci değineceğim nokta ise sonuçlarınız fazla iyimser. Öyle ki kurallarınızı saptamak için bir takım veriler kullanıyor, sonra da dönüp böyle oluşturduğunuz kurallarla söz konusu verilerin geçerliliğini (validasyon) irdeliyorsunuz.

Nathan Vastesaeger: Ben, hiçbir şekilde “geçerlilik ölçüyoruz” demedim. Ancak tabii bu yapmak istediğimiz bir şey. Dediğinizi anlıyorum. Dediklerimin epidemiyolojik sonuçlarına değineceğim. Bu arada belirtmek isterim ki biz bir ilaç şirketiyiz ve elimizde sadece bizim yaptığımız çalışmalarla ilgili veri tabanları var. Bahsettiğim çalışmalar da bir yerde ön çalışmalar.

Emmanuel Lesaffre: Hazırladığınız matrisin geçerliliğini bağımsız bir veri tabanında doğrulamanın başka yolları da var.

Nathan Vastesaeger: İlk sorunuza geri dönecek olursak, istediğimiz en iyi öngöründe bulunmak mıydı, yoksa klinik uygulamada anlam ifade eden ve doktorlara yardımcı olabilecek bir şeyler yapmak mıydı? Sonuçta elde ettiğimiz değişkenlerin klinik uygulamada kullanıldığını ve dolayısıyla da klinik uygulama açısından önemli olduğunu düşünüyorum.

Emmanuel Lesaffre: Bir doktor olarak önemsiz değişkenlerle uğraşmayacağınız için hepsi önemlidir. Anlamlı şeylerle başlarsanız anlamlı sonuçlar elde edersiniz.

Nathan Vastesaeger: Bu noktada haklısınız; bu tartışmaya daha sonra devam edelim, ancak bunlar, klinik uygulamada elde edilenlerden ziyade klinik çalışmalardan toplanan veriler ve belki de entezit skoru dışında tamamı klinik uygulamalardan derlenmiştir. Bence doktorlar, enteziti skorlamaktan ziyade var veya yok şeklinde değer-

lendiriyorlar. Bu da modelin zayıf yönünü ortaya koyuyor. Öngörü göstergelerinin birbirleriyle ilişkileri ise modelimizde belirleyici oluyor. İşin iyi tarafı, modelimizin sadece 12. haftadaki bazı ölümlerin değil 24. haftadaki hastalık durumu için de öngöründe bulunmamıza olanak tanıyor. Burada BASDAI50'nin ölçümü ASAS20'den daha zor, hesabı biraz daha zor ancak her ikisi de esasta hastalardan elde ettiğiniz verilere dayanıyor. Bunun yanında ASDAS, hasta yakınmaları ve CRP değerlerinden oluşuyor. CRP'nin önemini demin vurguladım. O nedenle de ASDAS20 ile elde ettiğimiz sonuçlarda doğal olarak önemli de etkisi olacaktır.

Olasılık oranlarına (Odds Ratio / OR) bakarsanız, ASDAS veya BASDAI cevaplarının anti-TNF tedavisi ile çok yüksek derecede ilişkili olmaları yanında, ASDAS'ın cevapla ilişkisini gösteren OR çok daha yüksek. Dolayısıyla anti-TNF ile tedavi edilen hastaların, BASDAI'larından çok ASDAS'larında önemli iyileşme gözliyoruz. Bu çok önemli. Çünkü şimdiye kadar yapılmış anti-TNF çalışmalarında kuvvet hesapları hep BASDAI50, ASAS20 veya benzeri bir göstergeye göre yapılmış. Hâlbuki gösterdiğim gibi ASDAS, anti-TNF yanıtına çok daha hassas. Bu nedenle üstünlük göstermek amaçlı ilaç çalışmalarımızın kuvvet hesaplarını ASDAS bazında yaparsak anti-TNF'lerin etkinliğini daha az sayıda denek çalışarak gösterebiliriz. O nedenle bundan sonra böyle çalışmalarda ben ASDAS kullanılmasını öneriyorum.

Bir Katılımcı: Anti-TNF'ye karşı geleneksel uygulama karşılaştırmasında, kaba (crude) OR mi yoksa modeldeki diğer değişkenlere göre düzeltilmiş (adjusted) OR mi kullandınız ?

Nathan Vastesaeager: Kaba OR.

Bir Katılımcı: Öyleyse, elinizde ayrıca düzeltilmiş bir değer de var mı?

Nathan Vastesaeager: Düzeltilmiş değer, az ya da çok, tasarladığımız modellerde zaten yer alıyor; çünkü tüm parametreleri siz seçiyor ve tüm bilgileri de siz giriyorsunuz. OR için özel bir düzeltme yapmadık.

Bekleneceği üzere CRP, ASDAS formülündeki haliyle, sadece hasta yakınmalarına dayanan BASDAI50'ye kıyasla, ASDAS değerleri ile çok daha yakın ilişkilidir. Bunun yanında BASDAI ve BASFI değerleri, aralarında yüksek derece ilişki gösterirler. Bu, modelimizde nasıl yorumlanmalıdır? Öncelikle, BASDAI'nın BASFI ile çok daha yakından ilişkili olduğundan bahsetmiştim. Bu ilişkiyi bu iki yöntemle elde edilen %81-55'e karşı %81-72'lik yüzdelere göre görüyoruz. Ancak ASDAS'ın, anti-TNF'ye cevabın değerlendirilmesinde çok daha iyi sonuç veren bir araç olduğunu söylemiştim. İşte ASDAS'ı kullandığınızda hastaların cevap oranları %40'lara kadar çıkabilir. Burada unutulmaması gereken, cevap oranları genelde çok daha düşüktür. Beklendiği şekilde anti-TNF tedavisi ve NSAİİ ile tedavi arasındaki fark belirgin olacak. Üçüncü nokta CRP'nin etkisidir. Düşük CRP grubunda bir ASDAS cevabı elde edilmesinin son derece güç olduğu görülürken, aynı durumda bir BASDAI cevabı elde edilme olasılığı hâlâ çok yüksektir. Bunlar önemlidir. Konuştuklarımız, bir ilaç çalışmasında seçeceğimiz birincil sonlandırma noktası için doğru hasta grubunun belirlenmesi ve seçiminde size yardımcı olacaktır.

Şimdi bir derleyelim. Anlattıklarım ışığında CRP'yi dikkate alarak hasta seçmenin AS deneylerinde çok şey ifade ettiği düşüncesindeyim. Halen AS tanısı konulan hastalarımızda yapılan çalışmalarda hastalık aktivitesinin ölçütü olan BASDAI'nın 4'ün üzerinde olması gerekmektedir. Bu şemaya CRP eklemenin –RA'da da olabileceği gibi– ilaç çalışmalarında yararlı olabileceği kanatımda. Ayrıca ASDAS yanıtının, daha küçük çaplı çalışmalar için daha iyi bir sonuç ölçütü olabilir. Etkinlik (effectiveness) değerlendirmesi yapar ve olaya tedavi masraflarını destekleyen kurumlar açısından bakar ve ayrıca söz konusu hastaların genç, yaş ortalamalarının da 39 olduğunu göz önünde tutarsak işaret ettiklerimin hayati ve çözüm bulmamız gereken sorunlar olduğunu düşünüyorum. Ayrıca parasal verimlilik açısından elde edilecek kazanç da oldukça yüksek olacaktır, diye düşünürüm. Elimde henüz verimlilik verileri yok ancak bu konuda çalışmalarımız devam ediyor.

12. hafta BASDAI50 cevabının karşılaştırılması

CONTINUE NSAID +/- DMARD				ADD ANTI-TNF THERAPY			
HLA-B27 neg.				HLA-B27 pos.			
CRP 0.6-2.0			%21	540		%70	no
							Enthesitis
TGS = 1 / (cevap hızlarının mutlak farkı)							
NNT = 1 / (0.70 - 0.21) = 2.0							
BASFI [4.5-6.5]				BASFI [4.5-6.5]			

1. Şekil

Şimdi bir de NNT sayılarına bakalım (1. şekil). Bir grup hastada NNT 22.7. Peki, bu hastaları bir anti-TNF ajanı ile derhal tedavi etmeye başlamak istiyor muyuz? Bilmiyorum, ancak daha dikkatli ve kapsamlı bir değerlendirme yapmak isteyebilirsiniz. Burada size örnek olarak sık görülen bazı hastalıklarda NNT sayıları vermek istiyorum. Antibiyotiklerle 14 gün süreyle otitis media tedavisi = 7; 38 hafta boyunca üçlü AIDS tedavisine karşı ikili tedavi = 19 (12 ila 50 arasında değişiyor); Sağ bitlenmesi = 1. Bunlar yanında tedavi masraflarını karşılayan kurumların bazı ilaçlar için istedikleri bazı NNT değerleri var.

2. şekilde 72 alt grubun her birisinde tedavi edilmesi gereken hasta sayının verildiği bir modelimiz var ve diğer hastalıklar için NNT değerlerini de verdim. Bakınız, ankilozan spondilit tedavisinden genel anlamda tüm alt grupların fayda sağladığı net olarak görünüyor. Tedavi edilmesi gereken hasta sayısına bakıldığında büyük çoğunluk için değer 5'ten düşük olduğunu ve ancak çok az sayıda alt grupta 10'un üzerine çıktığını görüyoruz. Yine burada farklı ASAS sonuçları görüyorsunuz. Tedavinizle hastalıkta tam remisyon bekliyorsanız bunun sağlanmasının son derece güç olduğunu biliyoruz. Ancak, yine de çoğu hastada remisyon için tedavi edilmesi gereken hasta sayısının sadece 2 olması da gerçekten şaşırtıcı.

Buradan giderek modellerimizin ileride uzman hekimlerin ve tedavi karşılayıcı kurumların verecekleri kararlarda yardımcı olacaklarını söyleyebiliriz.

Konuştuklarımızın günlük uygulamaya nasıl etki edeceği konusuna dönelim. Deminden beri konuştuğumuz hastalar esasında ilaç çalışmaları-na katılan, seçilmiş hastalar. Günlük uygulamada kime tedavi verip vermeyeceğimize nasıl karar vereceğiz? Bu amaçla İspanya'da yapılan REGISPONER ve Belçika'da yapılan ASPECT araştırmasından elde edilen sonuçları bir araya getirdik. Bunlar, romatoloji uzmanları tarafından fiilen izlenen hastalar üzerinde yürütülen epidemiyolojik araştırmalar olduğundan AS nüfusu ve romatoloji uzmanı tarafından tedavi edilen hastalar konusunda fikir verirler. Güncel tedavi kılavuz-

12. hafta BASDAI50 cevabı için tedavi edilmesi gereken sayı

CRP>2.0				CRP 0.6-2.0				CRP<0.6			
HLA-B27 neg.				HLA-B27 pos.				HLA-B27 neg.			
3.6	2.2	2.1	540	2.3	2.0	2.1	no	9.3	4.0	3.7	no
4.5	2.5	2.4		2.7	2.0	2.0	yes	12.8	5.2	4.7	yes
5.5	2.8	2.6		3.0	2.1	2.1	no	16.3	6.3	5.8	yes
7.3	3.4	3.1		3.8	2.3	2.2	yes	22.7	8.5	7.7	yes
TGS				TGS				TGS			
<3				<3				<3			
3-5				3-5				3-5			
5-10				5-10				5-10			
>10				>10				>10			
H: Hayır				H: Hayır				H: Hayır			
E: Evet				E: Evet				E: Evet			
BASFI [4.5-6.5]				BASFI [4.5-6.5]				BASFI [4.5-6.5]			

2. Şekil

ları sadece 4'ün üzerinde BASDAI skoru olan hastalarda biyolojik kullanılmasını önerdiklerinden biz de böyle hastaları analiz etik. Bu 2 çalışmayı birleştirdiğimizde, özet olarak şunları gözledik: 1. Hastalar genellikle 40 yaşın üzerindedir; 2. Hastaların çoğunda CRP değerleri yoktur, bazı hastalarda hafif bir CRP yükselmesi vardır; az sayıda hastada ise CRP çok yüksektir. 3. Hastaların BASFI dağılımları bizim ilaç çalışmalarında gözlediklerimize çok benzemektedir. Çok ilginç olarak hastaların %48 gibi oldukça büyük bir yüzdesinde NNT 3'ten düşüktür. Diğer yandan BASDAI skorları 4'ün üzerinde olan hastaların hemen hepsi tedaviden yararlanmışlardır. Diğer çok önem verdiğim bir nokta ise BASDAI skoru 4'ün altında olan (ki kılavuzlara göre bu hastalar anti-TNF alamaz) bazı hastalar vardır. Bunların, aynı zamanda, BASFI'leri 4.5 üzerinde olup, HLA B27'leri pozitif ve çok yüksek CRP değerleri vardır. İşte biz bu hastaların tedavi kılavuzlarına göre anti-TNF alabilmeleri için hastaları değerlendirirken BASDAI değil, ASDAS kullanılmasını öneriyoruz.

Sonuç olarak 6 öngörücü öğeden oluşan matrisimizin, ROC'lerle gösterdiğimiz üzere, gerek RA

gerekse AS'de, çeşitli zaman noktalarında, tedaviye verilecek yanıtı öngörmekte yararlı olduğunu söylüyoruz. Hangi hastaya hangi tedaviyi vereceğiniz konusunda bunlardan yararlanabilirsiniz. Aynı öngörücüler tedavi – destek kurumlarına da yardımcı olabilir. Bundan sonraki EULAR toplantısında kılavuzlardaki BASDAI skorlarını ASDAS skorlarına çevirtmeye çalışacağım.

Bir Katılımcı: NNT hesaplamasını RA bağlamında yaptığımızda aldığımız kıstas erozyonları önlemek. AS'de bunu yapmamızın amacı, %50'lik bir cevap oranına ulaşan hastaları görmek, doğru mu?

Nathan Vastesaeger: Doğru bir noktaya değindiniz; ancak benim göstermek istediğim, hastalar açısından tüm sonuçlar için NNT sayısını hesaplayabileceğinizdir. “Bu veya şu sonucu kullanın” demiyorum, ne size ne de ilaç tedavilerini karşılayan kurumlara. Ancak, size tüm bilgiler ve bunlardan elde edilen sonuç ve seçiminize göre uygun gördüğünüz kararı alabilirsiniz. NNT sayıları klinik ilaç çalışmalarının hepsinde hesaplanıyor ve tedavi-destek kurumlarına yardımcı oluyor. Bu son konu üzerine çalışmalarımız da devam ediyor.

Davetiniz için teşekkür ederim. Bu toplantılara belki de Türkiye'nin en nezh epidemiyoloji ve istatistik tartışma ortamı diyebiliriz. Bizim için de bir tür kamp gibi, yaz öncesi, ne yapmışız diye tartışıyoruz. Bazı çalışmaları metadolojik açıdan gözden geçirmek çok faydalı oluyor.

Metaanaliz aslında bir yönüyle oldukça basit, bildiğimiz, kullandığımız veya okuduğumuz bir yöntem. Ben birazcık 'bu yönteme ne zaman, ne kadar güvenebiliriz,' buna değineceğim. Sonuna doğru da vakit kalırsa Türkiye'den bir takım örneklerden, kendi yaptığımız çalışmalardan söz edeceğim.

Metaanalizde, küçük küçük çalışmaların, farklı yönlerle doğru giden, etki tahmini, çok farklı yönlerini görüyorsunuz (1. slayt). Tüm bu borazanları, tek bir borazan haline getirmek, hedef bu. Bu çalışmaların tek bir vücutta toplanması ve tek bir sonuç olarak sunulması. Tabii her türlü irili ufaklı çalışmalardan böylesine güzel, net bir sonuç çıkmayabilir. Hedef bu aslında, metaanalizden beklentimiz küçük küçük çalışmaları keşke bir araya getirsek, ortak bir sonuca gitsek, ne güzel olur şeklinde.

Metaanaliz temel olarak klasik derlemelerden, İngilizce 'narrative review' denilen derlemelerden farklı olarak bir kuantifikasyon, yani sayılarla sentez yapmaya imkân tanıyor. Diğer 'conventional review' veya klasik derleme dediğimiz şeye kalitatif özetleme de diyebiliriz. Kalitatif olarak işte şu yazar bunu söyledi, bu yazar bunu söyledi diye farklı farklı çalışma sonuçlarını sunmak olabilir. Bundan farklıdır. Bir de verilerin bir araya getirilip analiz yapılmasından yine farklıdır. Aslında verileri farklı merkezlerden toplayıp bir araya getirip bir sonuç yapabiliriz. Örneğin multi center çalışmalar böyle olabilir veya eş-dost küçük verilerin bir araya gelmesinden oluşan çalışmalar böyle olabilir ama yine de bundan da farklıdır. Çünkü normalde verilere ulaşmadan, yani veri tabanına ulaşmaksızın, elinizde makale olduğunu varsayarak bu makalelerden yola çıkarak yapılan, yeni bir analiz yöntemidir. İlk çıkışı 1944 Fisher ile olduğu söylenir. Daha sonra biliyorsunuz, Cochrane modeli gelişmiş ve çok da popüler olmuştur.

METAANALİZ

Önder Ergönül

*Enfeksiyon Hastalıkları Anabilim Dalı
Koç Üniversitesi Tıp Fakültesi*

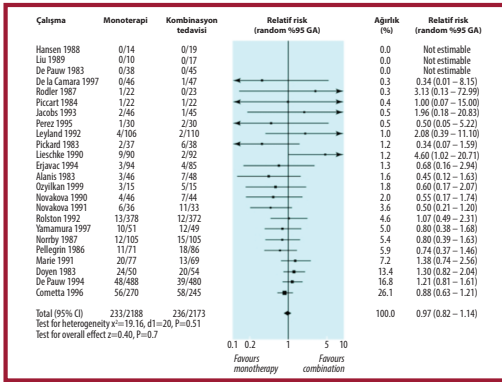


1. Slayt

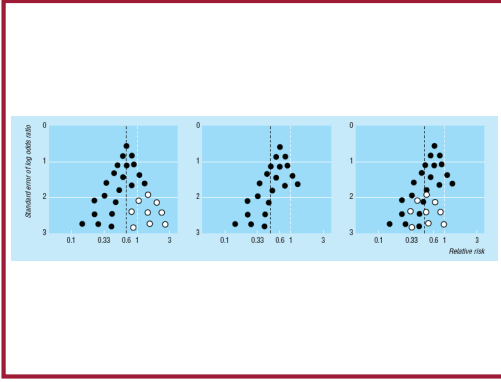
Metaanalizi evidence-based hiyerarşi içerisinde nereye oturtalım? Yaklaşık 10-15 sene önce metaanaliz ilk yaygınlaşmaya başladığında en üstte oturtuluyordu şüphesiz. Herkes acaba bu konuda bir metaanaliz var mı, diye soruyordu. Eğer metaanaliz varsa sanki orada laf bitmiştir, son söz söylenmiştir gibi bir algı vardı. Ama son yıllarda bunun azaldığını görüyoruz haklı olarak. Artık, metaanaliz varsa da nasıl acaba, diye düşünmeye başladı insanlar. Sadece Türkiye değil, dünyanın her yerinde metaanaliz hemen piramitin en üstüne oturtulmuyor. Daha alt değerlerde yer alabilir. Hatta bazen bir olgu sunumu bile çok kıymetli olabilir. Kanıta dayalı bir hiyerarşi modeli oluşturuldu, bu çalışmalar tarif edildi ama sonuçta en altta yer alan uzman görüşü, açıklayıcı tanımlamalar ve olgu serilerinin de yine hakkını yememek gerekiyor. Bu açıdan bakıldığında tarihsel olarak önemli çalışmalardan bir tanesi streptokinaz çalışması. Metaanalizi önemli kılan, bugüne gelmesini sağlayan, cesaretlendiren çalışma. 1959-1988 arasında 33 tane streptokinaz aktivitesine yani pıhtılaşmayı engelleyici aktivitesine dair 33 çalışma yapılmış. Bunların 6 tanesi anlamlı, 27 tanesi de anlamlı olmayan şekilde bulunmuş ve bu kafa karışıklığını acaba nasıl giderelim, diye düşündüklerinde bunları bir araya toplamış ve bir metaanaliz çalışmasında çok etkin olarak anlamlı olduğunu göstermişler. Sanırım 1991’de, Lancet’te yayımlanan bir çalışmada, streptokinazdaki kullanım, daha sonra birçok çalışmaya da örnek teşkil etmiş. Tabii özellikle küçük çalışmalar varsa, yeterli örneklem büyüklüğüne ulaşamıyorsa, ki romatoloji alanında bu tür hasta grupları söz konusu olabilir, örneğin Türkiye’den küçük bir çalışma, Arjantin’den, Yunanistan’dan başka yerlerden, bütün bunları bir araya getirdiğimizde acaba yeterli örneklem büyüklüğüne ulaşabilir miyiz? Metaanalizin yapılmasındaki en önemli motivasyonlardan bir tanesi bu ama hepsi bu değil. Ciddi motivasyonlardan bir tanesi ulaşamadığımız örnek, örneklem büyüklüğüne acaba bir araya getirerek, çalışmalarını bir araya toplayarak ulaşabilir miyiz? Önemli bir motivasyon oluşturuyor. Böylelikle sample size’a neden ulaşıyoruz? Belli bir örneklem büyüklüğüne ulaştığımız zaman, istatistik anlamlılığı yakalama ihtimalimiz oluyor. Eğer küçük çalışmalarla sınırlı

kalırsak istatistiksel anlamı yakalayamayız, p değeri atfedilen değerden yüksek kalır, yüksek kalırsa da etkisizmiş gibi bir sonuca ulaşabiliriz, oysa o anlamlılığı yakalamak için çalışmalar yapıyor. Klasik bütün metaanaliz çalışmalarında yer alan 'forest plot' yani bir tür 'orman grafiği' diye çevirebiliriz Türkçe'ye. Ben aslında birçok kavramı Türkçe'ye çevirme konusunda çaba sarf ediyorum ama çok da başarılı olduğum söylenemez.

2. slaytta her bir çalışma bir satır olarak yazılıyor, daha sonra her çalışmanın karşısına bir kare görüyoruz ve bir de ok veya bir çubuk görüyoruz. Çubuk çalışmanın güven aralığını gösteriyor. Ve kare, çalışma ne kadar büyürse güç açısından yani örneklem sayısı ne kadar büyürse, doğal olarak güven aralığı o kadar daralıyor. Yani ortaya konan sonuç, daha stabil, bizim daha çok güvenebileceğimiz bir tahmin yapmamıza neden oluyor. Güven aralıkları daraldığı zaman, aynı şekilde karenin de büyüdüğünü görüyoruz. Ortadaki kare de giderek büyüyor, sonuçta elmas şekli bütün çalışmaların sonucunu özetlemiş oluyor. Bir odds ratio, hazard ratio kullanıyorsak 1'in altı koruyucu, üstü koruyucu olmayan derecede önemi işaret ediyor. Son yıllarda bazı çalışmalarda gerek hazard ratio kullanan, yani effect estimate olarak hazard ratio veya odds ratio kullanan çalışmalarda metaanaliz olmasa da farklı sonuçlarını, parametrelerini burada yer alan değişkenlerini bu tarzda sunabiliyorlar. Böyle değişik bir gösterime rastlıyoruz son yıllardaki makalelerde. Her bir metaanaliz sonucunda böyle bir tablonun olması mutlaka gerekliliktir. Ama aynı zamanda her bir metaanaliz makalesinde olmasa da ideal bir şekilde farklı uçlardan, farklı taraflardan yayınlar varsa, yani ilk başta gösterdiğimiz gibi borazanın bir kısmı sol tarafta, bir kısmı sağ taraftaysa, ortaya çıkan sonuçlar her iki yönde ise o zaman metaanaliz daha da anlam kazanır. Eğer hepsi aynı yöndeysen ortaya çıkan metaanaliz var olan çalışmaları daha da güçlendiren yöndedir ama farklı yönlerde farklı görüşler söz konusuysa bunlardan oluşturulacak metaanaliz kıymet kazanır. Bu amaçla da funnel grafik denilen yani ters huni grafiği denilen bir grafik yapılmasında fayda vardır (3. slayt). Hepsinde göremezsiniz, makalelerin bir kısmında vardır. Bu şunu vermeye çalışır. Ne kadar simetrik yayılmış-



2. Slayt



3. Slayt

sa burada çıkan sonuçlar diyelim rölatif risk ya da odds ratio'da olabilir. O ölçüde çalışmaya dâhil edilen çalışmaların sonuçları birbirine yakın veya birbirinden uzak gösterilebilir. 3. slaytta görüldüğü gibi çok asimetrik şekilde kalmış olabilir ama neyi gösterir, yapılamaz mı, yapılabilir. Bunların her durumda da metaanalizi yapılabilir hatta çok farklı, çok basit durumlarda da metaanaliz yapılmaya başlanmıştır, yapılmaktadır.

Metaanaliz nasıl yapılır? Öncelikle çalışmaların seçilmesi sorunu var. En önemli sorunlardan bir tanesi, kimleri nasıl dâhil edeceğiz, hangi çalışmaları alacağız? Elbette birinci sırada randomize kontrollü çalışmalar. Varsa eğer randomize kontrollü çalışma oldukça önemlidir. Ama hep şu vurguyu yapmak gerekiyor. Hiçbir yöntem randomize kontrollü çalışma veya metaanaliz veya bir başka yöntem fetiş haline getirilmemelidir, tıpkı p değeri gibi. p değeri fetişizmi dediğimiz klinisyenlerde görülen bir yaklaşım. p değerine bakınca “Tamam, ne kadar güzel işte, her şey tamam” gibi bir bakış açısı var. Benzer şekilde randomize kontrollü çalışma için de “Randomize kontrollü çalışma, ne kadar hoş”. Tamam, ama doğru mu yapıldı, yanlış mı yapıldı, tasarım oldukça önemli. Randomize kontrollü çalışma olunca ismi tasarımdaki hatalardan ya da taraflılıklardan ya da bias’dan uzaklaşmış olmuyoruz. Gözlemsel çalışmalar dâhil edilebilir mi? Biraz sonra biraz daha ayrıntılı konuşacağız. Evet, bence edilebilir, bu hep tartışılan konulardan bir tanesidir. Öyle gözlemsel çalışmalar olabilir ki randomize kontrollü çalışma kalitesine yaklaşabilir, zaten literatürde yapılmakta. Yayımlanmamış veriler kaynak göstererek yine yapılabilir, abstraktlar da yapılabilir hatta editoryal, rakam varsa içerisinde nadiren rakam olur aslında ama editoryallar da düzgün bir editoryal yazılmışsa elbette, bütün bunlar olabilir. Makalelerin seçimi oldukça meşakkatli bir süreç, bunların akış diyagramı içerisinde toplam ne kadar tarandı, belli kaynaklardan Medline’den veya Cochrane’den ya da başka veri tabanlarından tarananlar. Bunların ne kadarının atıldığı, neden atıldığı yani hariç tutma kriterleri, dâhil etme kriterleri bunların çok detaylı yazılması lazım. Mümkünse ikiden çok araştırmacının olması gerek çünkü tek kişi bazen subjektif olabiliyor. Bazı makaleleri dâhil

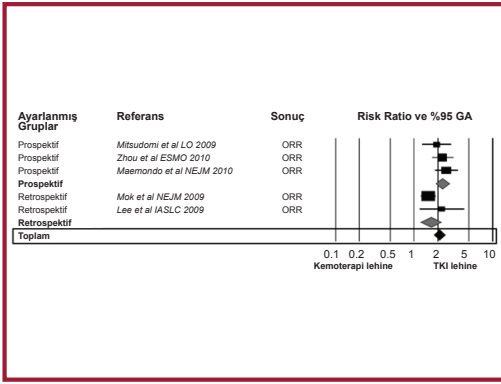
edip etmemek konusunda objektiviteyi tutturmak için ve mümkünse her bir çalışmanın, etki tahminleri, ‘effect estimate’ ya da bazen ‘effect size’ diye kullanılan etki tahminleri, bunlar odds ratio olabilir, rölatif risk olabilir, hazard ratio olabilir, bu tür etkilerin saptanması belirlenmesi oldukça önemli. Bu tür etki tahminleri neler olabilir, ortalama da olabilir veya yüzde bile olabilir.

Elimizde kontrol grubu olan bir veride analiz yapılabilecek olan bir makalede diyelim ki belli bir kesim bir ilaca maruz kaldı, bir ilaçtan aldı. O ilacı aldıktan sonra, basit kohort gibi düşünelim, yan etki veya ölüm gerçekleşti ya da gerçekleşmedi, iki olasılık var. İlaç almayan grupta da yine gerçekleşti veya gerçekleşmedi, toplam 4 olasılık. Bu 4 olasılık varsa ve yüzde üzerinden verilmişse yine bunlarla da metaanaliz yapılabilir ya da en popüler olanı risk ratio, odds ratio veya hazard ratio olması. Ya da korelasyon varsa yine korelasyondan da metaanaliz devam ettirilebilir. Yalnız burada sorun şudur. Eğer ki birinden gidiyorsak, mesela odds ratioyu kullanıyorsak, diğerlerinin sonuçlarını da odds ratioya çevirmemiz gerekir. Bunun için bir ara işlem yapmak gerekir. Yoksa hepsini beraber aynı şekilde katarsak bir sonuç çıkmaz. Ya da hazard ratio ile içerisine zaman faktörü girdiği için survivor analizinde kullanılan, odds ratio ile aynı havuza atamayız. Bu söz konusu olamaz ama yüzdeden veya buradaki risk ratiyodan bir odds ratio yaratabiliriz. Bu tür modifikasyonlar yapmak gerekebilir, hatta korelasyondan bile bunu yapmak mümkün, bunun bir takım matematiksel formülleri var. Çok kompleks bir iş değil.

İki temel yöntemle çalışmakta. ‘Fixed effect’ ve ‘random effect’ diye iki temel sistemi var metaanalizin. Fixed effect veya common effect deniyor. Burada dâhil edilen makaleler çok daha düz, birbirine çok daha benzer, çok daha yakın bu farklı makalelerin katılmış olması sadece örneklem büyüklüğünü arttırmış oluyor, güçlendirmiş oluyor. Ama diğer yandan random effect’te çok daha birbirinden farklı, çok farklı özellik gösteren makaleler var. Bunların bir tür ortalaması sisteme girmiş oluyor. Yani matematiksel arka planda bu varsayımlara göre farklı hesaplamalar yapılmakta.

Örnek olarak, 2011’de Annals of Oncology’de bir onkoloji ilacı için yapılan çalışmayı ele alalım (4. slayt). İlginç olarak bütün çalışmalar zaten genel olarak çalışmanın sonucunu destekliyor, yani bütün hepsi aynı yönde. Burada ‘fixed effect’ yani sabit, hepsinde ortak özellik olan bir yer kullanılmış durumda. Hepsinde ortak kullanılan etki tahmini hazard ratio. Böyle bir analiz yapmak son derece kolay. Hazard ratiosu olan 3 tane, 5 tane makale var ortada, bunların softwarerler kullanarak yapılması mümkün. Değilse bunlar birbirine çevrilebilir. Farklı bir şekilde kullanılabilir ama dediğim gibi bazılarını çevirmek oldukça zor olur. Bir kısım sonucu hazard ratio veriyorsa diğeri odds ratio veriyorsa bunları aynı havuza koymak hatalı sonuçlara neden olabilir. Daha önemlisi verinin kalitesinin belirlenmesidir. Buna göre elmayla armudu yan yana koymak yerine, eğer ki çok farklı farklı veriler varsa, farklı farklı makaleler varsa bunları mutlaka gruplamak gerekir. İlla bunların hepsini metaanalize koyacağım, diye bir şart olmayabilir. Sonuçta şunu unutmayalım, ana evrenimiz sistematik derleme. Sistematik derlemenin içerisinde bir kısmını belki metaanaliz yaparız. Yapılabilir, böyle şeyler var, ama elimizde bir sürü de olgu sunumu vardır örneğin. Olgu sunumlarını katacak değiliz ya da tek kollu çalışmalar katılamaz metaanalize ama onları da farklı bir başlık altında irdelemek mümkün. Yani onu irdelersek bunu neden metaanaliz yazısına kattınız gibi bir soru olmaz. Analize katmadan kullanabiliriz. Burada tabii ‘precision’ çok çok önemli, Türkçe’de kesinlik diye kullanalım, kesinliği nasıl ölçeriz? Güven aralığı ne kadar darsa o kadar kesin anlamına gelir. Veya daha bir descriptive bir veriden söz ediyorsak ‘standard error of the mean’, SEM diye kısaltılan bir şekilde gösterilebilir ya da 1/varyans üzerinden de gösterilebilir. Güven aralığı zaten bir tür varyanslardan elde edilen bir şeydir. Ve bu kesinliği nasıl sağlarız esas olarak? Sample size ile. Sample size’ı büyüttüğümüz ölçüde elbette güven aralıkları küçülür. Ve çalışma tasarımıyla bunu sağlayabiliriz.

Sistematik ve random bias çok önemli, bias olup olmaması metaanalize dâhil edilen çalışmalarda oldukça önemli, eğer bir sistematik



4. Slayt

hata varsa örneklem büyüklüğünün artmasıyla bu değişmez. Hata, hata olarak kalmaya devam edecektir. Ama eğer sistematik olmayan bir hata varsa örneklem büyüklüğünün artmasıyla azalabilir. Bu önemli bir şey ve ne kadar örneklem artarsa güven aralığı da küçülür. O yüzden bu bias'leri tarif etmek oldukça önemli. Ama metaanalizde önemli olan yine sistematik bir hatanın bias'i olup olmaması. Buna göre seçilen çalışmalar, her bir çalışmanın seçimini de yine tıpkı bir çalışmaya seçilen örnekler gibi düşünebiliriz, denekler gibi düşünebiliriz. Ve bunların nasıl bir tarzda oluştuğu, nihayet publication bias çok önemli, çünkü genelde sonucu pozitif olan çalışmalar yayımlanır, negatif sonuçlar yayımlanmaz. O yüzden burada ciddi bir publication bias'den bahsedebiliriz. Yine metaanalizde, metaanalize başlamadan önce de mutlaka güç analizinin yani örneklem büyüklüğü hesabından kaynaklı bir güç analizi yapılması önerilmektedir. Fixed effectten bahsederken söylediğimiz gibi bu oldukça kolay, 10 çalışma varsa her birinde 100 denekten bu kabaca 1000 eder gibi bir mantıkta düşünebiliyoruz ama daha heterojen çalışmaların olduğu random effect modelde bu daha güç olabilir. Burada şöyle bir şeyden söz edelim, öncelikle metaanalizin mantığı nasıl? Diyelim ki 3 çeşit çalışma var. Biz bu 3 çalışmayı dâhil ediyoruz. Bunların hepsi exposure ve no exposure diye sınıflanmış olsun. Exposure burada ilaç olabilir veya herhangi bir etki, bir maruziyet. Bu etkiye maruz kalanların bir kısmında olay sonuçlanır, bir kısmında sonuçlanmaz. Bu çalışmaları karşılaştırmamız için önce böyle bir tablomuzun olması lazım. Elimizde toplam 4 adet olasılığa dair rakamların olması lazım mutlaka. Eğer böyle 4 rakam elde edebiliyorsak bu çalışma her durumda metaanalize dâhil edilebilir. Ama bunu elde edemezsek tek bir yolla olursa metaanalize dâhil edemeyiz.

Çalışma	Etken var		Etken yok		Etki (örneğin OR)	Güven aralığı
	Sonuç var	Sonuç yok	Sonuç var	Sonuç yok		
A	5	15	3	30		
B	10	20	5	20		
C	5	25	2	10		
	Σ=20	Σ=60	Σ=10	Σ=60	Σ OR	Manthel-Hanszel

5. Slayt

Bu açıdan bakıldığında şöyle basit bir mantıkla düşünebilirsiniz. Diyelim ki metaanaliz hiç olmasa ben bunları alt alta yazsam buradaki var olan ve olmayan sonuçları toptasam, buradan bir odds ratio elde etsem, acaba bu metaanaliz aynı sonucu verir mi vermez mi, ne dersiniz? Vermiyor. Eğer verse zaten metaanalize gerek kalmayacak. Bu klasik hoş bir hesap, yani bunu yaparsınız, her

bir çalışma için bir satır, bunun odds oranı bellidir ama hepsini yapıp, ortaya çıkarıp da odds elde ederseniz metaanalizden bulduğunuz oddstan farklı. Bunun adı Simpson's Paradox. Ve dolaşısıyla halktan verilerin toplanması metaanaliz sonucundan farklıdır, metaanalizde uygulanan sistem odds oranları arasında bir tür toplanma, Mantel Haenszel denen bir yöntem ama (5. slayt) elbette bu odds oranlarını toplayıp üçe bölerseniz bir ortalama o olmaz, bir tür ağırlıklı ortalamaya benzer bir şekilde hesaplama yapılmakta. Elbette elle de yapılır, hep bilgisayar programlarına atıfta bulunuyoruz ama elle de yapılabilir. Elle yapılma mantığı ters, $1/\text{odds ratio}$ gibi giden ağırlıklı ortalamanın tersi üzerinden bir hesaplama. Ama tabii ki bunun formülünü verip hadi bunu hesaplayalım demek çok yanlış değil.

Bir Katılımcı: Ben yanlış mı biliyorum, aslında ideal olanı eğer elinizde ham veri varsa bunu yapmak, eğer ham veri elimizde yoksa, odds ratio olarak almaya mecbursanız o zaman diğerini yapmak.

Önder Ergönül: Burada ham veri yok aslında. Bu sadece makalelerden yeni elde ettiğimiz veri.

Bir Katılımcı: Ham veri derken onu demek istiyorum odds ratio değil de alan almayan, effect olan olmayanın sayısı varsa eğer çalışmaları böyle pool edip yapmak daha doğru değil mi?

Önder Ergönül: O şekilde pool edersek sonuç farklı çıkıyor.

Bir Katılımcı: Farklı çıkıyor ama tercih edeceğimiz sonuç bu değil mi?

Önder Ergönül: Yok, değil aslında.

Metaanalizde bunları kombine etmektense effect size'ları kendi içinde bir hesaplama, birden toplayıp bölerek değil de kombinasyonu tercih etmemiz gereken yol. Bir takım kendi içlerinde olan, varsa eğer, confounderları görmek, odds oranlarını tek tek görmek önemli. Ve o farklılığın nereden çıktığını, hangisinin az hangisinin çok olduğunu göstermek bu işin gösteriş ya da sunum

biçimi, ama sunum biçiminin ötesinde odds oranı da daha farklı çıkmakta. Yoksa metaanaliz denen yöneme gerek kalmazdı. Bilgisayarda da yaparsınız, nasıl yaparsanız yapın ama bunları bu şekilde toplayıp yaparsak belki tesadüfen aynı sonuçlara ulaşabilir ama çoğu kez farklı sonuçlara gidebiliyor o yüzden paradoks diye kabul edilebiliyor. Öbür yöntemi tercih etmek zorundayız. İşte şurada veri havuzlandığında sonuçlar aynı olmaz gibi bir uyarı yapmış olalım.

Şimdi bir de başka bir tabir kümülatif metaanaliz var. Kümülatif metaanaliz bütün hepsinin tek tek toplanarak ilerlemesi, tarihsel açıdan bunu yapabilirsiniz ama bu daha çok veriyi sunmak için, veriyi göstermek için ama analiz amaçlı kullanılmayan bir başka kavram. Yine bir başka kavram, meta-regresyon. Meta-regresyonda belirli faktörlerin sizin için önemli görebileceğiniz değişkenlerin tıpkı lojistik regresyon veya linear regresyon, yani çok değişkenli analiz gibi kullanılması. Mesela yine bu örnekte görüldüğü gibi akciğer kanseri ile ilgili bir metaanaliz, işte kadınla, non-smokerlar belli bir mutasyon. Bu faktörleri seçerek bir tür meta regresyon yapılmış. Meta regresyon, metaanalizden tamamen farklı. Kümülatif metaanaliz de ondan farklı kavram olarak. Daha sonra belki deşinebiliriz.

Metaanalizde sıkça sorulan sorular nelerdir? Bu çalışmalar acaba birleştirmek için uygun mudur? Bu sorulabilir, ana rakamlar olduktan sonra mümkün.

Farklı dizaynlar bileştirilebilir mi? Aslında bunlar birleştirilebilir nihayetinde.

Metaanaliz yapmak için kaç çalışmaya ihtiyaç var? Yazarların ortaklaştığı ikiden fazla çalışma varsa tabii ki yapılabilir hatta en az iki olması söylenmekte.

Eleştiriler ve karşı görüşler de vardır hep metaanalizde. Eleştiriler varken karşı görüşler de vardır. Mesela bütün alanda yapılacak metaanalizde bir tek rakamla her şeyi özetleyebilir miyiz? Hayır, özetlemiyoruz, metaanaliz her şeyin sonu değil. Belki birçok şeyin başlangıcı bile olabilir.

Publication bias olmuyor mu? Her konuda vardır zaten, klasik derleme de yapsanız olabilir. Burada biz elmalarla armutları topluyor muyuz? Bu her zaman geçerli olan şeyler, tamamen araştırmacılara bağlı. Garbage in ve garbage out şu, değerli ve değersiz şeyi aynı yere koymuyor muyuz burada? Doğru, olabilir ama güçlü çalışmalar her zaman etkiler aslında. Güçlü çalışmanın yanına zayıfı koyarsak o güçlü çalışmaya biraz haksızlık olmuyor mu, gibi görüş vardır ama onun ağırlığı zaten olacaktır. Randomize kontrollü çalışmanın yerine mi geçiyor? Geçmez. Çok iyi yapılmış randomize kontrollü çalışmalar varsa eğer, tasarımları güçlüye iyi yapılmış bir metaanaliz onlardan daha iyidir bence. Bu tamamen değerlendirmeye bakar. Metaanalizler son dönemde çok kötü yapılıyor, evet kötü yapılıyor ama yapılacak bir şey yok, bunu yapanlar insanoğlu sonuçta. Uzmanlar. Klasik değerlendirmeyi yapanlar da uzmanlar.

Kabaca farklarını özetlersek burada öne çıkan şu oluyor: Metaanalizin klasik derlemeye göre avantaj kazandığı nokta; bir kere sayıyı devreye sokuyorsunuz, sayılar işin içine giriyor. Yani her bir çalışmayı sayısal anlamda gözden geçirmiş oluyorsunuz. O sayısal gözden geçirme ile, klasik derlemedense işin içine daha çok girmiş oluyorsunuz. Yani derleme makale yazacak olsanız, çoğu kez belki detaylı bile okumazsınız birçok makaleyi. Özetinden bile yola çıkarak conclusionları ile idare ediyorken bu kez metaanalizi yazıyorken en azından sonuç, bulgular tablolarıyla haşır neşir olmak zorundasınız ve bunu iyi şekilde sunmak zorundasınız. Her ikisinde de uzman görüşü çok önemli. Ve söylenen çok doğru bir söz, 'metaanalizi kesinlikle abartmamak lazım,' metaanalizi yapan kişinin aynı zamanda süreci de çok iyi bilmesi gerekiyor. Yani klinik süreci bir hastalık üzerinde yapıyorsanız, klinik sürecin çok iyi bilinmesi sayesinde ancak doğru metaanaliz yapılabilir. Felaket metaanalizler de var bu arada. Bir Cochrane örneği var. Mesela ben buna şok oldum. Bir tek yapılan bir metaanaliz. Şaka değil açıp bakabilirsiniz. Şöyle ki ilgimi nereden çekti. Benim ilgilendiğim kendi alanım. HIV infekte kişinin eline iğne batarsa profilaksi verelim mi vermeyelim mi? Klasik olarak veriyoruz. Yani HIV'li hastamız yatıyor, diyelim hemşirenin eline iğne battı, ne yapacağız? Etkili olduğunu biliyor

muyuz? Tam olarak bilmiyoruz ama profilaksi bir ay boyunca bir sürü yan etkisi olan, iki üç ilacı veriyoruz. Vermek durumundayız şu anda. 1996 yılında bir tek çalışma yapılmış olduğunu biliyorum ben bu konuda. O da 'historical control' kullanan bir vaka kontrol çalışması yapılamadı. Etik olarak doğru değil, yapılamadı. Bunu görünce çok şaşırdım. Metaanaliz yapılmış, ne güzel diye okumaya başladığım zaman şunu gördüm, sadece o çalışmayı kullanıyor. Evet, tamam başka çalışma yok, yine o çalışmaya atfen anlatıyor, yorum yapıyor. Tıpkı buradan yola çıkarak. Aslında bir derleme sonuçta ama Cochrane veri tabanında yer almış bir çalışma. Yani metaanaliz okurken dikkatli olmalıyız, çok analitik bir şekilde değerlendirmek zorundayız. Cochrane iyi bir marka ama marka demek her zaman iyi bir şey demek değil. Metaanaliz her zaman son sözü söylemez. Mesela 2006 yılında önemli bir tartışma olmuştu, bir debate yaşandı. İsrail'den bir grup, Yunanistan'dan bir grup sürekli fabrikasyon tarzında metaanaliz üreten ekipler vardır. Onlardan bir ekip bir çalışma yapmıştı fakat dedi ki biz metaanalizi hiç önemsemeyiz, biz kendi hastanelerimize daha çok güveniriz. Bu da bir yöntem, bunu söyleyen kişi net olarak güvenilir buluyor, metaanaliz dert değil, dedi. Bizim son olarak Kırım Kongo'da iki sene evvel bir bias üzerinde tartışmıştık. Tartışma sürüyor fakat hoş tarafı şu, yöntemler geliştikçe daha bir üst düzeyden sürüyor. Kırım Kongo'da ribavirin konusunda üç tane metaanaliz yapıldı. İki sene evvel burada konuştuğumuz zaman, o biasler yeni çıkmıştı, ben burada söz ettim ve sonra bir editoryal yazdım. O editoryal yazıya dergi de geniş yer verdi, hiç kısaltmadan yayımladılar, 2-3 sayfalık. Ama karşı görüşe de yer verdi, debate olarak duruyor şu an literatürde. Ama editör o anlamda kendisi bir yorum yapmadı. Eksiktir bence, yanıtlar hiçbir şekilde karşılanmadı. Çok hakaretler de edildi, kişisel laflar söylendi ama editör hiç dokunmadı. Tabii bizim de bakacağımız başka bir ton iş olunca çok da üzerine gitmedik orada öyle duruyor ama bu tartışma sürüyor.

Bu arada İngiltere'den bir metaanaliz yayımlandı. İngiltere'den yayımlanan metaanaliz Kırım Kongo'da ribavirinin kullanımı, bir tablosunda ilk başında biasleri tarif etmişler İngiliz arkadaşlar, yapan da bir firma aslında. Bir ghost writer ya ilk

kişi ama diğer insanları da tanıyoruz. Sonuçta bias'lı olan yayınları yeşille belirtmişler.

Aslında güzel bir sunum. Bir metaanaliz için fena değil ama onu ne kadar doğru tarif ettiler o ayrı. Yani bu bias'ler ne kadar bias ve sonuçta bunu yaptıkları zaman görüyorlar ki Kırım Kongo'da ribavirin etkili. Overall yanıtı baktıkları zaman ki bizim için en önemli şey erken kullanımı. Tıpkı romatoid artritteki erken tedavi gibi, ki viral hastalıkta bu çok önemlidir, erken kullanımı oldukça önemli. Ama erken kullanım diye ayırmamışlar, ayırmaksızın yaptıklarında bunu etkili bulmuşlar aslında. Buna rağmen bu alanda randomize kontrollü çalışma olabilir diyorlar. Mesela bu yanlış bir şey, çok düz mantıkla yapılan, buradan işte şu vurguyu yapmak istiyorum. O konunun metaanalizini yapanlar konuyu çok iyi bilmeli, mesela 'length of stay' yoğun bakım için önemli bir hastalık, kronik hastalıklar için önemlidir. Yani bir kişinin hastanede kalma süresi kronik hastalıklarda önemli ama burada önemli değildir, çünkü akut bir hastalık var, zaten ağır hasta erken ölüyor. Yani burada ağır hasta diyelim ki Yozgat'tan, Çorum'dan Ankara'ya geldi, zaten o 1-2 günde ölüyor. Length of stay'i çok küçük. Length of stay'i yüksek olan kim, iyi hasta. İyileşen hastanın length of stay'i iyi oluyor. Dolayısıyla length of stay'i burada bir prediktif parametre ya da gösterge, outcome parametresi olarak alamazsınız. İngiliz arkadaşlar bu hataya düşmüşler bence. Bu yanlış olmuş o anlamda.

Servet Akar: O tabloda iki şey anlamıyorum. Birisi fixed effect birisi random mudur, yoksa başka bir şey mi ifade ediyor onlar?

Önder Ergönül: Random effect model kullanmış. Farklı şeylere göre, length of stay hepsi vermemiş, onu metin içinde açıklıyor, bazıları için birini kullanmış, diğerleri için öbürünü kullanmış, yani orada bir detayı kullanıyor.

Servet Akar: Bir öncekinde de vardı...

Önder Ergönül: Şu iki tane sistem burada RCT ve nonRCT için kullanıyor.

Servet Akar: Yukarıda RCT bir tane var, ondan yapmak doğru mudur metaanalizi? Ondan bir diamond elde etmiş.

Önder Ergönül: RCT'den yapmış, sonra buradan yapmış. Aslında buradaki RCT de yanlış. RCT denemeyecek bir çalışma. Ne kadar yanlış yönlenildiğine örnek olarak getirdim ben bunu. Şöyle ki, RCT denen bir çalışma, etik hocaları bu arada yapılamaz, diye rapor verdiler, Berna, Arda, Murat Civaner, leading medical ethicists dediler ki Türkiye'de randomize kontrollü çalışma yapılamaz, bunları yazılı olarak sundular Sağlık Bakanlığına. Fakat bu arada Karadeniz'den bir hocamız, "ama biz yaptık" dedi. E nasıl yaptınız, etik kurul nasıl aldı? Çünkü etik kurul dedikleri orada, fakültede. Karadeniz'in içinde almışlar bir şekilde. Ama zaten tasarım hatalı. Tasarım şu, ortalama kalış süresi, bu çalışma için söylüyorum, yüksek hastalar geç geliyor, bizim meselemiz orada zaten. Yani tedavinin erken ya da geç verilmesi çok önemli. Geç gelen hastalara bakıyorlar, geç gelen hastalarda zaten ribavarinin etkili olduğunu iddia eden yok, bir haftayı geçtikten sonra ki ortalama kalış süresi 5 günü geçiyor orada ve hiçbir şey de yok ayıriyeten, ikna edici değil. Ben kesinlikle ikna olmuyorum, çünkü şöyle ki burada çok bariz bir şekilde 'confounding by indication' var. İlk başta hastalık çıktığında ağır kimselere daha çok ilaç verildi. Trabzon bunu nasıl yapmadı, onu bilmiyorum. 2004 yılında yaptık deniyor. Bunun ötesinde de bir şey demiyor, yani nasıl yaptınız? Ama Türkiye'den bu arada çıkan bir başka tartışmalı metaanaliz, bu Türkiye'den ikinci metaanaliz, geçen haftalarda yayımlandı. Bu da aynı ekipten ama Trabzon'u eleştiriyor zaten. Hasta dağılımı uygun değildir, sorun vardır aslında diye eleştiriyor. Yine randomize kontrollü çalışma, fakat buradaki sıkıntı da ne? Yeni bir çalışma, bu arkadaşlar hepsi randomize kontrollü çalışmanın yapılması gerektiğini savunan kişiler. Türkiye'de mutlaka yapılmalı, diyorlar. Ama çok ciddi bir sıkıntı var, İran'ı kattıkları zaman bu kez yön ters. İran çok net bir şekilde ribavirin lehine buluyor. Ama onları ayrı tutup Türkiye çalışmaları katılıyor. Türkiye çalışmalarında da dâhil edilen çalışmalarda sıkıntı var, ben kendi çalışmam olduğu için biliyorum.

Mesela en önemli hata, biz seri olarak bir sürü yayın yaptık, en çok katıldığımız vakalar alınmamış ama geçmişten başka bir şey alınmış. Burada niyet farklı tabii ki, oysa bu temelde çalışmaları seçerken yine erken kullanım diye ayırmak mutlak şart. Bizim yaptığımız çalışmada, bütün hepsine baktığımızda zaten gene anlamlı, p-değeri de çok anlamlı ve odds oranı da 0,475 yani ribavirin alanlar almayanlara göre iki kat daha az fataliteye sahipler. Erken kullanım yaptığımız zaman ise 5 kat daha az fataliteye sahipler. Yani erken kullanım oldukça önemlidir, diye söylüyoruz. Şimdi böylesi bir durum özetle şu, metaanaliz yapınca sorunlar çözülüyor. Biz de bunu yayımlayacağız tabii.

Servet Akar: Hatırladığım kadarıyla RevMan sadece RCT'ler için kullanılıyor değil mi?

Önder Ergönül: Hepsini kullanabilirsiniz, fark etmez.

Servet Akar: Yok kantitatif olarak mesela, exposure nonexposure yerine parametrik çalışmamak için.

Önder Ergönül: Olabilir. Hepsinde yapabilirsiniz.

Servet Akar: Çünkü ben RevMan'ı indirdim, sizin çalışmalar için belki yapılabilir ama yanıt verenler vermeyenler, veya yanıt tahmini için?

Önder Ergönül: Belli bir menü seçeneği var. Proportion mı, ortalama mı, bunları seçiyorsunuz zaten. O da herhalde onun gibi bir şeydir, bilemiyorum. Ama STATA da kullandığınız zaman, sürekli bir şeyler ekledikleri bir program. Oldukça iyi sayılabilir.

Bir Katılımcı: NCASES de aslında çok iyi bir meta-analiz programı, onu da öneririm.

Hasan Yazıcı: Çok teşekkürler. Aynı konuda bir çok metaanaliz yapıldığı zaman, başka şeyler çıkabiliyor, tamam, şimdi bunun tersini alalım, software'ini de veriyorsun, metaanalize hangi verilerin girileceğini de veriyorsun, şu makaleleri alacaksın diyorsun, sen seçiyorsun, tamam mı? Software'ini de beraberinde iki ayrı gruba veriyor-

sun. O iki ayrı grubun arasında o metaanalizde ne farklılık çıkıyor ortaya, değişkenlik ne var? Bunun için çalışmalar var mı? Ya da softwareler arasında değişiklik olduğu zaman çıkan sonuçta değişiklik var mı?

Önder Ergönül: Hiç öyle bir şeye rastlamadım.

Hasan Yazıcı: Böyle bir çılgın proje yapılmamış mı?

Önder Ergönül: Aslında güzel fikir, şöyle ki bunları kattığımız zaman acaba hepsinde denediğimizde nasıl bir sonuç çıkıyor? Ben şimdi bu NCASES, RevMan madem indiriliyormuş, bunları denerim sıcaklığına.

Hasan Yazıcı: Nathan'a teklifim şu olacak. Ben 'federal authority' olsam derim ki A firmasının verilerini, çıkmadan evvel B firması analiz etsin. Rakip firma B'yi de C analiz etsin.

Önder Ergönül: Burada temel sıkıntı bakış açısından kaynaklanıyor.

Bir Katılımcı: Tamamen farklı çıkabilir. Bir, makale seçme aşamasında Vedat Hocayla TNF'lerin sistematik reviews'ini yaparken 100 küsur makalede benim bulduğum onun bulmadığı, onun bulup benim bulmadığım, o aşamadan da olabilir. İkincisi ve daha önemlisi outcome measurementları seçmekte. O Cochrane deneyimimizden de gördük, o konuda çalışmayan insanlar tamamen farklı outcome parametreleri seçme eğiliminde oluyor.

Bir de o ilk sözünü ettiğim İngiliz makaledeki son isim Cochrane'ın adamı. Onu da söyleyim. Cochrane sayılır yani ama yayımlandığı dergi çok parlak bir dergi değil.

Bir Katılımcı: Cochrane çok mekanik yaklaşıyor bu işe, biraz da duygu katmak lazım herhalde.

Önder Ergönül: Uzmanlık katmak diyelim.

Bir Katılımcı: Ben istatistikçi olmadığım için metotla ilgili bir şey söylemeyeceğim ama klinisyen-

lerin metaanalizlerle kandırıldığını düşünüyorum. Çünkü klinisyenlerde şöyle bir inanç var, meta-analiz datası varsa klinisyen bu artık tartışılmaz bir etki gibi düşünüyor. Hâlbuki o kadar minimal, ihmal edilebilir, çok özel hasta gruplarına ait yazarlar, metaanalizler, defalarca aynı çalışmalar, iyi gelen çalışmalar metaanalize konuluyor. Şimdi klinisyeni gördüğünüzde bir kanıt var, diyor. Alzheimer'lıya Aricept gibi ilaçlarla ilgili, klinikte faydası son derece tartışmalı, hangi hasta grubunda nasıl bakıldı, neye bakıldı, çok minimal. NNT'si 100ler, 200lerin üstünde, osteoporozda da aynı durum var. Etki son derece marjinal ama klinisyene metaanaliz datası sunulup bir A bir B, bir sürü çalışmanın metaanalizi var denildiğinde bu klinisyenleri, özellikle gençleri çok yanıltıyor. Bir kanıt var, bu tartışılmaz. Bir de bu guidelinelara da çok giriyor, onun için metaanalizleri bazıları kötü amaçlı kullanıyor gibi geliyor bana. Bilmiyorum siz bu konuda ne düşünüyorsunuz?

Ayhan Dinç: Ben de genel prensip üstüne bir şey söylemek istiyorum. Şirketlerin evrimiyle ilgili bir hikaye vardır, 10 kişi çalışıyor, sonra şirketler büyüdükçe yönetim kademesi büyüye büyüye bir bakıyorsunuz sonunda yönetim kademesinde 90 kişi var yine altta 10 kişi üretiyor. Şimdi bu aynı, çalışanın sayısı artmıyor ama yorumlar sürekli artıyor. Burada bir etik sınır var mı? Yani bir dergiye bir metaanaliz gönderildiği zaman orada tutum nasıl oluyor? Çoğu zaman metaanaliz ve diğer reviewler için de geçerli, çok eleştirecek bir şey bulamıyorsunuz, yayımlayıp yayımlamaması tamamen derginin insiyatifi. Yani çok çalışma üretilmeyip de sürekli birilerinin bu konuda yorum yapması nasıl karşılanıyor bu metaanaliz camiasında?

Önder Ergönül: Metaanalizin tabii kendisinin de orijinal olması gerekir. Biz şimdi 3. metaanalizi göndereceğiz, göndermek durumundayız. Ne diyeceğiz, böyle bir farklılığı var demek durumundayız. Metaanalizin bir orijinalitesinin olması gerekli. Yoksa saçma olur.

Banu Çakır: Önder teşekkür ederim çok güzeldi. Sözel olarak birkaç katkı yapmak istiyorum, bazı bölümler hızlı geçti. Önder vurguladı ama tekrar vurgulamak istediğim birkaç konu var. Bir, ben

hep metaanaliz anlatırken diyorum ki bir elbise yapıldığı kumaş kadar kaliteli olabilir. Siz onun üstüne ne kadar güzel bir format yaparsanız yapın, kumaş bozuksa bir yerden patlayacaktır, yani metaanalizlerde en önemli özellik tabii ki malın kalitesi. Yani hiçbir metaanaliz ya da hiçbir uzman yok ki, ancak istatistikçiler, epidemiyologlar bu aşamada dua edebiliyorlar, veri kötüyse sizin yapabileceğiniz bir şey yok. Onun için metaanalize giren materyalin iyi olması, biz bu materyali kaliteli yaptığımız zaman geliştirmek açısından önemli. İkinci, metaanaliz hep yıllarca bize aslında örnek hacmini büyütme için yapılacak bir şey gibi takdim edildi. Biz de bu şekilde kullandık. Bu önemli. Örnek hacmimiz küçük olduğu zaman ya da cevap verilmediği zaman, biz belirli bir sonucu bulamıyoruz. Hatta daha kötüsü, örnek hacmi küçük çalışmalarda bir ilişki bulunduğu zaman bu ilişkinin gerçek olma olasılığındansa yanlış olma olasılığı daha yüksek. Çünkü bizim her zaman p hata payımız var, %5. Dolayısıyla 100 kişi sağlamsa siz bu 100 kişinin içinde çalışma yaptığınızda zaten beşini hasta buluyorsunuz, yani çalışmada örnek büyüklüğü yeterli değil ama ilişki buldum, demek de hoş değil. Ben metaanalizi özellikle klinisyenlerle paylaşırken daha başka açıdan görmeye çalışıyorum. Metaanalizin güzel yanı, örnek hacmini genişletmek ve tek bir değer vermek, hani Önder'in Kırım Kongo'sunda olduğu gibi, herkes farklı bir değer veriyor. Bu değeri vermenin ortalaması mümkünse, daha önemlisi belki bunun sınırlarını vermek, alt ve üst sınırını vermek. Ama bana göre metaanalizin daha önemli bir kısmı, bütün yayınları okuyup, bu yayınlar arasındaki heterojeniteyi fark etmek. Şimdi random model formülü geliştikten sonra bu da sıkıntıya düştü. Fixed effect modeli anlaşılır, modeller benzer, dolayısıyla ortalama bir değer veriyorsunuz. Ama modeller heterojen olduğunda, o zaman etkinin gücü, kimi zaman artı kimi zaman eksi, gelin biz bunun ortalamasını alalım demek çok doğru değil. Aslında metaanalizde ben birkaç örnekte düşündüm, Kırım Kongo'da da düşündüm. Bu farklılığı açıklamak daha önemli. Yani hangi çalışmada böyle tuhaf bir farklılık var, acaba popülasyonun altında başka bir denge mi var, başka

bilmediğimiz bir karıştırıcı faktör de mi var? Simpson's paradox bizim sık karşılaştığımız bir şey. Küçük çalışmalardan elde edilen çalışmalar bir araya gelince karıştırıcı faktör vs. de olmadığı zaman yanlış yönde bir bilgi veriyor ama ben de kendi adıma çalışma yaparken ham veriyi kullanmayı seviyorum. Çünkü ham verinin şöyle bir avantajı var. Ham veride sizin gördüğünüz karıştırıcı faktörler de var, yani çalışma o karıştırıcı faktör için hiç kontrol etmemiş olabilir ya da bir çalışma etmiş, diğeri etmemiş olabilir. Ama siz ham veriyi alırsanız ve doğru dürüst metaanaliz yapıyorsanız, o karıştırıcı faktör için kontrol edip daha doğru bir sonuç verebilirsiniz ama bugün aklınızda kalmasını istediğim şeylerden bir tanesi metaanaliz sadece bir değer elde etmek ya da örnek büyüklüğünü geliştirmek için değil, aynı zamanda heterojenliği görmek için. Eğer romatoid artrit tedavisinde aynı ilaç Avrupa toplumuyla Asya toplumunda farklı bir etki gösteriyorsa, belki bu Avrupa ve Asya toplumundaki genetik alt yapı farkı olabilir, subgrup analizleri yapılabilir. Bir bunu vurgulamak istedim bir de şu çok önemli, biz forest plotları veriyoruz da funnel grafikleri hiç vermiyoruz. Aslında o Önder'in gösterdiği funnel grafikte onun külâh görünümünde olması çok önemli. Niye önemli, çünkü altında bir boşluk var, o boşluk küçük çalışmalar. Dolayısıyla o küçük çalışmalar yok yani yayımlanmamış, biz onların sonuçlarını bilmiyoruz. İki şey yapabiliyoruz, birincisi çok kolay olarak hepimizin yapabileceği bakıp, asimetrikse belli ki burada çalışmalar yayımlanmamış, bu bir yöntem. İkincisi bazı yöntemler var, matematiksel olarak da oraya, onları da varmış gibi, onları da koysak acaba ilişki hangi yöne gider, diye. Bu tür şeylere bakmak için de metaanalizleri iyi değerlendirmek lazım. Yani burada yine okuyucu okuyup, yazana bu yanıtı verebilmeli. Yani siz bize bunu verdiniz ama bakın, funnel grafiğinizde birçok yayın paylaşılmamış. Dolayısıyla bu verdiğiniz sonuç doğru değil, ya da burada bir random effect modeli kullanılmış ama o kadar ciddi bir gruplar arası fark var ki bunu belki kullanmamanız lazım, belki bunu ayrı ayrı rapor etmeniz lazım diye biz okuyucu olarak da buna geri dönüş yapabilmeliyiz. Metaanalizleri o açıdan ben iyi buluyorum, her

şeyi okuyup bir yana koyuyorlar siz bunu değerlendirebiliyorsunuz, bu anlamda bakarsak daha değerli olabileceğini düşünüyorum.

Servet Akar: Peki hocam heterojeniteyi gördünüz, siz neyle değerlendiriyorsunuz? e^2 ile mi x^2 ile mi, bazı programlar heterojeniteyi anlayabilmek için bir takım istatistik testler tanımlıyor. Heterojeniteyi gördünüz, buna rağmen analizde mi tutuyorsunuz yoksa onları çıkarıp mı analizi tekrar ediyorsunuz. Hangisini öneriyorsunuz?

Banu Çakır: Farklı yöntemler var gerçekten ama her ikisini de yapabilirsiniz, belirli çalışmalarda şöyle yöntemler de olabilir mesela, tıp da ilerliyor, teknikler değişiyor vs. Eğer subgrup analizleri yapabilirsiniz ayrı ayrı değerlendirirsiniz. Heterojenlikteki en kolay yöntem ayrı değerlendirmektir. Her anlamda en kolay. Eğer kadında ve erkekte farklıysa ve hiçbir şey bilmiyorsanız, kadın ve erkek için iki şey yapmanız en kolay açıklamadır. Çünkü heterojenliği tarif etmemiz lazım. Karıştırıcı faktör de kontrol ediyorum.

Heterojenlikte tanımlamamız lazım, çünkü heterojenlik bir zenginlik aslında, gösterilmesi gereken bir şey. En kolayı subgrup analizleri ayrı ayrı verebilirsiniz. İkinci bir şey, onlu, onsunuz verebilirsiniz ya da random model kullanıp açıklamasını yapabilirsiniz.

Emmanuel Lasaffre: Doğru mu anladım emin değilim ama soru; fixed mi random effect meta-analiz mi idi, değil mi? Ben her zaman random effect metaanaliz kullanırım. Sebebi şu, heterojenite için herhangi bir testin gücü çok zayıf ve ben homojeniteye inanmıyorum çünkü her çalışmanın protokolü farklı ve farklı coğrafi alanların kapsadıklarını ya da sahip olmadıklarını göz önünde bulunduruyor. Heterojeniteye rıza göstermeli ve çalışmanın söylediklerini dinlemeliyiz, bu mümkün olursa tabi.

Önder Ergönül: Şöyle söyleyelim, alt grup analizi oldukça önemli, o da hipotezle ilişkili. Bizim örnekte ribavirini erken kullanmak bizim için çok önemli, hipotez, bu da alt gruptan ortaya çıkan bir şey. Eğer öyle kullanmış olan başkaları varsa ama

İran'daki çalışmayı nereden bileceğim. Bilemeyeceğim ama öyle yazılmış, onu kabul ederim ya da etmem. Bence de sonuçta farklı heterojeniteleri matematiksel testlerle anlamak işi çok kompleksleştirebilir.

Banu Çakır: Lütfen açıklamama izin verin. Random effect modeli sevmediğimden değil, belirtmek istediğim eğer saf heterojenite görebiliyorsanız, bazı alt gruplar, alt popülasyonlar, aykırı gruplar, random effect modeli kontrol etti diyen tek bir karar uygun olmayacaktır. Bu durumda alt grup prezantasyonları yapmak tek bir karar vermekten

çok daha iyi olacaktır. Öte yandan fixed, hiçbir şey yokken fixed effect modeli kullanmanın hırs yapmak olduğu konusunda size katılıyorum. Ben random effect modeli kullanma kısmına katılıyorum ama tek tek çalışmalarda yüksek heterojenite varsa alt grup analizi de yapmak isterim, sadece heterojenitenin başka bir kaynaktan gelmiş olup olmadığını görmek için.

Emmanuel Lasaffre: Yani diyorsun ki, eğer yüksek heterojenite varsa önce random effect metaanaliz yapalım, bunu da meta regresyon gibi bir yöntemle açıklamak istiyorsun.

MALİYET ANALİZLERİ

Necdet Süt

1. Giriş

Sağlık teknolojilerinin hızlı gelişimi sonucu çeşitli hastalıkların tedavisinde kullanılan ilaçlar, tedavi yöntem ve araçları çok pahalı elde edilebilir bir noktaya ulaşırken sağlığa ayrılan kaynaklar sınırlı kalmıştır. Oluşan bu yapı, sağlık sisteminde hastalara yönelik sağlık hizmet ve ürünlerinde tercih yapılması durumunu yaratmıştır. Temel ilke olan ‘sağlığı iyileştirici her tedavi maliyetine bakılmaksızın verilmelidir’ ilkesi yerine çeşitli hasta popülasyonları için en uygun (optimal) tedavinin hangisi olduğunun dikkatlice incelenmesinin daha yararlı olduğu kanısı edinilmiştir. Sağlık harcamalarının hızla büyümesi, tedavi yöntemlerinin ve programlarının özellikle de sağlık ürünlerinin ekonomik değerlendirmelerinin yapılmasını gerekli kılmıştır. Maliyet analizi yöntemleri, uygulanan tedavi yöntem/programlarının etkinliklerinin ve faydalarının değerlendirilmesinde karar vericilere yol gösterici rol oynamaktadır. Bir başka ifade ile sağlık ekonomisinde kullanılan maliyet analizi yöntemleri eldeki sınırlı ekonomik kaynakların hasta/sağlık sistemi açısından rasyonel karar vermede ve maliyet-sonuç dengelemesinde nasıl kullanılması gerektiği konusunda yol gösterirler.

Maliyet analizleriyle sağlık sisteminde tedavi yöntemlerinin maliyetlerinin ölçümü ve bu maliyetlerle sağlanan sonuçlar değerlendirilmektedir. Temel amaç, tedavi yöntemlerinin maliyet ve sonuçlarının tanımlanması, ölçümü ve karşılaştırılmasıdır. Ayrıca bu analiz yöntemleri yardımıyla her hastanın kendisi için en uygun tedaviyi almasının sağlanacağı bir düzende sağlık kaynaklarının rasyonel tahsis edilmesi için yol gösterilmektedir. Maliyetler uygulanan tedavi yöntemlerinin girdi kaynakları, sonuçlar (çıktılar) ise bu tedavi yöntemleriyle elde edilen sağlık durumlarıdır.

İyi bir maliyet analizi değerlendirmesinde, incelenen sağlık sorununa ilişkin tüm maliyetler kapsanmalıdır. Maliyetler ‘doğrudan’, ‘dolaylı’ ve ‘görülme – intangible’ olmak üzere üç kategoria sınıflanabilir.

- Doğrudan maliyetler, hastanede tedavi görenin getirdiği maliyetler olarak ta-

nımlanmakta olup kendi içerisinde ‘Tıbbi maliyetler’ ve ‘Tıbbi olmayan maliyetler’ olmak üzere ikiye ayrılmaktadır. Doğrudan tıbbi maliyetlere örnek, laboratuvar testleri, ilaçlar ve hastanede yatma giderleri verilebilirken, doğrudan tıbbi olmayan maliyetlere örnek, hasta nakil giderleri, hastanenin elektrik su gibi giderleri ve personel giderleri verilebilir.

- **Dolaylı maliyetler**, uygulanan tedavi yönteminin ya da hastalığın kişinin üretim kapasitesinde meydana getirdiği değişim olarak tanımlanabilir. İş günü kaybı, üretkenlikte azalma, sağlık durumunda değişme, yaşam kalitesi kaybı dolaylı maliyetlere örnektir.
- **Görülmeyen (Intangible) maliyetler**, uygulanan tedavi yöntemi ya da hastalık nedeniyle oluşan ağrı ve acının getirdiği maliyeti ifade etmekte olup ölçülmesi çok zordur.

Bir sağlık sorununa yönelik uygulanan tedavi yönteminin sonuçları ise, ‘*linik*’, ‘*insancıl*’ ve ‘*ekonomik*’ sonuçlar olmak üzere üç farklı kategoride sınıflandırılmaktadır.

- **Ekonomik sonuç**, tedavi yönteminin hastaya, sağlık kuruluşuna, sosyal güvenlik kuruluşlarına ve topluma getirdiği mali yük olarak tanımlanmakta olup parasal değerlerle (TL, USD, Euro vb.) ifade edilmektedir.
- **Klinik sonuç**, tedavi yönteminin bireyin sağlık durumunda meydana getirdiği kazanımları gösterir.
- **İnsancıl sonuç**, tedavi yönteminin kazanımlarının hastanın fonksiyonel durumunda ve yaşam kalitesinde meydana getirdiği değişimlerdir. Örneğin, yaşam kalitesi, hasta memnuniyeti, hasta tercihlerindeki değişim gibi.

Klinik araştırmalarda etki (efficacy), araştırma çıktılarının değerlendirilmesinde etkinlik (effectiveness) ve maliyet analizi araştırmalarda verimlilik (efficiency) değerleri esas alınmaktadır. Bu kavramlar aşağıda tanımlanmaktadır.

- **Etki (Efficacy)**: İdeal koşullarda (Faz III klinik deneylerde) bir ilacın ne kadar iyi oldu-

ğunu gösteren ölçüttür. Belli bir toplumdaki bireylerin bir sağlık sorunu için uygulanan tıbbi teknolojiye ideal kullanım koşullarında yararlanma olasılığı olarak da ifade edilebilir.

- **Etkinlik (Effectiveness)**: Normal klinik uygulamalarda (gerçek yaşamda) bir ilacın ne kadar iyi olduğunu gösteren ölçüttür.
- **Verimlilik (Efficiency)**: Normal klinik uygulamalarda (gerçek yaşamda) bir ilacın klinik etkinliği kadar maliyet etkinliğinin de iyi olmasıdır.

2. Maliyet Analizi Yöntemleri

Maliyet analizi, herhangi bir sağlık sistemin işleyişine yönelik kaynak ayırma işleminde analitik yöntemlerin kullanılması olarak tanımlanabilir. Bu yöntemlerin temel amacı tedavi yönteminin maliyetini belirlemek ya da alternatif tedavi yöntemlerinin sonuçlarını ve maliyetlerini karşılaştırmaktır (yeni tedavi yönteminin standart tedavi yöntemlerine göre maliyet etkinlik/fayda/kar açısından değerlendirmelerinin yapılması gibi).

Yapılacak olan herhangi bir maliyet analizde bulunması gereken ana elemanlar aşağıda sıralanmaktadır.

a) Analiz yönteminin tanımlanması: Çalışmada hangi maliyet analizi yönteminin (Maliyet Analizi, Maliyeti En Aza İndirgeme Analizi, Maliyet-Etkinlik Analizi, Maliyet-Fayda Analizi, Maliyet-Kar Analizi gibi) kullanıldığının açıklanması gereklidir. Seçilecek yöntemin çalışılacak konuya uygun olması gereklidir. Eğer alternatif iki tedavi yöntemi birbiriyle örtüşen yarar/etkinlik ve güvenilirlik profiline sahiplerse bu durumda en uygun değerlendirme ‘Maliyeti En Aza İndirgeme Analizi’ ile yapılır. Tedavi yöntemleri sağladıkları yaşam kalitesi ile değerlendirilecekse ‘Maliyet-Fayda Analizi’ ile değerlendirme ‘Maliyet-Etkinlik Analizinden’ daha doğru olur.

b) Bakış açısının (perspective) belirlenmesi: İlgilenilen soruna uygun çözüm bulabilmek için öncelikle hangi maliyet analizinin uygulanması

gerektiği belirlendikten sonra analizin bakış açısına karar verilmelidir. Bakış açısı temel alınarak doğru maliyetler ve sonuçlar belirlenebilir. En yaygın kullanılan bakış açısı toplumsal bakış açısıdır. Ayrıca bu bakış açısının dışında hasta, sosyal güvenlik kurumları, sigortacılar, ilaç endüstrisi ve işverenlerin bakış açılarına göre de maliyetler hesaplanabilmektedir.

c) Maliyet ve sonuçların ölçülmesi: İncelenen çalışma konusuyla ilgili olan tüm maliyetler değerlendirilmede yer almalıdır. Alternatif iki tedavi yöntemi söz konusuysa bu yöntemlerin maliyetleri farklı olabilir (yan etkilerin maliyeti, hastane maliyeti vb.), yapılacak değerlendirmede tüm bu maliyetlerin kapsanması gerekir.

- Maliyet analizlerinde tedavi yöntemleriyle ilgili olan tüm maliyetler (doğrudan, dolaylı) kapsanmalıdır.
- Maliyetlerin belirlenme yöntemleri (Friction yaklaşımı, Human Capital yaklaşımı) açıklanmalıdır.
- Maliyet analizlerinde tedavi yöntemleriyle ilgili sonuçlar (Klinik sonuç ölçümler, QALY ya da utilities gibi yaşam kalitesi skorları) belirlenmelidir.

d) Tedavi yönteminin seçimi: Sağlık sorununa yönelik klinik uygulamalarda uygulanabilir en iyi tedavi yöntemi seçilmelidir.

e) İndirgeme (Discounting) oranının seçimi:

Maliyet ve sonuçların indirgenmesi, şimdiki değerlerin geleceğe uyarlanması işlemidir. Tedavi yöntemlerinin maliyetleri ve faydaları farklı zamanlarda meydana geldikleri için 1 yılı aşkın sürenin söz konusu olduğu çalışmalarda para ve sağlık sonuçlarının uyarlanması gereklidir.

Maliyet ve sonuçlar farklı zamanlar için doğru uyarlanırlarsa, karar vericiler aynı zaman çizgisi üzerinde karşılaştırma yapabilirler. Bireylerin tercihleri, şimdiki kazanımları (ya da kayıpları) karşısında gelecekteki kazanımlarını (yada kayıplarını) yansıtmalıdır. Literatürde uygun indirgeme oranları hakkında kesin bir kanı olmamakla birlikte Maliyet-Etkinlikte önerilen indirgeme oranı

%3'tür. İndirgemeye örnek olarak, 1 yıl sonra 1000 TL olan maliyetin %3 indirgeme oranına göre şimdiki değeri 970 TL, 5 yıl sonra 1000 TL olan maliyetin %3 indirgeme oranına göre şimdiki değeri 860 TL olarak verilebilir.

f) Duyarlılık (Sensitivity) analizi: Bir sağlık programının ekonomik değerlendirmesi kapsamlı ve iyi yapılsa bile maliyet ve sonuç verileri çeşitli derecelerde belirsizlik ve bias içerebilir. Farklı kaynaklardan (klinik deneyler, deneyimler, gözlemsel çalışmalar vb.) çok sayıda bilginin bir araya getirildiği ve bilginin kesinlik düzeyinin kısıtlı olduğu maliyet analizlerinde alternatif tedavi yöntemlerinin doğru Maliyet-Etkinliği hakkında bazı kuşkular oluşur. Bu nedenle incelenen sorunun ana elemanlarına ilişkin kısıtlı bilgi verilerek yapılan tahminlere her zaman güvenilmez. Bu kuşkuyu gidermek için standart yöntem duyarlılık analizi uygulamaktır. Duyarlılık analizinde, maliyet etkinlik oranı hesaplanırken bazı kritik parametreler değiştirilir ve oran bu değerlere göre yeniden hesaplanır, elde edilen yeni oran bir parametredeki değişimin sonuçlara yansımaları ve duyarlılığı ne ölçüde değiştirdiğini gösterir. Duyarlılık analizi kendi içinde iki başlıkta ele alınabilir;

1. Tek değişkenli (Univariate) duyarlılık analizi:

Bir zaman çizgisinde tek bir değişkenle yapılan duyarlılık analizidir,

2. Çok değişkenli (Multivariate) duyarlılık analizi:

Bir zaman çizgisinde birden fazla değişkenle yapılan duyarlılık analizidir.

Maliyet Analizi Yöntemleri 5 başlık altında sınıflandırılmıştır.

- Maliyet Analizi (Cost analysis / Cost of illness analysis)
- Maliyeti En Aza İndirgeme Analizi (Cost-Minimization Analysis)
- Maliyet-Etkinlik Analizi (Cost-Effectiveness Analysis)
- Maliyet-Fayda Analizi (Cost-Utility Analysis)
- Maliyet-Kâr Analizi (Cost-Benefit Analysis)

Bu 5 farklı tipin temel özellikleri ayrıntılarıyla aşağıda açıklanmaktadır.

2.1. Maliyet Analizi (Cost analysis / Cost of illness analysis)

Maliyet analizi belirli bir hastalığın, tanımlanmış bir toplum üzerinde toplam maliyetinin tahmin edilmesi ve tanımlanmasına yönelik ekonomik değerlendirme yöntemidir. Hastalık maliyeti '*cost of illness*' olarak da adlandırılmaktadır.

Bu değerlendirmede belli bir hastalığa atfedilebilir doğrudan ve dolaylı maliyetler yer alır. Herhangi bir hastalığın getirdiği mali yükün tahmin edilmesinde kullanılan maliyet analizi alternatif tedavi yöntemlerinin karşılaştırılmasında kullanılan bir yöntem değildir. Bir ölçüde tüm ekonomik değerlendirme yöntemlerinde tedavi yöntemlerinin Maliyet analizleri de yapılmaktadır. Maliyet analizinde karar verilmesi zor olan iki durumla karşılaşılmaktadır, bunlar:

- Tedavide üst maliyet sınırının ne olacağı ve
- Farklı zaman kesitlerinde maliyetlerin değişimidir (indirgeme).

Maliyet analizleri doğrudan maliyetler ve dolaylı maliyetler olmak üzere iki bileşenden oluşmaktadır. Buna örnek olarak Türkiye'de Behçet Sendromunun maliyet analizi çalışması verilebilir. Burada yıllık ortalama doğrudan maliyet 2203 USD, dolaylı maliyet ise 1023 USD olarak hesaplanmıştır. Yıllık toplam maliyet 3226 USD olarak hesaplanmış olup toplam maliyetin %68'ini doğrudan maliyet giderleri oluşturmaktadır.

Amerika'da 2003 yılında yayımlanmış bir başka çalışmada RA hastalarında doğrudan medikal maliyetler incelenmiş ve RA'nın yıllık maliyeti 9519 USD olarak saptamıştır. Bu maliyetin %66,4'ünü ilaç maliyetinin oluşturduğu bildirilmiştir.

Osteoartrit ve RA'da hesaplanan dolaylı maliyetlerin literatür derlemesinin yapıldığı bir başka çalışmada maliyetlerin hesaplanmasında farklı yöntemler kullanıldığı ve bu nedenle maliyetlerin hesaplanmasında bir standardizasyon gereklili-

ğinden bahsedilmektedir. Maliyetlerin belirlenmesinde farklı yöntemlerin kullanımı geçerlilik ve güvenilirlik sorunlarını yaratmaktadır. Dolaylı maliyetleri içeren 32 çalışmanın (25'i RA, 7'si OA) maliyetlerinin değerlendirildiği çalışmada RA'nın yıllık maliyetinin Friction yaklaşımıyla belirlendiğinde 886 USD ile 3585 USD arasında olduğu, Human Capital yaklaşımı ile belirlendiğinde ise 2618 USD ile 18.778 USD aralığında olduğu bildirilmiştir. Yöntem belirtilmeyen çalışmalarda ise Tayland'da 760 USD ile en düşük Amerika'da ise 32.155 USD ile en yüksek olduğu bildirilmiştir. Her ne kadar ülkelerin kişi başına düşen milli gelirleri farklı da olsa yine de RA'nın maliyetleri arasında uçurum olduğu, bu nedenle bir standardizasyon yapılması gerektiği vurgulanmıştır.

2.2. Maliyeti En Aza İndirgeme Analizi

Eşit etki/etkinlik/güvenilirlik (efficacy/effectiveness/safety) profiline sahip alternatif tedavi yöntemleri içindeki en ekonomik tedavi yönteminin hangisi olduğunu belirlemede kullanılan ekonomik değerlendirme yöntemidir. Gerçek dünya koşullarında eşit etki/etkinlik/güvenilirlik profiline sahip bir durum bulmak oldukça zordur bu nedenle bu yöntem nadiren uygulanabilmektedir. Maliyeti En Aza İndirgeme Analizine bir örnek, aynı tedavide kullanılan aynı içeriğe sahip ilaçların ve markaların karşılaştırılması olarak gösterilebilir. Diğer bir ifadeyle ticari bir ürünü (bir ilaç) etki eşitliği (biyoyararlılık eşitliği) bilinen jenerik bir benzeri ile fiyat açısından kıyaslamak ve uygunsa fiyatını azaltmaya yönelmek gibi.

Gerçekte çok az çalışma başlangıçta Maliyeti En Aza İndirgeme Analizi olarak planlanır. Çalışmalar ya Maliyet-Etkinlik Analizi olarak planlanır ve sonuçlar eşit etki/etkinlik/güvenilirlik profiline sahip olduğunda Maliyeti En Aza İndirgeme Analizine dönüşür ya da önceki klinik araştırmaların bilgisi doğrultusunda Maliyet Analizi olarak tasarlanır ve sonuçların eşit etki/etkinlik/güvenilirlik profiline sahip olduğu gösterilir.

2.3. Maliyet-Etkinlik Analizi

Maliyet-Etkinlik Analizi farklı etkinlik ve güven

profiline sahip aynı amaca yönelik alternatif tedavi yaklaşımlarının maliyetlerinin ve sonuçlarının birlikte değerlendirildiği kapsamlı bir ekonomik analiz yöntemidir. Tedavi yöntemlerinin maliyet/etkinlik açısından karşılaştırmalarında sonuç çıktıların aynı ölçüm biriminde olmaları gerekmektedir.

Maliyet-Etkinlik Analizinde maliyetler parasal değer olarak, sonuçlar ise klinik göstergelerle ifade edilirler. Maliyet-Etkinlik Analizinde birbiriyle ilişkisiz maliyet ve sonuçlar karşılaştırılmamalıdır. Maliyet-Etkinlik Analizinde sonuçlar maliyet/etki (cost/effect) olarak ifade edilmelidir. Örneğin, RA'da semptomun gözlenmediği her bir günün kazancı 100 TL gibi. Maliyet-Etkinlik Analizinde etkinlik ölçüsü olarak seçim yapılırken, etkinlik ölçüsünün sonuç sağlık çıktılarıyla her zaman ilişkili olmasına özen gösterilmelidir.

2.3.1. Maliyet etkinlik oranı

Maliyet Etkinlik Oranı (Cost Effectiveness Ratio), bir tedavi yönteminin getirdiği ilave maliyetin sağladığı fazladan etkinliğe oranıdır. Tedavi programlarının bağımsız olarak ele alınmasına göre ve programın diğer bir alternatifiyle karşılaştırılmasına göre iki farklı şekilde hesaplanmaktadır.

a) Bağımsız tedavi programlarında maliyet etkinlik oranı

Bağımsız programlarda, yani tek bir tedavi yönteminde Maliyet-Etkinlik Analizi uygulandığında, Maliyet-Etkinlik Oranı (MEO) aşağıdaki şekilde hesaplanır;

$$MEO = \frac{C_T}{E_T}$$

Formüldeki, C_T Tedavi yönteminin maliyetini, E_T Tedavi yönteminin etkinliğini göstermektedir.

b) Alternatif tedavi programında artan maliyet etkinlik oranı

Aynı sağlık sorununa yönelik kullanılan alternatif tedavi yöntem/programlarının ekonomik değer-

lendirmelerinin yapılmasındaki amaç, bu yöntem/programlarının görece maddi değerlerini (maliyet-etkinlik) belirleyerek tedavi yöntemi seçiminde karar vericilere yol göstermektir. Hastanın sağlık sonuçlarını iyileştiren fakat aynı zamanda toplam maliyeti de arttıran bir tedavi yönteminin hastalar üzerinde uygulanıp uygulanmayacağına karar vermede belirleyici maliyet-etkinlik ölçüsü Artan Maliyet-Etkinlik Oranıdır (R -Incremental Cost Effectiveness Ratio). Pahalı ve özel tedavi yöntemlerinin standart tedavi yöntemlerine göre olan etkinliklerinin araştırıldığı Maliyet-Etkinlik Analizi çalışmalarında R oranının hesaplanması gerekir. R oranı, ilave bir sonucun yarattığı maliyet artışı olarak tanımlanır. Diğer bir ifadeyle bir tedavi yönteminin (A) alternatifi olan diğer bir tedavi yöntemine (B) göre getirdiği ilave maliyetin sağlayacağı fazladan etkinliğe oranlanmasıdır. Bu oran aşağıdaki şekilde hesaplanır.

$$R = \frac{C_A - C_B}{E_A - E_B} = \frac{\Delta C}{\Delta E}$$

Formüldeki, C_A : A tedavisinin toplam maliyetini, C_B : B tedavisinin toplam maliyetini, E_A : A tedavisinin etkinliğini, E_B : B tedavisinin etkinliğini göstermektedir.

R oranının hesaplanmasında kullanılacak verilerin incelenen hasta popülasyonunun rastgele bir örneklemeden elde edilmiş olmaları gerekmektedir. Eğer veriler bu yolla elde edilmişse R örneklem ortalamalarını temel alır ve Artan Maliyet-Etkinlik Oranı,

$$R = \frac{(\mu_{CA} - \mu_{CB})}{(\mu_{EA} - \mu_{EB})}$$

formülüyle hesaplanır. Formüldeki,

(μ_{CA}) ; A tedavi yönteminin gerçek toplum maliyetinin ortalaması,

(μ_{EA}) ; A tedavi yönteminin gerçek toplum etkinliğinin ortalaması,

(μ_{CB}); B tedavi yönteminin gerçek toplum maliyetinin ortalaması,

(μ_{EB}); B tedavi yönteminin gerçek toplum etkinliğinin ortalamasıdır.

Maliyet etkinlik analizinde karar vermede maliyet etkinlik planından yararlanılmaktadır. Bu planda tedavi yöntemlerinin maliyet farkıyla, etkinlik farkı gösterilmektedir. Bu planda aşağıda ifade edilen 4 farklı seçenekle karşılaştırılması olasıdır:

- I Standart tedavi ile yeni tedavi yönteminin maliyet-etkinlik açısından karşılaştırıldığı bir çalışmada yeni tedavi yöntemi daha etkin ve daha pahalıysa bu durumda *Maliyet-Etkinlik Analizi* ya da *Maliyet-Fayda Analizi* kullanarak karar verilmesi gereklidir.
- II İkinci durum, en kötü senaryodur. Yani yeni tedavi yöntemi daha kötü fakat daha pahalı ise yeni tedavi yöntemi RED edilmektedir.
- III Üçüncü durumda yeni tedavi daha az etkiye sahip fakat daha ucuz ise *Maliyet-Etkinlik Analizi* ya da *Maliyet-Fayda Analizi* kullanılarak karar verilmelidir.
- IV Dördüncü, en iyi senaryo. Yeni tedavi yöntemi daha etkin ve daha ucuz ise yeni tedavi yöntemi kabul edilir.

Burada I ve III no'lu durumlarda *Maliyet-Etkinlik Analizi* ya da *Maliyet-Fayda Analizi* kullanılarak karar vermek gereklidir. Bunun için bir *Artan Maliyet Etkinlik Oranı* eşik değeri belirlenir. Hesaplanan oran bu eşik değerinin üzerindeyse yeni tedavi yöntemi reddedilir, altındaysa yeni tedavi yöntemi kabul edilir. Bu kararı verebilmek için *Maliyet Etkinlik Kabul Edilebilirlik Eğrisi* kullanılmaktadır. Karşılaştırılan tedavi yöntemlerinin maliyet-etkinlikleri ve eldeki mali olanaklar göz önünde bulundurularak en doğru kararın verilmesine yardımcı olur.

Yapılan bir çalışmada “Adalimumab+MTX” tedavisi almaya devam edenlerde, HAQ-DI iyileşmesi gözleendiği, fakat maliyetlerin de anlamlı oranda arttığı, 44.000 Euro eşik değerinde “Adalimumab+MTX” tedavisinin rastgele örneklemin %83’ünde maliyet-etkin gözükürken aynı eşik

değerde ‘Etanercept Monotherapy’ rastgele örneklemin %72’sinde maliyet-etkin olduğu bildirilmiştir.

2.4. Maliyet-Fayda Analizi

Maliyet-Fayda Analizi herhangi bir hastalık ya da sağlık sorununa ilişkin tedavi yöntemlerinin maliyetleri ve elde edilen sağlık sonucunun “kalitesi” üzerine odaklanan bir değerlendirme yöntemidir. Maliyetlerin parasal değer olarak ve sonuçların hastanın yaşam kalitesi ölçülerini içeren klinik terimlerle ölçüldüğü durumlarda kullanılır.

Maliyet-Fayda Analizinde yarar ölçüsü olarak çoğunlukla Kalite Düzeltmeli Yaşam Yılı “Quality Adjusted Life Year - QALY” birim olarak kullanılır. QALY yaşama ilişkin niceliksel ve niteliksel değerleri içeren bir ölçüttür. Sonuçların ölçüsü olarak QALY kullanımı, farklı ekonomik analizlerden elde edilen maliyet/yarar oranları için doğrudan karşılaştırma yapmaya olanak sağlar.

1 QALY, kişinin mükemmel sağlıkla bir yıl yaşaması ya da kişinin %50 sağlıkla 2 yıl yaşaması, %25 sağlıkla 4 yıl yaşaması gibi değerlere eşittir.

‘Yarar’, ‘Değer’ ve ‘Tercih’ kavramları bazen birbirinin yerine kullanılabilir gibi görülmektedir fakat gerçekte bu üç kavram birbirinden farklıdır. Hangi kavramın kullanılacağı ölçümlerin nasıl yapılacağına bağlıdır. Ölçüm işleminde iki temel yaklaşım vardır. Birincisi sonuçların belirli ya da belirsiz olduğu durumlar, ikincisi ise bireylere seçim ya da ölçek uygulanıp uygulanmayacağıdır.

QALY katsayıları, hastalıklara özgü geçerli yaşam kalitesi ölçekleri (örneğin EQ-5D ölçeği), zaman ödünleşmesi (time trade-off) yaklaşımı, standart risk yaklaşımı ve görsel analog ölçek (Visual Analogue Scale) gibi yöntemler kullanılarak hesaplanabilir.

Zaman ödünleşmesi yaklaşımında sağlık durumunu iyileştirmek için bireylerin yaşamlarından gönüllü olarak feragat edebilecekleri süre miktarının cevabı aranır. Örneğin kronik bir sağlık sorunuyla 10 yıl yaşamak, tedaviyi alırsa ve tam

sağlığa ulaşırsa kaç yıl yaşamaya eşit olduğu sorusunun cevabı aranır. Her bir hastaya bu soru sorularak bir kalite skoru elde edilir.

QALY katsayılarını hesaplamada bir başka yaklaşım, *standart risk* yaklaşımıdır. Daha iyi bir sağlık durumuna erişebilmek için bireyin alabileceği risk düzeyinin belirlenmesinde kullanılan bir yaklaşımdır. Örneğin kişi kronik durumu seçerse kronik sağlık problemiyle yaşayacaktır. Fakat tedavi yöntemini seçerse p olasılığı ile tam sağlığa kavuşacaktır, 1-p olasılığıyla da bu seçtiği yöntemin ölüm ya da kronik durumdan daha kötü bir sonucu varsa o riskle karşı karşıya kalacaktır. Bu yöntemde tam sağlığa kavuşmak için kişinin alabileceği risk düzeyi belirlenerek bir kalite katsayısı hesaplanmaktadır.

RA başlayan hastalarda tedavi yöntemlerinin ‘Ardışık monoterapi (Sequential monotherapy)’, ‘Adımsal kombinasyon terapi (Step-up combination therapy)’, ‘Prednizonlu başlangıç kombinasyon terapi (Initial combination therapy with prednisone)’ ve ‘İnfliksımabl başlangıç kombinasyon terapi (Initial combination therapy with infliximab)’ Maliyet-Fayda Analizinin yapıldığı bir çalışmada İnfliksımabl başlangıç kombinasyon tedavisinin diğer yöntemlerden anlamlı olarak daha yüksek yaşam kalitesi sağladığı fakat maliyeti çok yüksek olduğundan Prednizonlu başlangıç kombinasyon tedavisinin tercih edilmesi gerektiği bildirilmektedir. Ayrıca, 1 QALY için ödenecek tutar 22.000 Euro’dan büyük tüm noktalarda ‘İnfliksımabl başlangıç kombinasyon tedavisi’ diğer tedavilerden daha üstün olduğu bildirilmektedir.

2.5. Maliyet Kar Analizi

Maliyet Kar Analizi herhangi bir programın/tedavinin girdilerinin (maliyetlerin) ve çıktılarının (sonuçların) parasal değer olarak ölçümlendiği çözümleme türüdür. Maliyet-Kar Analizi ile farklı tedaviler/programlar değerlendirilebilir. Uygulanan tedavi yönteminin bireye sağladığı yarar ve maliyetlerde sağlanan tasarruf analizin temel mantığını oluşturur. Değerlendirme sonucu, gelecekte ortaya çıkabilecek tıbbi bakım harcama-

larının azaltılmasını, kayıp işgününün azaltılması gibi sonuçları içerebilir. Maliyet-Kar Analizi bir “yatırımcının”, geniş bir ana konu çerçevesinde, farklı konumlarda, farklı girişimlerin sonuçlarının karlılıklarını doğrudan finansal birimlerle kıyaslamasıdır. Örneğin, sağlık yatırımları yapılması açısından MI sonrası rehabilitasyon tesisi veya kalp transplantasyon merkezi yapılması seçeneklerinden birine karar verilirken yapılan değerlendirme gibi.

Maliyet-Kar Analizi artan program maliyetleriyle artan program kazanımlarını karşılaştırır arasındaki fark programın net sosyal karıdır. Daha basit ifade etmek gerekirse analizin amacı programın karının programın maliyetini aşip aşmayacağını belirlemesidir. Maliyet-Kar Analizinde değerlendirme yapılırken ‘Net Kazanç’ değerinin belirlenmesi gerekir bu aşağıdaki formül yardımıyla hesaplanır,

$$NK_i = \sum_{t=1}^n \frac{b_i(t) - c_i(t)}{(1+r)^{t-1}}$$

NK_i : i projesinin net kazancı (indirgenmiş)

$B_i(t)$: i projesinin t yılındaki karı

$C_i(t)$: i projesinin t yılındaki maliyeti

$1/(1+r)$: yıllık indirim faktörü

n : proje süresi

$i=1,...,l$ olası yatırımlar

Maliyet-Kar Analizinin birincil amacı projenin net kazancının sıfırdan büyük olmasıdır ($NK>0$). Oran birden büyükse eylem yapılmaya değerdir. Bir başka deyişle, maliyeti en az, ancak getirisi en fazla olan yolu seçmek de mümkün olabilmektedir. En iyi ekonomik analiz olarak incelenmesine karşın hastanın hayatına ve elde edebileceği sağlık sonuçlarına parasal değer yüklemenin zor olmasından dolayı eleştirilmekte ve bu nedenle de sağlık alanında kısıtlı uygulamalara sahiptir.

3. Sonuç

Son yıllarda sağlık ürünlerinde ve tedavi yöntem/programlarında gözlemlenen hızlı gelişmeler karşısında sağlığa ayrılan bütçelerin sınırlı kalması bu tür ürün ve hizmetlerin ekonomik değerlendirmelerinin yapılmasını gerekli kılmıştır. Ekonomik analizler sağlık ürünlerinin, hizmetlerinin ve tedavi yöntemlerinin değerlendirmelerini yaparlar. Sağlık sisteminde tedavi yöntemlerinin değerlendirilmesi için tanımlayıcı ve analitik yaklaşımlara sahiptirler. Hastalık yönetimi, tedavi, tanılama ve önlemede tedavi yöntemlerinin klinik, insancıl ve ekonomik bakış açılarını belirtirler. Bu nedenle karar vericilere, sağlık kaynaklarının en iyi nasıl kullanılacağı konusunda çözümleyici bilgiler sağlarlar. Maliyet, Maliyeti En Aza İndirme, Maliyet-Etkinlik, Maliyet-Fayda ve Maliyet-Kar analizleri gibi çözümleme yöntemleriyle kısıtlı kaynakların etkin kullanımı konusunda sağlık sistemindeki karar vericilere yol gösterirler.

Gelişmiş ülkelerde maliyet analizlerinin yaygın kullanıldığı ve büyük bir öneme sahip olduğu gözlenmesine karşın, ülkemizde bu tür analizlerin yetersiz olduğu görülmektedir. Bu durum ülkemizde sağlık ürünlerinin ve tedavi yöntem/programlarının maliyet-sonuç dengelemelerinin en iyi (optimal) şekilde yapılmadığının göstergesidir.

Sonuçta, bu tip analizlerin ülkemizde de yaygınlaşmasıyla sağlık ürünlerinin ve tedavi yöntem/yöntemlerinin maliyet-sonuç dengelemelerinin daha iyi yapılarak, hastaya, sağlık sistemine ve topluma en iyi yarar sağlayacak klinik zemin daha doğru belirlenebilir ve hasta/sağlık sistemi açısından daha doğru kararlar alınabilir. Bunun sonucunda da eldeki kaynaklar en iyi şekilde kullanılarak olası kaynak israfının önüne geçilebilir.

Kaynaklar

1. Berger ML, Bingefors K, Hedblom EC, Pashos CL, Torrance GW. Health Care Cost, Quality and Outcomes, International Society for Pharmacoeconomics and Outcomes Research Press, NJ, 2003.
2. Drummond MF, O'Brien BJ, Stoddart GL, Torrance

GW. Methods for the Economic Evaluation of Health Care Programmes, New York, Oxford University Press, 1997.

3. Drummond MF, McGuire A. Economic Evaluation in Health Care, New York, Oxford University Press, 2001.
4. Sutton M. How to get the best health outcome for a given amount of money, BMJ 1997; 315: 47-49.
5. Goeree R, O'Brien BJ, Blackhouse G, Marshall J, Briggs A, Lad R. Cost-Effectiveness and Cost-Utility of Long-Term Management Strategies for Heartburn. Value Health 2002; 5: 312-328.
6. Hunink M, Glasziou P. Decision Making in Health and Medicine, Cambridge University Pres, 2001.
7. Michaud K, Messer J, Choi HK, Wolfe F. Direct medical costs and their predictors in patients with rheumatoid arthritis: a three-year study of 7,527 patients. Arthritis Rheum. 2003 Oct;48(10):2750-62.
8. N J Bansback, A Brennan, O Ghatnekar. "Cost effectiveness of adalimumab in the treatment of patients with moderate to severe rheumatoid arthritis in Sweden" Ann Rheum Dis 2005;64:995-1002.
9. Palmer S, Byford S, Raftery J. Types of Economic Evaluation, BMJ 1999; 318: 1349.
10. Raftery J. Costing in Economic Evaluation. BMJ 2000; 320: 1597.
11. Rapier CM. An Introduction to Outcomes Research, Brookwood Medical Publications, 1996.
12. Rychlik R. Strategies in Pharmacoeconomics and Outcomes Research, New York, Pharmaceutical Products Press An Imprint of The Haworth Press, Inc., 2002.
13. Süt N. Maliyet etkinlik ve maliyet fayda analizleri, RAED Dergisi, 2013; 5(1):20-24.
14. Süt N. "Hastalık maliyet analizi", RAED Dergisi, 2011; 3(3-4):1-5.
15. Süt N, Seyahi E, Yurdakul S, Şenocak M, Yazıcı H. A cost analysis of Behçet Syndrome in Turkey. Rheumatology (Oxford) 2006; 46: 678-682.
16. Süt N. Tedavi Yöntemlerinde Ekonomik Değerlendirme Yaklaşımları: Farmakoekonomi ve Araştırma Çıktılarının Değerlendirilmesi, Doktora Tezi, İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi, 2005.

17. Udvarhelyi S, Colditz GA, Epatin AM. Cost-Effectiveness and Cost-Benefit Analyses in the Medical Literature: Are the Methods Being Used Correctly?, *Ann Intern Med* 1992; 116: 238-244.
18. van den Hout WB, Goekoop-Ruiterman YP, Allaart CF, de Vries-Bouwstra JK, Hazes JM, Kerstens PJ, van Zeven D, Hulsmans HM, de Jonge-Bok JM, de Sonnaville PB, Dijkmans BA, Breedveld FC. Cost-utility analysis of treatment strategies in patients with recent-onset rheumatoid arthritis. *Arthritis Rheum.* 2009 Mar 15;61(3):291-9.
19. Venturini F, Johnson KA. Introduction to Pharmacoeconomic Principles and Application in Pharmacy Practice, *CJHP (California Journal of Health-System Pharmacy)*, 2002; 14: 6-15.
20. Walker D, Fox-Rushby JA. Economic Evaluation of Communicable Disease Interventions in Developing Countries: A Critical Review of The Published Literature, *Health Econ* 2000; 9: 681–698.
21. Xie F. The need for standardization: a literature review of indirect costs of rheumatoid arthritis and osteoarthritis. *Arthritis Rheum.* 2008; 15;59:1027-33.

“LETTER TO EDITOR, RHEUMATOLOGY”DE SON DURUM

Hasan Yazıcı

*Romatoloji Bilim Dalı
İstanbul Üniversitesi
Cerrahpaşa Tıp Fakültesi*

Bir reklam arasına geldik, bu konuşmamın esası bir reklam aslında. Letter to Editor Rheumatology’nin reklamı. Birkaç yıl önce Almanya’da bir çalışma, bir analiz yapmışlar. Pubmed’de 3 çeşit yazı çeşidi var, birincisi editöre mektuplar, ikincisi geleneksel klinik çalışmalar, üçüncüsü de derleme makaleleri. Derleme makaleleri sayıca en çok olan. Burada bir sorun var. Garfield 1994 yılında Tübitak davetlisi olarak Cerrahpaşa’daydı. O zaman herkes nasıl yayın yapalım diye Garfield’a soruyordu. Garfield’i tanıyorsunuz, Science Citation Index’in babası. Garfield, “İyi dergilerde derleme yazarsanız muaz- zam sitasyonunuz artar.” dedi. Garfield her konuş- ma yaptığı yerde aynı şeyi söylüyor. Böyle olunca da herkes bunu yapıyor. Bu tahammül edilebilir bir olay değil, yani biz kritik düşünce diyoruz, Garfield ne diyor! Bu arada tabii editöre mektup gidiyor da bu editöre giden mektupta ne yazıyorlar? Biz Emine’yle beraber hâlâ üzerinde çalışıyoruz. Bi- zim 3 tane ana dergimizi aldık, bu 3 ana dergide yazılan mektupların içinde ne kadar sert mektup var, sertlik derecesi birinci biraz kötü, ikinci çok kötü, üçüncü daha da kötü. Bir de benim yıllar önce yazdığım bir mektup var, onu örnek olarak kullandık, standardımız bu mektuptur, dedik. Buna göre baktığımızda yazar gibi görünüyorlar ama esasında kritik eden yok. Biz iyi iş yaptık diye bunu EULAR’a gönderdik, kongre kitabına bile kabul etmediler. Böyle olunca bir dergi başlatmak düşüncesi ortaya çıktı. O da www.lettertoeditor.org. Teknolojiden yararlanarak böyle bir dergi yayımlıyoruz. Bu siteye girdiğimizde “all the views that are fit to print” göreceksiniz, bu New York Times’ın mottosu, biz bunu aldık. Ondan sonra bir editörler kurulu kurduk. Aranızda bu editörler kurulunda olanlarınız da var. Büyük bir listemiz var, Türkler var, yabancılar da var. Editöre mektup gönderirken ne yapmamız gerektiği sitede açıklanıyor. Ve bu mektupların 600 kelime limiti var. Uzun yazarsanız derleme gibi yapıp 1000 kelimeye kadar gidebiliyor. Ancak hepsi hakeme gidiyor bunların. Hakeme gittikten sonra ayrıca orijinal yazara da gidiyor. Eğer bize Türkçe gönderirseniz, biz İngilizceye çevireceğiz. Şu anda 3 tane var, dördüncü de hakemlerde kabul edilmiş durumda. 7-8 tane de elimize mektup gelmiş. Son 3 tanesi hakemlere gidiyor, hakemlerden geldi, yazara gönderdik, ocaktan beri bu işle meşgulüz. Peki,

bunları yaptıktan sonra ne oluyor? Son bir hafta evveline kadar aktivitesi 278'di. Çok ahım şahım bir rakam değil ama geçen hafta bütün ACR membership'e gönderdik. Birden bire 222 arttı. Aralarında ilginç yorumlar yapanlar da var. Niye bu işi yapıyorsunuz, diye soruyorlar. Bir tane de çok güzel bir yorum geldi, 'Ben sadece hocalarımın dediğini dinlerim,' dedi. 'Ben editöre mektup yazmıyorum,' diyenler, 'Ne iyi bir fikir bu' diyenler var. En fazla Amerika'dan ve Türkiye'den geliyor.

Şimdi vereceğim örnekler New England Journal'dan. Niye? New England en çok okunan dergilerden. Meşhur bir makale var. Bir oral tirozin kinaz inhibitörünün romatoid artritte kullanımı. Bu makaleyi okuduk, makale "bu iyidir" diyor. Fakat iyidir derken tablo 1'e bakıyorsun, tablo metotreksata karşı iyidir, diyor. Tabloda plasebo var, 150 mg var, yeni ilaç 300 mg var. Ve ondan sonra başlangıç metotreksat dozu ne kadar diye baktığınız zaman, başlangıç metotreksat dozu verilmiyor. Şimdi eğer romatoid artrit iyi gelmiyor diyorsa, 7,5 mg'a mı iyi gelmiyor, 20 mg'a mı iyi gelmiyor? Bu bilgiyi vermeden, sonucu değerlendirmek tamamen imkânsız. Bilmek gerek. Metotreksata yanıt vermiyor diyen sürpriz makaleler de var. Burada Bolivya var, Romanya var, bir de Amerika var. Ve daha evvel de başlarına gelmiş, Bolivya'da, Romanya'da vs metotreksat dozu düşük veriliyor. Amerika'da 17,5 gr ortalama haftalık. Ama bu çalışma bunu vermiyor. Editöre ve yazarlara mektup yazdık, cevap gelmedi. Cevap verirlerse yayımlarız. Hâlâ bekliyoruz hatta ben ACR'de yazarlara niye cevap vermediklerini soracağım. Biraz yaş ilerleyince bunları daha rahat yapabiliyorsun.

Burada biz metotreksat dozunu arıyorduk ama bu cetvelde kenarda p değerleri var. Bu p değerlerini görünce, bizim co-editörümüz Maarten Boers kızdı, CONSORT'a göre randomize kontrollü çalışmalarının hasta özelliklerini gösteren 1. tablolarına p değeri koyamazsınız, dedi. Niye koyamazsın, çünkü randomizasyon olduğu zaman hastalar ve kontrol grupları arasındaki başlangıç özellikleri zaten şansa bağlıdır, bunları buraya koymak biraz göz boyamak dedi. Bunun üzerine yazı yazdı. Bu yazdığı yazıda diyor ki, New England Journal'da 2010 yılının Eylül ayında yaklaşık 7 tane randomi-

ze çalışma yayımlanmış. Bunların hepsinde Cetvel 1'de p değeri var. Siz CONSORT'a bağlısınız, bu ne çelişki. Bu mektup onu söylüyor. New England Journal'a gönderdik. New England Journal editörüyle özel yazışmalarım var benim. Özetle onlar da yanıt vermediler. New England Journal of Medicine'da bir sessizlik var ama herhalde günün birinde cevap verecekler. Şimdi, demek Cetvel 1'de bunları vermek akılsızlık, göz boyama ve CONSORT'ta veremezsin diyor. Bunu yapanlar var, bu yanlışlık değil de mantıksızlık diyor. Böyle olunca da istediğin zaman bunları kullanabilirsin tabii.

Ondan sonra bir başka örnek, New England Journal'da bir yazı çıktı. Tüberküloz tanısı koymak için yeni bir yöntem. Gülen, yazarlara kibarca 'Bu hesaplarınız doğru değildir.' yazdı. Fevkalade güzel, efendice cevap geldi. Yani insanlar güzel yazıldığı zaman buna cevap veriyor. Tartışmanın devam etmesi gerekiyor.

Niye mektup yazıyoruz?

1. Bir şeyi teklif etmek için, önermek için,
2. Katılmamak için,
3. Düzeltmek için,
4. Yeni bilgi vermek için,
5. Öğretmek için yazıyoruz.
6. "Ben de düşünüyorum." Bu da çok önemli. Eğer bir grup insan "Ben de düşünüyorum." derse kendini rahatlatıyorsun.
7. Bazen de "Katılıyorum." diye yazıyoruz. Katılıyorum demek esas amaç değil. Katılıyorum diye habire mektup yazmak tatsız bir olay.

Peki, nelere dikkat etmek lazım? Bir defa bu işe hevesli olmak lazım. Girin internete bakın, neredeyse 100 seneden beri makale tartışmaları var. Türk üniversitelerinde makale tartışmaları yok veya çok az var. Bizim eğitimimizde muazzam bir eksik bu. Bu makale tartışmalarının daha çok olması lazım ki editöre mektuplar daha fazla çıksın, düşünce çıksın. Bunu yazarken ne olur klişelerden kaçının. Bütün mektuplarınız "We read an interesting...", tamam, interest olmasa zaten yazmayacaksınız. Böyle başlamayalım. Kısa ve öz olmalı.

Gülen'in yazdığı mektupta başka şeyler de vardı. Hepsini birden söylemeye kalkmayın, en önemli

bir veya iki nokta vermek lazım. Olabildiğince spesifik ve rakamsal olmak gerekiyor. Esas orijinal makalenin söylediğini tekrarlamaya gerek yok. O zaten orada var, çok kısa bir kısmı olabilir. Kibar olun. İngilizce en kızdığım laflardan biri “balanced view”. Balance politikacıların işi. Biz burada ‘balance’ için değiliz. Editörler bazen “Balanced view vermiyor” yazıyor. Balanced view olsa zaten sana göndermem ben bunu. Balance tüccarın, politikacının işi. Biz öyle değiliz. Onu istemiyoruz. Ondan sonra, bir de, ‘Benim makaleme neden referans vermedin?’. Yani olabildiğince kendini beğenmekten, narsizmden kaçınmak lazım. Ama çok aşikârsa, hakkını yemişse öyle değil, tamam. Ondan sonra, eğlendirmek güzel bir şey, komik olabilmek, çok ciddi yazmaya gerek yok. Duruma göre komik olabilirsiniz. Tabii dergi ne istiyor, ona göre yapın. Mektubu ilk başta kendi kendine sonra da başkalarına yüksek sesle okumak iyi bir fikir. Ondan sonra düşünün üzerinde. Arkadaşlarınıza söyleyin mektup yazdım diye. Her yazdığınız mektubu da göndermek zorunda değilsiniz. İnsan kızar eder ama bazı mektupları yazıp kendinizde de tutabilirsiniz. Popper’ın bir lafı var, diyor ki, “Akılcı olmanın yegâne yolu başkalarının kritiğini dinleyebilmek.” Romatolojide çok az kritik dinliyoruz. Prof. Lesaffre’nin konuşmasında var, hiçbir istatistiği beğenmiyor, %100 hakkı var bunda. Böyle bir durum var, kendi yaptıkları iyi, onu istemiyoruz.

Özetle arkadaşlar, bu da bizim sloganımız, Don Kişotluk yapmak istiyoruz. Benim söyleyeceklerim bu kadar, lütfen omuz verin ve bu işi götürelim. Teşekkürler.

Yaşar Karaaslan: New England’daki tirozin kinaz inhibitörleri ile ilgili çalışmada siz p değerlerine baktınız, ben de bu çalışmada böyle hastaları nereden buldular diye takıldım?

Hasan Yazıcı: Bu hastaları bir kısmı Amerika’dan, Orta Doğu’dan, Güney Amerika’dan. Avrupa’dan hasta yok burada. Amerika ve bir miktar geri kalmış ülkeler. Onların arasında ilaca veya plaseboya randomize ediyorlar.

Yaşar Karaaslan: Böyle hastaların pratikte hiç olmayacağını düşünmüştüm. Bu tablonun aynısı-

nı görmüştüm, o kadar eski hastalardı ve o kadar az DMARD almışlardı ki ben klinisyen olarak bu tirozin kinaz inhibitörleriyle ilgili çalışmada hastalara baktım. Pratikte buraya alınan hastalar, Arthritis’de galiba, bununla ilgili bir çalışma, sanırım bunun bir sonraki aşaması yayımlandı. Hastalar o kadar yaşlı, 17-18 yıllık hastalar ve çok az ilaç almışlardı. Pratikte düşündüm acaba böyle hasta nerede bulunabilir diye. Güney Amerika’da acaba bu hastalar dağda yaşayıp hiçbir şekilde romatolog görmemiş hastalar mı? Yani çalışmaya alınan hastalar pratikte bulunabilir hastalar değildi.

Bir Katılımcı: Benim de 2 gözlemim olacak. Bir tanesi Gülen’in çalışmasına cevap veren kişi, sanırım o kiti geliştiren firmadan biri aynı zamanda. Şöyle düşündüm açıkçası, akademisyenlerden çok, böyle letterlardan kaygısı olan endüstri daha çabuk cevaplar verip daha dürüst olabilir.

Hasan Yazıcı: Çok güzel bir nokta. Onlar, hatta editoryali yazan da öyleydi ama biz dedik ki bunu tek bir noktaya getirelim çünkü ondan 15 gün evvel New England Journal’la öbür Maarten Boers mektubu için aramızda belki 4 tane email gidip geldi, yani New England Journal, Lettertoeditor kendine savaş mı açtı diye düşünmesin istedik ve böylece balanced view (!) yapmaya çalıştık.

Bir Katılımcı: İkinci söyleyeceğim, son 15 gün içinde ben iki çalışma reddettim, ikisinde de metotreksata dirençli hastalara yeni bir biyolojinin kullanılmasıyla ilgili faz 3 çalışma tekniği idi. Şöyle bir argüman yürüttüm; metotreksata dirençli hastaya benim biyolojik vermem için o hastaya yalan söylemem lazım. Nedir şu an için durum, Türkiye’de anti-TNFlar var, rituksimab var, abatasept var, siz bu ilaçlar dururken bunlardan daha etkin olduğu kanıtlanmamış güvenirliliği de tartışmalı, bunu etik anlamda da soruyorum artık, tartışmalı bir ilacı kullanır mısınız? Şimdi ben bu hastalardan informed consent alacağıma inanmıyorum ve firmaya bu çalışmaya hasta bulamam dedim. Ayrıca bu iş etik açıdan da sorunlu. Bilmiyorum böyle bir letter ister misiniz ama bence kontrollü ilaç çalışmalarının Batıdan Doğuya kaymasının nedeni de bu.

Hasan Yazıcı: İsteriz ama referans da isteriz. Biliyorsun geçen sene Arthritis Research and Therapy yayımladık, dedik ki lütfen informed consentler herkes tarafından bilinsin. Eğer böyle olur ise zaten sorunların çok büyük kısmı hallolur.

Bir Katılımcı: Ben şu an itibariyle biyolojilere dirençli hastalarla ilgili bir çalışma dışında bir ça-

lışmayı etik olarak kendi hastalarım önermemeyi düşünüyorum.

Önder Ergönül: Enfeksiyon hastalıklarında da mektup yazmak kabul oluyor mu?

Hasan Yazıcı: Eğer sen bize Kongo kanamalı hummasının lupusla karışma ihtimali var dersen düşünürüz.

Neredeyse 10 yıl kadar önce Prof. Yazıcı beni odasına çağırdı ve “Vedat, kayıt tabanı (registry) oluşturmamız eve git, hazırlan ve 48 saat sonra gel” dedi. Biyolojik ilaçların Türkiye’de piyasa çıkması bekleniyordu. Bir firma direktörü Prof. Yazıcı’ya gitmiş ve Türkiye’deki biyolojikler için bir kayıt tabanı oluşturma fikrini belirtmişti. O zamanlar kayıt tabanları pek popüler değildi, sadece yeni başlanan İngiliz ya da İspanyol kayıt tabanları biliniyordu. Benim bildiklerim de COX2 inhibitörleri kayıt tabanlarıyla sınırlıydı. Araştırdım, 2-3 haftada bir şeyler hazırlamaya çalıştım. Fakat tamamladığımda ilaç firması vaz geçmişti, biz de projeyi unuttuk. Bugünlerde tekrar kayıt tabanı projesi gündeme geldi. Bu nedenle bu oturumu yapıyoruz.

İlaç yan etkileriyle ilgili anlık bildirimler yapmakta ancak bunlar çok uygun değil. Uzmanlar gördükleri yan etkileri her zaman rapor etmiyorlar ve kontrol grupları da yok. Gözleme dayalı çalışmalar ise retrospektif ya da prospektif oluyorlar ve kayıtlar diğer metotlarla karşılaştırıldığında daha uygun. Ancak daha pahalılar, daha fazla zaman ve daha fazla çaba gerektiriyorlar. Bu kayıtlar gerçek hayatı daha iyi yansıtırlar, çünkü tüm hastaları kayda alıyorsun, kontrollü çalışmalarda olduğu gibi hastayı dışarıda bırakmıyorsun, komorbiditeli hastalar dâhil. Böyle olunca da hasta sayısı çok olduğu için gerçekleri yansıtıyor ve sonuçlar tüm hastalar için genelleştirilebiliyor.

Yönlendirme yanlılığı (channeling bias or confounding by indication): Bir ilaç için kayıt yapıyor ve hastayı bunun için seçiyorsun. Kontrol grubunu tespit etmek zordur çünkü kontrol grubundaki hastalar biyolojik kullananlarla benzerlik göstermelidir. Hastanın takibi de zordur çünkü kontrol grubu da aynı şekilde takip edilmelidir. Eğer uzman ilacı tanıyorsa, örneğin hasta anti-TNF kullanıyorsa, metotreksata kıyasla tüberküloz ya da fırsatçı enfeksiyona daha fazla önem verecektir. İstatistik problemler de kayıt sistemlerinin zorluklarındandır. Eğer yan etkiler çok nadirse ulusal kayıt sistemi yeterli olamayacaktır ve diğer ülkelerin kayıt sistemleri ile birlikte çalışmak gerekecektir.

Kayıt sistemlerinin olmazsa olmazları vardır ve bunlar Türkiye için de zorluklardır. Veri girişi has-

RAED HASTA KAYIT VERİTABANI

Vedat Hamuryudan

*Romatoloji Bilim Dalı
İstanbul Üniversitesi
Cerrahpaşa Tıp Fakültesi*

Kayıt sistemlerinin zorluk - zayıflıkları

- Yönlendirme yanlılığı (Channeling bias & confounding by indication)
 - Etkinlik değerlendirmesini zorlaştırır
 - Biyolojik ilaçlar daha ağır hastalığı olanlara yazılıyor.
 - Tedaviye devam edenler kesenlere göre daha iyi prognozlu
 - İlaç seçmede yanlılık
- Kontrol grubu oluşturmada zorluk
 - Kontrol grubu ilaç grubuna benzer özellikte olmalı
 - DMARD almış – almamış
- Hasta takibi
 - Kontrol grubu aynı şekilde takip edilmeli
 - Kullanılan ilacı bilmek belli yan etkiler için daha fazla dikkat doğuruyor.
 - Biyolojik ilaçlarla görülen yan etki daha fazla bildiriliyor.

1. Slayt

ta görüldüğü anda yapılmalıdır, tabloyu hazırlayıp veriyi girmesi için sekretere vermek yeterli değildir. Bunun anlamı datanın denetimi ve geçerliliği zayıf demektir. Her takip ziyareti kaydedilmelidir ve klinikler arasında işbirliği gereklidir. Gelecekte bu veri başka ülkelerle birleştirilebileceğinden daha başlangıçtan onların veri tabanlarıyla uyumlu olmalıdır.

Hangisi hasta veritabanı değildir?

1. ARTIS
2. DANBIO
3. CORRONA
4. DREAM
5. BSRBR
6. IMPACT
7. BIOBADASER
8. RATIO

Cevaplar çok iyi, %56 katılımcı IMPACT demiş. Bunların hepsi IMPACT hariç kayıt sistemi. IMPACT ise psöryatik artritte yapılmış bir kontrollü ilaç çalışması.

Bütün registrylerde genellikle hastalar başından kaydediliyor ve takip ediyorsunuz. Ama bunlardan sadece bir tanesinde veri tabanına sadece olay çıkınca, yani görmek istediğin olay, örneğin lenfoma, tüberküloz çıkınca sokuyorsunuz hastayı.

Aşağıdakilerden hangisinde hasta kaydı yan etki çıkınca yapılmaktadır?

1. RATIO
2. DANBIO
3. BSRBR
4. ARTIS
5. CORRONA
6. BIOBADASER
7. DREAM
8. RABBIT

%17 BSRBR denmiş. İngiliz veritabanı. Beşincisi Amerika'daki en eski veritabanlarından bir tanesi. Rabbit Almanlar'ın, bunların üçü de birbirine benziyor. Prospektif gidiyor. Buna karşılık bir tek RATIO Fransız veri tabanı. Bütün ülkeyi esasında

RAED hasta veritabanı nasıl olmalı?

1. Tüm hastalıkları içeren tek bir veritabanı
2. Her hastalık için ayrı veritabanı
3. İlaçlara göre düzenlenmiş veritabanı

2. Slayt

temsil etmiyor. Ama buna kaydolun hastalar var ve lenfoma, tüberküloz bir TNF inhibitörü kullanan hastada çıktığı zaman bunu ve buna yakın aynı havuzdan başka bir kontrol buraya bildiriliyor. Yani 1. şık RATIO doğru.

Bütün bu bilgilerden sonra, bir veritabanı hazırladığımız taktirde bu nasıl olmalı?

- Tüm hastalıkları içeren tek bir veritabanı mı olsun?
- Her hastalık için ayrı bir veritabanı mı olsun?
- İlaçlara göre düzenlenmiş veritabanı mı olsun?

Tüm hastalıkları dediğim zaman da belki lupusu ayırmak romatoid artrit AS diyelim yani biyolojikleri hedef alan bir şey. Her biyolojik için ayrı mı, AS için RA için ayrı ayrı mı, yoksa AS, RA, psöryatik artrit hepsi bir havuzda mı?

Bir kere, ilaçlara göre istenmiyor diye düşünüyorum. Ayrı ayrı ilaçlara göre olmasın, hastalıklara göre olsun. O da ikiye ayrılıyor. Her hastalık için biraz daha ağır basıyor.

Şayet böyle bir veritabanı yapacaksak, biraz evvel zorlukları da söyledim. Ama bütün bu zorluklara rağmen, kliniğe gelen bütün hastaları kaydetmek mi lazım? İdeal olan tabi bu. Yoksa bizim iş yükümüz fazla, her hastayı giremeyiz, belli bir randomizasyona göre biz hastalarımızın belli bir kısmını mı alalım? Her klinik için farklı. Yüküne göre öyle. Ne dersiniz bu şekilde olur mu? Yani randomizasyona karşı mısınız, ona göre cevaplayalım.

Hasan Yazıcı: Sorun şu ki, kayıt sistemi organizasyonları biraz kaotik. Sebebi şu, bütün ilaç firmaları registry istiyor, bu onların rekabeti için gerekli ama biz hastalarımızı etkinlik ve yan etki de takip etmek istiyoruz. Bu ilaçların bazı yan etkileri var. İdari makamlar, düzenleyiciler, bu ilaçlar için para ödeyenler bu kayıtları istiyor. Onları kontrol etmek istiyorlar. Hastaların bu registryleri isteyip istemediğini bilmiyorum. Biz iyi yayınlar istiyoruz, bu yüzden bu kayıtlara sahip olmamız gerektiğini biliyorum. Ayrıca tecrübelerimizi bi-

RAED veritabanına hangi hastalar alınmalı?

1. Tüm hastalar
2. Rastgele seçilmiş hastalar

3. Slayt

lime dönüştürme meselesi de var. Hastalarımızı gerçek hayat koşullarında görüyoruz ve bunu da bilime çevirebilmeyi umuyoruz ve bu da meziyet gerektiriyor. Ben bu romatoid artritli hastaları gördüğüm için çok iyi bir data havuzum var ve çok iyi de istatistikçilerim. ACR için hazırladığım bir tebliğim var, eğer her şeyi doğru koymayı başarırısam. 1970'lerde romatolojiyi öğrenmeye başladığımda akademik inanış her zaman şuydu: Datayı her zaman hipotezine göre topla. Bu bir yolu, yapmak istediğin her çalışma için ayrı data oluştur ve tam tersi datayı topla. Sonrasında o datayla ne yapacağını düşün. Daha da ileri düşünebiliriz, tümünden gelime karşı tüme varım diyebiliriz. Dernek olarak her iki yaklaşımı da uzmanlardan alıyoruz. Oğlum New York'da sizin bir çoğunuzla işbirliği içinde, New York'lu hastalarla İstanbul, Konya vs hastaları karşılaştırıyor. Ankara, İstanbul ile de yapmayı istiyoruz. Bunu yaparken birkaç önemli konu var. En iyi dergilerimizde bile registryler çok kötü ele alındı. Bazılarını tartıştık. Kayıt sisteminde belirlenenle ulusal datanın karşılaştırılmasının istatis-

tikleri çok kötü. Bütün anti-TNF ajanları, birçok isim tarafından mortalite ve morbidite vs'ye bakan pek çok yayın var. Bu ilaçları kullandığınız zaman bunun için gayet iyi bir sebebiniz var çünkü bu ilaçlar için bütün double blind çalışmalar gösterdi ki eğer anti-TNF denerseniz ve 6 ay, 1 ay, ya da 1 yıl double blind çalışmayla karşılaştırırsanız ilaç grubunda daha fazla maliniteler olacaktır. Ama kayıt sistemlerine gittiğinizde bu yok olur. Üç yıl kadar önce, bu malinitelerin yok olması için açıklama şuydu; "Tabii ileri derecede hastalar var, ilerlemiş romatoid artrit lenfomaya sebep olur, lenfomayı tedavi edersin, artrit ve lenfoma yok olur." Mantıklı ama sonra gülünç datalar da görmeye başlarız. Örneğin bir tanesi gösteriyor ki TNF inhibitörü lenfomayı tedavi ediyor, daha da kötüsü 5 yıl boyunca lenfomaya karşı koruyor. Yakın zamanda yeni bir data oluşturduk ve bize bunun neden böyle olduğu ile ilgili ipucu verecek. Bu demek değildir ki kayıt sistemlerinden uzak duralım, iyi kontrol gruplarıyla iyi registryler yapmak ve geliştirmek istiyoruz.

Biyolojik registry'lere 10 yıllık bakıştan sonra fark ettik ki dizaynları ile ilgili sorunlar var. Haklarında düşünmeye başladığımız için bu iyi bir şey olabilir.

Vedat Hamuryudan veritabanı diye adlandırdı ama biz kayıt kütüğü de diyoruz.

Neden kayıt kütükleri? Çünkü ilaçların nadir, beklenmeyen, uzun süren yan etkilerini tespit etmek istiyoruz. Yeni bir ilaç ruhsatlanmadan önce sadece 2000-3000 hasta ilacı kullanıyor. Eğer 2000-3000'den az olduğunda düşük yan etki eğilimi görüyorsak kayıt kütüğüne ihtiyacımız var. Tabii bu yan etkileri rapor eden spontan raporlama sistemleri de var ama biliyoruz ki ilaç kullanımındaki tüm yan etkileri ortaya çıkarmakta etkili değil. Ayrıca piyasadaki ilaçların farklı dozlarını, kombinasyonlarını ve ilaç değişim nedenlerini de bilmemiz gerekiyor. Biliyoruz ki randomize kontrollü çalışmaların dış geçerliliği düşük. Epidemiyolojik kohorta gereksinimimiz var çünkü spontan raporlama sistemleri yetersiz ve ilacın gerçek hayattaki performansını göstermiyor. Avrupa'da çok sayıda kayıt kütüğü var. Aşağıdakiler onların temel özellikleri:

- Ulusal romatoloji dernekleri ile başlatılmaları,
- İlaça özgü değil, tüm biyolojikleri kapsıyor olmaları,
- Epidemiyolojik kohort şeklinde olmaları,
- Kaybedilen hastanın biyolojik ilacı kesilse bile takip ediliyor.
- Hastaların büyük bölümü kaybediliyor.

Hemen hemen o ülkede, o ilacı kullanan herkesi dâhil ediyorlar. Prof. Yazıcı'nın söylediği gibi o ülkelerin çoğunun karşılaştırmalı kohortları var. Almanya'da karşılaştırmalı kohort kayıt kütüğünün bir parçası şeklinde. İngiltere'de özel merkezlerdeki hastalardan kütük dışı kohort, İsveç ve Danimarka'da ulusal kanser ve ölüm kayıtları ya da hastaneye yatırılma kayıtları gibi diğer veriler. Kayıt kütüklerinin bir diğer özelliği hepsinin ilaç firmalarının koşulsuz desteği sağlanarak hazırlanıyor olmaları. Tasarımı, yönetimi ve verilerin yayımlanması bağımsız ama yayının gönderilmesinden önce firmaların görme ve yorum yapma hakları var. İsviçre, Almanya, İngiltere gibi bazı

BİYOLOJİK KAYIT KÜTÜKLERİ

Ayhan Dinç

*Patio Özel Romatoloji
(Ankara)*

ülkelerde kayıt kütükleri 6 ayda bir ilaç firmaları-na, kayıt kütüğü ve karşılaştırmalı kohortlardaki yan etkiler ile ilgili bildirim yapıyor. Bu kısa bildiri Manchester Template olarak adlandırılıyor. Bu gerekli bir raporlama çünkü firmalar da EMEA'ya gerekli bilgiyi yolluyorlar. Danimarka'da daha sık, yılda 4 kez tüm gruplardaki tüm ciddi yan etkileri ilaç firmalarına bildiriyorlar. Tüm Avrupa'da çok sayıda kayıt kütüğü için yılda bir kez ortak toplandı yapılıyor. Birbirleriyle yakın ilişkide ortak çalışmalar yürütüyorlar. Ayrıca tüm dünyadan başka kayıt kütüklerini de davet ediyorlar. İngiltere'deki bir toplantıya ben de bir kez katıldım.

British Society of Rheumatology Biologics Register (BSRBR) Ekim 2001'de çalışmaya başlamış-tır ve Manchester ARC Epidemiyoloji ünitesinden 15 kişilik ekiple yürütülmektedir. UK'deki tüm biyolojik hastalar dâhil edilmiştir. UK'de yaklaşık %8 RA hastası biyolojik kullanmaktadır. NICE guideline'a göre ilaç başlayacak hastanın aktif hastalığı ve 2 DMARD duyarlılığı olmalıdır. Başlan-gıçta her ilaç için en az 4000 hasta ve 5 yıllık takip planladılar. Bu hesabı yaptılar çünkü bu rakam lenfoma insidansında 2 kat artışı belirleyebiliyor. 2002'de NICE Institute bu kaydı zorunlu hale getirdi. Sonra sayıyı arttırdılar. Şimdi 19.000 hastaları var, 16.000'i RA hastası. Sadece DMARD alan 3000 RA'lı hasta ile karşılaştırmalı kohortları var. ACR ve DAS28 klinik izlem verisi 6 ayda bir yollanıyor. Ayrıca ilk 3 yılda 6 ayda bir demografi HAQ hasta verisi hazırlanıyor. Ömür boyu malignite ve ölüm yönünden izlem yapıyorlar.

RABBIT Registry Mayıs 2001'de başladı. Berlin German Rheumatism Research Center'in Epidemiyoloji ünitesi tarafından hazırlanıyor. Altı biyolojik hastayı kapsıyor. İlginç bir de karşılaştırmalı kohortları var. Altı biyolojik ilaçtan birine veya öncesine yetersizlik nedeniyle DMARD başlayan hastalar karşılaştırmalı kohort olarak dâhil edilmiş. On yıl takip ediyorlar, kontrol hastaları da var. Bu çok önemli, kontroller aynı merkezlerden. Örneğin merkez 100 hasta yolluyor, hemen hemen aynı sayıda aynı merkezden kontrol hastası yol-luyor. 6000 kayıtları var, diğer ilaçlara son verip 2007'den beri sadece Rituksimab ve Abatasept verilerine devam ediyorlar. İlk yıl 3 ayda bir veri

ekliyorlar, sonra 6 ayda bir. Klinik data (ACR; DAS28) ve hasta verisi (Hannover functional score) dâhil ediliyor.

Swedish Biological Register; 1999'da başladı ve Swedish Rheumatology Association tarafından yürütülüyor. İlaça başlayan her hasta kayıt kütüğüne dâhil ediliyor. Raporlar her 3, 6 ve sonra da 12 ayda yollanıyor. Klinik data (ACR; DAS28) ve hasta verisi (Demografi, HAQ) rapor ediliyor. 2003'ten beri verilerine websitesi aracılığı ile giriş seçeneği sunuyorlar. İki ulusal kayıt kütüğü tarafından destekleniyor; Stockholm ve South Swedish registers. Başlangıçta sadece RA'lı hastalarken, şimdi tüm hastaları kapsıyor-lar. Şu anda 15.000 kadar hastaları var. Karşı-laştırma grubu diğer kayıt kütüklerinden biraz farklı, bazı yatan RA'lı hastaları da alıyorlar. Verilerini diğer ulusal kayıtlarla karşılaştırıyor-lar, ölüm, yatan hasta, ayaktan tedavi gören ve kanser kayıtları gibi.

Spanish Register; ilginçtir, 2000 yılında başladı ama çok fazla klinik veri yok. İlaça başlama, son-landırma ve tarihler, fakat takip sırasında klinik ya da hasta verisi yok. Hastalar sistemden yan etki bildirim formuna yönlendiriliyor. 2000 ve 2006 yıl-ları arasında merkez kısıtlaması yok sistemde sa-dece %50 hasta kayıtlı. İzlemiden sonra, 2010'da kayıt kütüğüne dâhil edilecek merkezleri azalt-maya karar verdiler ve takibin sıklığını arttırdılar. Yan etki bildirim raporlaması için WHO kriterlerini kullanıyorlar. Karşılaştırmalı kohort kullanıyorlar (EMECAR). Bu kayıt kütüğünden farklı bir kohort, kabaca bu kohortta 800 hastaları var.

Denmark Registry; 2000 yılında başladı. Danish Society of Rheumatology ve Institute for Rational Pharmacotherapy tarafından yürütülüyor. Kayıt kütüğünde 26 romatoloji merkezi var. 3500 RA, AS ve PSA hastaları. Hiçbir hasta bu kayıt kütüğün-de sadece klinik veri olarak yer almıyor. DMARD almış bazı erken RA hastalarını karşılaştırmalı kohort olarak alıyorlar. İki veri giriş sistemi kulla-nıyorlar. Bir online, bir kağıt, aynı zamanda diğer ulusal kayıtlara link ile bağlantılılar (ölüm, yatan hasta, kanser).

Norwegian DMARD Register (NOR-DMSRD); 2000 yılında 5 romatoloji merkeziyle başladı. Sadece biyolojiklerle sınırlı değil tüm hastaları dâhil ediyorlar.

Dutch Rheumatoid Arthritis Monitoring Register (DREAM); 2003 yılında başladı ve Radboud University Nijmegen Medical Center tarafından 11 merkez ve yaklaşık 100 hastayla yürütülüyor. Karşılaştırmalı kohort Nijmegen, erken RA inception, ama bu hastalar henüz DMARD almamış hastalar. Advers olaylar hastalarca bildiriliyor, advers olay kodlaması yok.

Swiss Clinical Quality Management Database; başlangıçta kayıt kütüğü değildi, kalite kontrol sistemiydi. Hastalar biyolojik başlamadan önce kayıt ediliyor. Standart izlem süreleri yok. Klinik, hasta kökenli verileri var.

Avusturya’da kayıt sisteminde sadece Viyana’daki 2 romatoloji kliniği var.

Başka kayıt sistemleri de var, Kanada’da Ontario System yeni başladı. ACR/DAS28 demography ve HAQ data içeriyor.

ABD’de çok sayıda sistem var ama hepsi kayıt sistemi değil, bazıları sadece ulusal veri bankalarındaki hasta verileri. CORRONA kayıt sisteminde veri, uzman ve hasta tarafından giriliyor. Yaklaşık 15.000 hasta var. Uzmanlar hasta bilgi ve bulgularını online görebilirler.

İngiliz, Alman, İspanyol kayıt kütüklerinde advers olay bildirimine baktığınızda MedRA kodlama sistemi kullanıyorlar. İsveç’de ICD-10 ve daha birçok kodlama sistemi kullanıyorlar. Aşağıdakiler biyolojik kayıt kütüklerinden veriler:

Günlük pratikte anti-TNF ajanların etkinliği

- İlaçta kalma oranları
- DAS28, ACR yanıt oranları ile etkinlik.
- İlaç değişimlerinin etkinliği
- Remisyon oranı
- Yanıtı kestiren faktörler
- Maliyet etkinlik analizleri

Güvenlik verileri;

- Tüberküloz
- Kardiyovasküler olaylar
- Lenfoma
- Solid tümörler
- Gebelik
- Ölüm

Vedat Hamuryudan’ın da söylediği gibi ‘channeling bias’ ve ‘confounding by indication’ da bazı sorunlar var:

- Kılavuzlar, biyolojik ilaçları ağır ve DMARD dirençli olgularda öneriyor, baştan kötü prognozlar.
- Enfeksiyon geçirenlerin biyolojikleri kesildikçe enfeksiyon oranı düşüyor.
- Time dependent risks

O zaman ne yapılmalı?

Karşılaştırmalı kohorta ihtiyacımız var. Bu kohorta; yaş, cinsiyet, hastalık süresi, hastalık aktivitesi ve şiddeti, tedavi öyküsü benzer olmalı. DMARD-IR gruba ihtiyacımız var.

Ayrıca takip sistemine de ihtiyacımız var. Kontrol grubu da izlemde aynı usulleri kullanmalı. Diğer kaynaklar ile doğrulama yapmalıyız. Hasta bildirimleri karşılaştırmada kullanılabilir. Veri analizi desteğine de ihtiyacımız var.

Türkiye’de şu anda Gözlemsel İlaç Çalışmaları Kılavuzu var. Bu guideline’ı kendi sistemimizi dizayn ederken kullanabiliriz. Bu kayıt sistemini en başında Sağlık Bakanlığına rapor etmeliyiz. Bu guideline’da diyor ki, ilave ilaç için gerekli onay baştan alındığı için ayrıca kayıt kütüğü onayına gerek yoktur.

Bir diğer user guide’da ABD’den. Bu tür bir guide hazırlamışlar çünkü biyolojikler için onlar da kayıt kütüğü hazırlamak istiyorlar. Kalın bir kitap çünkü yeni bir kayıt istemi için tüm detayları dâhil etmişler.

Size ilk soruyu ben sormak istiyorum, kayıt kütüğüne ihtiyacımız var mı?

Hasan Yazıcı: Bunu İstanbul’da çok tartıştık, bazı dataları toplamaya da başladık ama takım çalışması da gerekiyor. Lenfoma vs.’ye bir bakış açısı, British registry diyor ki “Anti-TNF kullananlarda lenfoma sıklığının iki kez arttığını onaylıyoruz.” Bu kabul edilemez. Çünkü romatoid artritte anti-TNF ile ya da anti-TNF’siz lenfomada 2 kat artış olur. Bu 30 yıllık bilgi birikimi. Yani evet, soru bu ilaçlar azaltıyor mu artırıyor mu, bunu bilmiyoruz. Bir diğer yol, lenfomalı hastalara bak. Lenfoma ya da akciğer kanseri kayıt kütüğüne bak sonra anti-TNF ya da benzeri ilaçları kullanmış ya da kullanmamış olanlarla matrak geç. Bu etkiye dair bir data var mı? Bu eski bir soru, östrojenin vulva kanserine neden olduğu, bu 35 yıl önce çok tartışılıyordu, 40 hastanın çocuklarıyla sorunun cevabını bulduk, evet, neden oluyor. Belki biraz para harcayıp hematoloji ve onkologlarla birlikte çalışabiliriz. Cerrahpaşa’da yapmaya başladık ama büyütülebilirdi, tüm Türkiye’de, o gruptaki RA hastaları alınabilir. Sedat hali hazırda bazı hemotolog ve onkologlarla iletişime geçti. Biz bunu onun izniyle daha büyük bir proje haline getirebiliriz. Belçika da bize katılır mı?

Emmanuel Lesaffre: Ben de tam bu konunun istatistiğin çok ayrı bir bölümü olduğunu düşünüyordum. Bazen çalışma arkadaşlarım bana danışmak için bir soru sorar; “Hayır, bu hayata dair bir soru” derim. İstatistikçiye sorma nedeni problemin iç yüzüne girmek, çünkü bu epidemiyoloji ve istatistik kombinasyonu. Bu problemi çözmek için bir grup istatistikçi olması lazım sadece bir tane değil. Bu kolay değil, daha önce çözümlenmiş insanlar da olabilir. Genç bir istatistikçi yerine ortak çabayla çözümü bulmalısın.

Hasan Yazıcı: Emmanuel seni 2004 yılında FDA’nın yaptığı bir dataya götürebilirim. Arthritis&Rheumatism’de de yayımlandı. Anti TNF kullanan ilk 26.000 hastada erken lenfoma FDA’dan rapor edildi. Editöre mektup yazdım. Yayımlandı. Şöyle diyordu, “Oyunun başında olanlarla sonunda olanları karşılaştıramazsın”. Nisan 2004’te FDA toplantısı vardı, birisi bu çalışmanın editörüne sordu, “Oluş sıklığında artış yok, zamanla ilişkisi

var mı?” Cevap pervasızca; “Hayır biz saptamadık,” oldu. Yani bu sadece insanların görüşlerini saklamaları değil, en iyi dergilerimizde yer alan tartışmalarda bu hiç tartışıldı mı?

Sadece Suissa gibi birkaç kişi bunu yaptı, “Nasıl olur da kontrollü klinik çalışmalarda var ama kayıt sisteminde olmaz?” dedi. Bu çok basit bir soruydu ama cevaplanmadı. Az önce söylediğim 40 yıl önceki vulva kanserinde de aynı şey oldu. Mayo klinikte çok önemli çalışmalar vardı, müzakere şu konuydu; hamileyken östrojen kullanmış kadınların çocukları vulva kanseri oluyorlar. Çalıştıkları 20.000 hastada endikasyona rastlanmadı. Boston’dan birisi 40 hasta ve çocukları aldı, 20 vulva kanserli, 20’si değil. Ki kare 8,5 gibi bir şeydi. Gayet basit. Korkum şu, başka şeylerin de karıştığı başka basit ilişkilerle de uğraşmak zorunda kalabiliriz.

Emmanuel Lesaffre: Demek istediğim kayıt sistemlerini benzer küçük karşılaştırmalı çalışmalarla birlikte araştır. Çünkü herkes data toplamak için çok fazla efor sarf ediyor, sonra da bir şey bulamıyor. Peki, kayıt sistemlerinin bu küçük çalışmalarla aynı etkiyi yapması için ne yapman gerekiyor, en azından yapmak istediğin. Bu çok karışık ve dikkat çekmek istediğim de buydu. Bazı izole vakaların küçük vaka kontrol çalışmalarıyla ya da küçük kohortlarla çözüleceğini düşünüyorum, böyleyken tüm bu eforu hele de küçük çalışmaya üstün gelemeyeceksen, kayıt sistemine harcamak yazık olur.

Hangi hastalar biyolojik kayıt kütüğüne dâhil edilmelidir?

1. Tanı koyulan her hasta
2. DMARD başlanan her hasta
3. Biyolojik başlanan her hasta
4. 2 ve 3 hastalarından bir örneklem

Biyolojik kayıt kütüğünün kapsamı nasıl olmalıdır?

1. Tüm romatoloji hekimleri
2. Belli romatoloji merkezleri
3. Tüm biyolojik kullanan hekimler (FTR dahil)

RA'da klinik veriler hangileri olmalı?

1. DAS 28 durumu
2. ACR 50 yanıtı
3. Şiş eklem sayısı
4. Doktor global

RA'da hasta bildirimleri nasıl olmalı?

1. HAQ
2. RAPID 3 veya 5
3. Hasta global
4. Yaşam kalitesi örgütleri

Hasan Yazıcı: Sen devam etmeden önce arkada oturan Merck ilaç firması temsilcisine birkaç sorum var. Didem, eğer sana, “Tüm Türkiye’de 1 yıl boyunca Remicade reçete edilmiş hastaların sayısını belirlemek için bir yolun var mıdır?” diye sorsam.

Didem Gümüş: Hayır fikrimiz yok.

Hasan Yazıcı: Uzmanları takip ettiğinizi biliyoruz. Reçetelerin üzerinde de isimleri yazıyor. Önerim şu, ilgileri farklı birçok rakip merkeze soracağımıza, size “Neden bize örnek vermiyorsunuz?” diye sorsak. Tüm reçetelerin 10’da birini, Türkiye’de ya da İstanbul’da. İyi tanımlanmış grup. Sonra o bireyleri Romatoloji Derneği olarak takip ederiz. Ciddiyim. Aynı şeyi Abbott, Pfizer ve diğer biyolojik firmalara da söyledim. Eğer gerçek data istiyorsanız ona yaklaşmak için başka bir yol bu. Soru şu, gerçekten bilmiyor musunuz, yoksa vermek mi istemiyorsunuz?

Ayhan Dinç: Bence Sağlık Bakanlığı firmaların hasta bilgilerini bize vermelerine izin vermez.

Hasan Yazıcı: Sorumluluğu üstlenerek Sağlık Bakanlığına, ilaçların kötü yan etkileri ile ilgili araştırma yaptığımızı söyleriz. Sağlık Bakanlığına gitmek de başka bir fikir olabilir. İlaç firmasına gideceğine, Sağlık Bakanlığına git. Onlar formları nasıl olsa alıyorlar ve muhtemelen kışın yakıyorlardır. Ciddiyim. İyi fikir, ilaç firmasından isteyeceğine onlardan iste.

Bir Katılımcı: Şeffaflık yasasına göre bakanlığa bize datayı vermelerini söyleyebiliriz.

Hasan Yazıcı: Önce etik komiteye mi gidiyor? Yoksa doğrudan Sağlık Bakanlığına mı rapor ediliyor?

Bir Katılımcı: Ayhan gösterdi, doğrudan bakanlığa gidiyor.

Hasan Yazıcı: Lokal etik komiteyi pas mı geçiyor?

Bir Katılımcı: Muhtemelen bakanlık etik komitenin onayını sorar, lokal olan zaten sorun değil, asıl olan büyük sorun. Sağlık Bakanlığı yeni bir etik komite hazırlama girişiminde.

Ayhan Dinç: Güvenlik izlem formlarındaki bilgi çok güvenilir değil. Kimse okumadığı için.

Banu Çakır: Bir görüş bildirmek istiyorum, bir de yorum. Önce bu etik olarak doğru değil. Sağlık Bakanlığından ya da ilaç firmasından almak, her ikisi de etik değil. Eğer hasta sensen datanı kimseyle paylaşmak istemeyebilirsin. Faydalı bir şey için olsa bile. Hasta herhangi bir yaklaşıma dâhil olmak istemeyebilir. Bu etik olarak doğru değil. Datayı bir şey için alabilir, sonra başka bir yerde paylaşabilirsin. Bu ilacın reçete edildiği hastaların sadece adını ve adresini almak, zaten ilacı reçete eden uzmana vermek. İlaç firmasından ya da bakanlıktan datayı almak yerine, kimler bu ilacı yazıyor biliyorsun, tek tek her hastaya ulaşabilirsiniz. Dernek aracılığı ile ulaşabilirsiniz.

Koray Taşçılar: Gayet güzel detaylı reçete raporlama sistemleri var. Zamanlama, sayı. Sosyal güvenliklikten reçeteleme bilgilerini alabilirsiniz.

Banu Çakır: Hayır bunu veremezler, etik değil. Yasal olarak da doğru değil.

Koray Taşçılar: Legal olarak değil ama sample sorabilirsin. Aynı zamanda reçeteyi yazan doktorun da bilgilerine sahipler.

Banu Çakır: Kohort çalışmalarında tüm bu genellemelere ihtiyacın yok. Eğer amaç sadece hastayı bulmak ve takip etmekse, bunun Tüm Türkiye’de

ilaç yazan hekimler olması da gerekmiyor. Şu sa-
londaki hekimler, biz bu çalışmada varız, bundan
sonra olmak isteriz diyorsanız, o kohortu takip
etmekte hiçbir sakınca yok.

Ayhan Dinç: O registry olmaz ama.

Banu Çakır: Ama şu önerilen şey de registry değil.
Registry'ye alternatif değil ama kolay bir yöntem.
Ben hasta gözüyle bakınca da bunu çok etik bul-
muyorum. Siz bir doktora gidiyorsunuz, bir ilaç
alıyorsunuz bu reçete bir şekilde ilaç şirketine
gidiyor ya da Sağlık Bakanlığına gidiyor. Sonra
sizi biri arıyor diyor ki ben sizi çalışmaya almak
istiyorum. Benim datam nasıl ortada dolanıyor,
bunu hastanın anlaması çok olası değil.

Ayhan Dinç: Baştan hastanın onayı alınacak.

Banu Çakır: Hasan Bey'in önerisi için fikrimi söy-
lüyorum.

Bir Katılımcı: Hocam web tabanlı mı bu uygulama?

Ayhan Dinç: Evet, web tabanlı, şu an webdeyiz.

Bir Katılımcı: Peki yedekleme yaptığınız kaç mer-
keziniz var. Ve de lokasyonları?

Ayhan Dinç: 2, Türkiye ve Kanada'da.

Bu aslında bir registry değil. Bu, bizim 3-4
sene önce tamamen elektronik hasta takibi için alt
yapı olarak kurduğumuz bir sistem. Yoksa registry
olabilmesi için söylediğim şartların olması lazım.
Amacı olması lazım, ne için kuruyorsunuz bunu,
kimler katılacak, bilgiler toplanacak vs. olması
lazım.

Bir hasta girdiğiniz zaman şöyle bir sayfa açı-
lıyor, adı, soyadı, cinsiyeti, doğum yeri, mesleği,
adresi, e-mail adresi. Daha sonra aile öyküsünü
girebilirsiniz, arttırılabiliyor, istediğiniz aile öykü-
sünü ekleyebilirsiniz. Diyelim ki diyabet girmek
istediğiniz zaman, babasında var diye girebilir,
not ekleyebilirsiniz. Ya da tüberkülozla ilgili bir
veri girmek istiyorsunuz diyelim ki eşinde tüber-
küloz olmuş, onu da not alabilirsiniz. Daha önce

kullandığı ya da devam ettiği ilaçları ekleyebilirsi-
niz. Komorbiditeleri varsa onları yazabilirsiniz. Tik
atmanız yeterli, ya da ameliyat ekleyebilirsiniz.
Alışkanlıklarını ekleyebilirsiniz. Aşılarını ekleyebi-
lirsiniz. Ama bu, dediğim gibi tamamen takip için
yapılmış, bu kadar bilgiyi isterseniz girer, iste-
mezseniz girmezsiniz.

Bir hasta kartı ilave edeceğimiz zaman, yeni bir
kişi giriyorsunuz, bunu başka bir kişi de girebilir.
Sizin adınıza, şifre verdiğiniz kişi girebilir. Onun
yetkilerini tanımlayabilirsiniz. İlaç silmesin, data
silmesin vs. Bu sizin elinizde olan bir şey. Kimlik
no girebilirsiniz. Girilsin mi girilmesin mi diye
tartışılabilir. Şimdi herkesin kimlik numarası bir
şekilde tespit edilebiliyor.

Hasan Yazıcı: Bunun mekaniğine benim hiçbir iti-
razım yok. Benim bütün çabam, diyorum ki bunu,
kontrollü random semptom için olabilir. Kontrol
grubu için. Bunu herkesin kullanabileceği tek bir
şey yapabiliriz. İsmail ve Yusuf'un kullandıkları
var bir tane. Bunların hepsini bir yerde, belki bu-
nun içine girebilir, veya bu ona göre adapte edile-
bilir, onlar da devam ederler öbürleri de.

Vedat Hamuryudan: Yaşar'dan bir şey öğrendim, o
birçok şeyi değiştirecek galiba.

Yaşar Karaaslan: SGK elektronik rapor diye bir
uygulamaya geçiyor. Diyabette ilk örneğini hazır-
ladılar. Diyabetik her hastayı, elektronik ortamda
hastanın kilosu vs giriliyor, bir zorunluluk. Diyor ki
diyet tedavisi aldı mı, şu mu, bu mu? Ama acayip
sıkıntılı bir şey. Ben kendim diyabet raporu çıkara-
cağım zaman, bitiremedim. Diyor ki hasta nerede
yaşıyor, o kadar uzun ki. Bir Glucophage yazacak-
sın. Bu çok saçma derken akşam bana e-posta
geldi. "Biz biyolojikler başlamak için böyle bir şey
istiyoruz, sizin de dâhil olduğunuz birkaç kişiden
bunu isteyeceğiz." Bilmiyorum başka kimden
istendi. İşte böyle bir form hazırlayın dediler. Ben
de hazırlayıp arkadaşlara e-posta ile gönderdim.
Herkes çok uzun dedi. Ondan sonra şu andaki
veriler olan adı, soyadı, yaşı, daha önce kullandığı
DMARD'lar ve RF pozitif mi negatif mi vs gibi bir-
kaç bir şey yazdım, sonra onu gönderdim, uzun bir
süre bir ses çıkmadı, geçenlerde bir daha aradılar.

Dediler ki sizin form çok basit olmuş. Bir de fizik tedaviden birine göndermişler, o da “A, böyle bir şey olur mu, içine fizik tedavi alındı mı, şu verildi mi, bu verildi mi?” demiş, o da ilaveler yapmış. Sonunda bir form geldi, ben bu form çok lüzumsuz, çok uzun, bu konuda dernekten görüş alın, dedim. Zaten kendim doldurmak istemiyorum diye sizin e-posta adreslerinizi yazıp verdim.

Hasan Yazıcı: O bürokratik olarak doldurulması zorunlu bir form gibi. İyi bir form. Lafımız geçer geçmez, ilaç yazmak için o formu doldurma zorunluluğu olabilir. Olmasa daha iyi. Ama ideal dünyada o formu bizim formla birleştirmek. Olmadı, gene biz kendimiz düşündüğümüz ideale yakın bir form geliştirebiliriz. Onun yanında da öbür form.

Yaşar Karaaslan: Hocam eğer SGK o formu zorunlu kılacaksa, oraya girdiğimiz formu biz bir yere kopyalayarak hepsini bir yere koyduğumuzda, demografik bilgilerini komorbid hastalıklarını, aktiviteyle ilgili bazı şeyler, ppd, akciğer grafisi gibi, şu anda zaten yazdığımız bilgileri elektronik olarak kopyalayıp o formu kullansak.

Ayhan Dinç: Orada şöyle bir sorun var yalnız, hasta tedaviye devam etsin diye aktiviteyi biz olduğundan çok yazıyoruz. Orada gerçek aktivite mi değil mi tartışması olacak bu sefer.

Hasan Yazıcı: Politikacılar seçim öncesi sevimsiz görünmek istemezler. Bir şeyler yapar gibi görünüp bir şey yapmayacaklardır eminim. Murat’ın bir formu var, etrafta dolaşıyor, İsmail ve Yusuf’un var, Ayhan’ın bu sistemi var. Başka form hazırlamış olan var mı? Benim bildiğim bu üçü var. Fizik tedavinin ayrıca var mı? İlaç firmaları “Fizik tedavi bunu yapıyor, biliyor musunuz” dedikleri form nedir Feride, onda kaç soru var?

Feride Göğüş: Onun içeriğini çok net hatırlayamayacağım. Demografi var, eklem, CRP, ankilozan spondilit vs, böyle 5 yıllık bir takip.

Hasan Yazıcı: Murat, İsmail, Ayhan bir de fizik tedaviden sen, bir araya gelin, ortak bir form hazırlamaya çalışın, ondan sonra bitirelim işi. Öyle değil mi Ayhan?

Ayhan Dinç: Olur hocam. Orada iki konu var. Birisi bu bilgiler, diğeri de bu bilgiler nasıl girilecek? Esas problem bu. Onu tartışmamız lazım. Bu bilgiler ayrıntılı oldukça bunun girilmesi çok vakit alıyor, o yüzden bunun için İngiltere’de olduğu gibi bir ekip kurulması lazım.

Hasan Yazıcı: Tamam ona para veriz. Endüstri de yapılmasını istiyor, endüstri de yardım edebilir öyle bir şeye, düzgün yapılırsa.

Ayhan Dinç: Bir de bu girilen bilginin, mesela bazı veri tabanlarında olduğu gibi hekimin de kullanımına sunulması güzel olabilir. Hekim de kendi hastalarının özellikleri neymiş, hangi tedaviyi başlatmışım, diye görebilir.

Hasan Yazıcı: Formun bir bölümünü kendi hastasına kullanır, bir bölümü de o istediğimiz registry olur.

Ayhan Dinç: Yani kendi hastalarını da görebilsin sonradan. Ama hastaların bilgilerine bilgi ilave etmemiz lazım. Zaten bizim burada uğraştığımız da oydu. Mesela başka bir yerde kullanılabilir mi dediğimiz şey güvenlik izlem formuydu.

Bir Katılımcı: Burada amaç güvenlikle ilgili bilgileri mi temelde izlemek? Hastalığın aktivitesini mi izlemek, yoksa genel bir şey mi yapmak? Kişisel görüşümü söyleyeyim, eğer hastalığın bütün prognozunu art tedavileri artı güvenliği izlediğimizde oluşacak olan şema, çok fazla veri içeren, gerçekten aslında Ayhan’ınkinin en geniş cover ettiği bir sistem. Ama kendi merkezimizden biliyorum, bu ciddi bir emek, en önemli sorun, fellowlar bu işi yapacak olan kişiler doğal olarak kendi araştırmaları kendi konuları çerçevesinde bu işi gönüllü yapmıyorlar.

Hasan Yazıcı: Onun için de diyoruz ki bu modüllerin bazıları için bir sample olabilir. 1’e 10 sample, 1’e 5 dedik. Türkiye bazında yapılırsa, Türkiye’nin romatoid artritini New York’la kıyaslamakla ama New York istiyorsa Cerrahpaşa’yla Marmara’yla kıyaslamak, onu da ayrı kıyaslayabilir. Bunun hakkında böyle her şeyi konuşuyoruz. Modülün içinde herkesin ayrı şifreleri ola-

bilir, ayrı şifreyle değişik modüllere girilir. Öyle değil mi Ayhan?

Ayhan Dinç: Peki, ortak görüş, bakanlığı da ikna ederek, onunla birlikte kullanılacak bir şey haline getirmek...

Bir Katılımcı: Hocam aslında şöyle bir teknoloji var. Mobil, küçük cihazlarla, hatta ekrandan elle girilecek şekilde, ilaçların da barkodunu tanıyabilen sistemler var. Belki hekim ya da hasta tıbbi sekreter olmadan da sağlık çalışanı herhangi birinin girebileceği gibi basit hale getirilebilecek bir sistem aslında.

Hasan Yazıcı: En fazla anti-TNF yazan biziz, en çok hasta bizde var ama hepsini birden tutmak bizi paralyze eder.

Ayhan Dinç: Mümkün değil o. Aynı anda hem klinik işi hem bunu girmesi mümkün değil.

Hasan Yazıcı: Onun için diyorum ki bir kısmını randomize edelim.

Ayhan Dinç: Mesela YusuP'un sisteminde her visite data girilmesi gerekiyor. Registry'de eğer 3-6 ay tespit edilirse o zaman farklı olacak o.

Feride Göğüş: Ben etkinlik değerlendirmek için bu registry'nin güvenilir bir şey olduğunu düşünmüyorum. Ancak güvenilirliği değerlendirebilirsiniz.

Ayhan Dinç: Gerekli de değil zaten.

Hasan Yazıcı: Neden değerlendirilemesin?

Gülen Hatemi: Mesela bizim o polikliniğe koyduğumuz kioskadaki sorunlardan bile tekrarlanabilirliği kötü. Umursamadıklarından.

Hasan Yazıcı: Birisinin işi sahiplenip götürmesi lazım.

Gülen Hatemi: 120 hasta görüyorsun Perşembe günü. Kaçı romatoid artritli, kaç lupuslu var vs, her seferinde 5 tanesi düzgün girilse, 2 Marmara'dan girilse, 3 başka yerden girilse 6 ayın sonunda muazzam etkinlik de var, güvenlik de var.

Ayhan Dinç: Bir de şunu söylemek lazım, merkezde birileri data girebilir. Benim bildiğim kadarıyla fizik tedavinin registry'sinde bir şirket vardı datayı giren. Çok pahalıya geldi diye onu online'a geçirmeye çalışıyorlar. Buna benzer bir sisteme geçelim diye. Ya da şöyle bir organizasyon yapabiliriz. Üniversitelerde bir kısım insanları eğitip dataları onların girmesini sağlayabiliriz.

Hasan Yazıcı: O zaman pilot çalışma yapalım. İsteyenler katılsın, 10'a bir, 5'e bir sample alalım. İki ay görelim. Kaç tane hasta girebiliyoruz yardımla. Dönelim endüstriye yardım et diyelim. Bakalım ne oluyor.

Bir Katılımcı: Nereye gireceğiz? Nerede toplayacağız?

Hasan Yazıcı: Tek bir formda toplayacağız. O formda topladıktan sonra 10'a bir veya 5'e bir sample alacağız. Cerrahpaşa'da hemşire veya yardımcı görevlendireceğiz. Hiç bulamazsak yardım isteyeceğiz endüstriden. Hiç olmazsa bir ay nasıl gittiğini görelim.

JAMA’da yeni yayımlanan bir çalışmadan söz edeceğim, başlığı “Yaşlı Kadınlarda Bifosfonat Kullanımı ve Subtrokanterik Bölge ya da Femur Gövdesi Kırıkları” (LY Park et al. JAMA 2011). Kişiler arasında kırık riskini karşılaştırma amacıyla yapılmış bir veritabanı çalışması. Elli yaş üstü kadınların yarısının geri kalan yaşamlarında osteoporozla bağlı bir kırık geçireceğini ve kırık geçiren her 5 kişiden birinin de bir yıl içinde öldüğünü düşünelim. Osteoporozu olanlara bifosfonat tedavi vermekteki amacımız işte bu kırıkları ve dolayısıyla ölümleri engellemektir. Son zamanlarda bifosfonat ilaçlarının kemik yapısını değiştirdiğini ve bunun sonucu basit travmalarda femurun subtrokanterik bölgesinde kırıklar oluşmasına sebep olabildiğini öne süren çalışmalar yayımlandı. Bu makalenin yazarları, Ontario’da yaşayan ve bifosfonat grubu ilaç kullanan 68 yaş ve üstü kadınların bulunduğu bir kohortta iç-içe vaka-kontrol çalışma yöntemi kullanarak (nested case control) bifosfonat kullanımı ile subtrokanterik bölge kırıkları arasındaki ilişkiyi araştırmış. Çalışma 6 yıllık bir dönemi kapsıyor.

Bir vaka-kontrol çalışmasının “iç-içe” olması, “yuvalanmış” olması ne anlama gelir? İç-içe vaka-kontrol çalışması, örneklemenin bir kohort içinde yapılmasıyla olur. Takip altında bir kohortun özellikleri de iyi tanımlandığı için bu yöntem vaka bulma kolaylığı sağlar, ayrıca çalışma için ihtiyaç duyulabilecek başka verilere de ulaşılmasını kolaylaştırabilir. Nitekim bu çalışmada da böyle oluyor.

Kohortla iç içe vaka-kontrol analizi, klasik vaka-kontrol analizinden farklıdır, konvansiyonel kohort yöntemi kadar güvenilir bir epidemiyoloji yöntemidir ve veri analizini yapmak kohort analizinden daha kolaydır. Bunu sağlayan başlıca özellik ise kontrollerin seçilme yöntemidir. Vakaların kohorta girişiyle vaka özelliği kazandığı zaman arasında geçen takip süresi hesaplanır ve kontrollerin her biri eşleşeceği vaka ile en az aynı süre boyunca takip edilmiş kohort üyeleri arasından raslantısal olarak seçilirse kohorttaki vaka insidansı hesaba katılmış olur ve koşullu lojistik regresyon yöntemi ile kaç kaç oranı (odds ratio) hesaplandığında bu oran konvansiyonel

DERGİ KULÜBÜ

Koray Taşçılar

*Romatoloji Bilim Dalı
İstanbul Üniversitesi
Cerrahpaşa Tıp Fakültesi*

kohort çalışmalarında Cox regresyonu ile hesaplanan tehlike oranına (hazard ratio) ya da klasik insidans hızı oranına (rate ratio, incidence ratio) eşdeğer olur. Bu örnekleme yöntemine “risk seti örnekleme” ya da “insidans yoğunluğu örnekleme” adı verilir. Çalışmada bu isim belirtilmiyor fakat yazarların vaka-kontrol seçimiyle ilgili detaylarda tarif ettiği yöntem bu yöntemdir.

Yazarlar Ontario’da birkaç veritabanını birbirine bağlayarak hastaları buluyor. “Ontario Public Program Database” adlı veritabanında tutulan reçete kayıtları kullanılarak bifosfonat kullanan hastalar tespit edilmiş. Eyalette hastaneye yatan herkesin epikrizlerini kaydeden “The Canadian Institute for Health Information Discharge Abstract Database” adlı veritabanı ise bifosfonat alan kişiler arasında ilaç kullanmaya başladıktan sonra subtrokanterik kırık geçiren kişileri tespit etmek için kullanılmış. “The Health Insurance Plan Database” adlı veritabanı kullanılarak da kohorta giren kişilerin doktor kayıtları kullanılarak genel sağlıkları hakkında bilgi toplanmış. Bunlara ek olarak gözlenen kırıklardan bazılarının kansere bağlı patolojik kırık olabileceği de düşünülerek Ontario kanser veritabanı ile bağlantı kurulmuş ve kohort üyeleri arasında kanseri olanlar tespit edilmiş. Eyaletin nüfus veritabanına göre çalışmanın yapıldığı bölgenin genel nüfusu tanımlanmış. Ayrıca Statistics Canada nüfus sayımı bilgileriyle eşleştirme yapılarak çalışılan grubun gelir düzeyi hakkında veri toplanmış.

Yazarlar Ontario’da yaşayan 68 yaş ve üstündeki kadınlar arasında bifosfonat reçete edilenleri çalışmaya dâhil etmiş. Bunlar arasında çalışma başlangıcından önceki son 10 yılda kanser olanları, kemik yapısı değişmiş olanları; yani önceden osteomalazi, osteoporoz, hiperparatiroidi, hiperkalsemi, epilepsi, Çölyak hastalığı, Paget hastalığı, renal osteodistrofi ve gastrik baypas ameliyatı olanları çalışma dışı bırakmış. Ayrıca çalışma başlangıcından önceki 5 yıl içinde raloksifen, kalsitonin, sodyum florür, klodronat, pamidronat ve zoledronat alanlar da çalışma dışı bırakılmış.

Bu bir vaka-kontrol çalışması; yazarlar önce vakaları belirlemiş. Hastane kayıtlarında ICD-10 tanı kodlarına bakılarak 2003-2008 yılları arasında

femurun subtrokanterik bölgesinde ya da gövdesinde kırığı olanlar belirlenmiş. Kırık sebebi olarak travma, trafik kazası ve düşme belirtilen kişiler çalışma dışı bırakılmış. Kalça kırığı olup sadece kalça protezi ameliyatı yapılarak tedavi edilenler de çalışmaya dâhil edilmemiş. Bunu niye yaptıklarını bilmiyoruz çalışmada sebebi belirtilmiyor. Her vaka için aynı kohort içinden subtrokanterik bölge ya da femur gövde kırığı sebebiyle hastaneye yatmamış kişiler arasından her vaka için beşer kontrol seçilmiş. Kontroller yaş ve kohort çalışmasına giriş yılına göre eşleştirilmiş.

Reçete kayıtlarına bakılarak tüm vaka ve kontrollerin bifosfonat kullanma süreleri de tespit edilmiş ve 3 aydan kısa, 3 ay-3 yıl arası, 3-5 yıl arası ve 5 yıldan daha uzun süre şeklinde sınıflanmış. Yani vaka ve kontrollerin her birinin ne kadar süre bifosfonat aldığı kaydedilmiş.

Bir veritabanından veri alınırken bilmek isteyeceğimiz şeylerden biri de verinin geçerliğidir. Yani her bir kişi için elde ettiğimiz bilginin gerçek durumu ne olasılıkla gösterdiğini bilmek isteriz. Bu amaçla yazarlar ayrı bir validasyon çalışması yapmış ve kullandıkları tanı kodlarının duyarlılığını ve pozitif prediktif değerini belirlemiş.

Yazarlar primer analizde subtrokanterik femur kırığı sebebiyle hastaneye yatmak ve kümülatif bifosfonat kullanım süresinin 5 yıl üzerinde olması arasındaki ilişkiyi değerlendirmiş. Bu amaçla koşullu lojistik regresyon analizi ile kaçakaç oranları ve bunların %95 güven aralıkları hesaplanmış. Analizde düzeltmeler yapılırken sonuçla mantıksal ilişkisi olmayan çok sayıda değişken sebebiyle sonucun yorumlanabilirliğinin bozulmaması (overfitting) için, ilaç ve sonuç arasındaki ilişkiyi alternatif olarak açıklayabilecek ilaçlar ve hastalıklar seçilirken endikasyonlara ve etki mekanizmalarına göre gruplandırılmış.

Burada asistanlarımız da olduğu için bu çalışmada kullanılan yöntemleri kısaca açıklamakta fayda var.

Regresyon analizi, gözlemler üzerinden oluşturduğumuz ilişkileri özetleyecek bir grafiğin

matematik formülünü bulmaya yarar. Sonuç verimiz lineer ise (bir ölçümün sayısal sonucu, örneğin kolesterol değeri, gibi) analiz yöntemi lineer regresyon, eğer sonuç değişkeni bir olayın olma ya da olmama olasılığı (dikotom) ise analiz yöntemi lojistik regresyon olarak adlandırılır. Regresyondan söz ederken bir regresyondaki uyumun iyiliğinden (goodness of fit), yani açıklayıcı verilerdeki değişimin sonuçtaki değişimi ne kadar iyi tahmin ettiğinden de söz ederiz. Bu da analiz yaparak bulduğumuz formüle açıklayıcı sebepleri koyarak formülden gelen sonuçların gerçekte gözlediğimiz sonuçlarla ne kadar farklılık gösterdiğine bakılarak anlaşılır. Eğer bir regresyon analizi yapılırken sonuçla sayısal olarak sıkı ilişkisi olan ancak aradığımız sebep sonuç ilişkisi açısından mantıksal anlamı olmayan çok sayıda veri kullanılırsa, araştırdığımız sebep-sonuç ilişkisini, fazla sayıda açıklayıcı veri kullanıldığı için, regresyon analizi ile görmek mümkün olmayabilir. Bu duruma “overfitting” adı verilir.

Bu çalışmada bifosfonat kullanımı ile subtrokanterik kırık olasılığı arasındaki ilişki belirlenirken analizde bazı demografik bilgiler, kullanılan ek ilaçların sayısı ve cinsi ve bazı ek hastalıklara göre düzeltmeler yapılmış. Ek olarak sağlık hizmeti kullanma sıklığına göre de düzeltme yapılarak çalışılan grubun genel sağlık durumu da göz önünde tutulmaya çalışılmış. Son olarak kemik yoğunluğu ölçümlerinin ne sıklıkta yapıldığına göre de sonuçlar düzeltilmiş.

Bunlara ek olarak çalışmada bazı ikincil analizler de yapılmış. Aynı veri tabanı kullanılarak bir başka iç-içe vaka-kontrol analizi ile bifosfonat kullanımı ile osteoporozla bağlı kırık riski arasındaki ilişki değerlendirilmiş. Yani yazarlar kurdukları sebep-sonuç ilişkisi modelinin aradıkları sebep sonuç ilişkisini (bifosfonat-subtrokanterik kırık riski) açıklamakta başarılı olup olmadığına karar vermek amacıyla aynı veritabanında aynı modeli kullanarak gerçekten var olduğu bilinen bir sebep sonuç ilişkisinin (bifosfonat – osteoporozla bağlı kırık riski) gösterilip gösterilemediğini anlamaya çalışmışlar. Buna ek olarak sebebe atfedilebilir kırık riski de (attributable risk) hesaplanmış. Atfedilebilir risk, olay riski oluşturan sebebi tamamen

ortadan kaldırdığınızda o sebeple ilişkilendirilmiş sonucu tam olarak ne kadar azaltabileceğinizi gösterir.

Atfedilebilir riski hesaplamak için basitçe kaç-kaç oranını birden çıkarıp kaçakaç oranına böleriz. Bu bulduğumuz atfedilebilir riski sebebe maruz kalan vaka sayısının oranıyla çarparak da toplumdaki atfedilebilir risk hesaplanmış olur.

Emmanuel Lesaffre: Ben atfedilebilir risk kavramını hiç kullanan biri değilim ama formülde kaçakaç oranı yerine aslen rölatif risk kullanılması gerekmiyor mu?

Koray Taşçılar: Hayır, bunu kaçakaç oranından da hesaplayabilirsiniz, aslında atfedilebilir risk kavramı değişik yerlerde değişik tanımlanıyor. Ama genel tanım şu: Bir risk faktörünü ortadan kaldırdığınızda ona bağlı sonuçtaki azalma miktarı.

Emmanuel Lesaffre: Ama bu kaçakaç oranının da tanımı.

Koray Taşçılar: Aslında atfedilebilir risk ve atfedilebilir risk kesiri (attributable risk fraction) kohort çalışmalarından gelen bir kavram çünkü önce bir riske ihtiyacınız var ki atfedilebilir riski ve atfedilebilir risk kesirini hesaplayabilesiniz. Bu kohortla iç-içe bir vaka-kontrol çalışması olduğu için burada risk oranı hesaplama şansı var, fakat çalışmayı yapanlar koşullu lojistik regresyon analizi kullanarak risk yerine kaçakaç oranları buluyorlar. Fakat yine tasarım itibarıyla bu bir iç-içe vaka-kontrol çalışması olduğu için lojistik regresyonda bulduğumuz kaçakaç oranları aslında bildiğimiz konvansiyonel kaçakaç oranı değil kaçakaç-risk oranlarıdır, böyle düşünmüş olmalılar. Yazıda bunu belirtmiyorlar fakat lojistik regresyonda buldukları kaçakaç oranlarına risk oranı muamelesi yapmışlar ve atfedilebilir risk yerine atfedilebilir risk kesirini hesaplamışlar. Ben bu çalışmadaki orijinal insidans hızlarından bir takım atfedilebilir risk kesirleri hesapladım. Bu sadece böyle bir çalışmada mümkün çünkü bu kohortla iç-içe vaka-kontrol çalışması, aksi takdirde aşına olduğumuz bir şey değil.

Kaçakaç sayılarınız düşük olduğu zaman elde ettiğiniz sayılar da riske yakın olur zaten. P/1-P p'ye yakındır. Ama sadece iç içe vaka-kontrol çalışması olduğu için. Klasik vaka-kontrol çalışmasında bu yapılamazdı.

Yedi yıllık çalışma dönemi boyunca 68 yaş üstü bifosfonat kullanan 205.000 kadın saptanmış. Bunlar arasında subtrokanterik bölgede ya da femur gövdesinde kırığı olan da 726 kadın saptanmış. Olgular 3580 kontrolle yaş ve kohorta giriş yılına göre eşleştirilmiş.

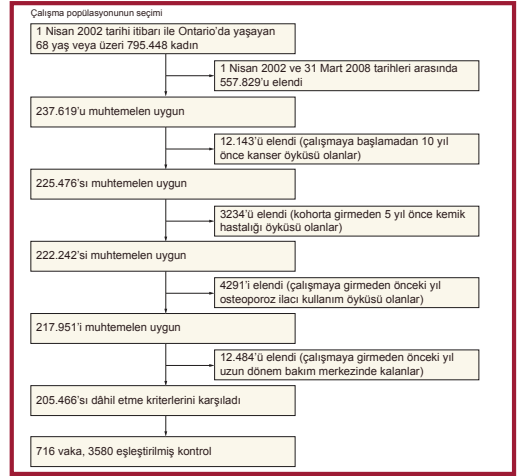
Burada olgu ve kontrollerin başlangıç özelliklerini görüyoruz. Gördüğünüz gibi geçmişte osteoporozla bağlı kırık açısından vakalar ve kontroller arasında fark var, ayrıca son 5 yıl içinde düşme sayısında da büyük fark var. Yazarlar lojistik regresyon modelinde kırık hikâyesi ve geçmişte düşme hikâyesine göre düzeltme yapmışlar ancak bu farkı hiç tartışmıyorlar, bence bu bir randomize kontrollü çalışma olmadığı için yazarların çalışma başındaki özelliklerle ilgili farkları tartışması gerekiyordu.

Çalışmadan önceki 5 yıllık dönem içindeki genel sağlık durumlarına bakıldığında da vakaların bu açıdan kontrollerden daha kötü olduğu görülüyor, gördüğünüz gibi vakalar arasında daha fazla demans, daha fazla konjestif kalp yetersizliği ve hipertansiyon var. Yani vakaların genel sağlık durumu genel olarak kontrollerden daha kötü.

Bu da verilerin geçerliği ile ilgili duyarlık analizi ve kullandıkları tanı algoritmasının subtrokanterik kırıkları %90 pozitif prediktif değerle saptadığını belirtiyor.

Primer analizde subtrokanterik bölge kırıkları için hesaplanan kaçakaç oranının 2,74 olduğunu görüyoruz. Kırık riskinde gözlenen bu artışın geçerliğini teyit etmek için de bifosfonatların osteoporozla bağlı kırık riskinde beklenen azalmayı sağlayıp sağlamadığını anlamaya çalışıyorlar. Böyle yaparak bir tür kontrol olayı kullanıp kohortu doğru değerlendirdiklerinden emin olmak istiyorlar. Bu değerlendirmede de bifosfonatların kendilerinden beklendiği üzere intertrokanterik kırık riskini azalttığı görülüyor.

**Yaşlı Kadınlarda Bifosfonat Kullanımı
ve Subtrokanterik Bölge ya da Femur
Gövdesi Kırıkları**
JAMA, 2011;305(8):783-789



1. Slayt

Bu çalışmanın farklı ve özel bir yanı iki ayrı araştırmayı aynı makalede sunmaları (1. slayt). Her bir çalışma, ilaç kullanımının farklı bir yönünü ve tedavideki yararını anlamamızı sağlıyor. Daha önceki toplantılarda çalışmanın tasarımları hakkında konuşmuştuk, ama konuyla ilgili birkaç hatırlatma yapmak istedim. Daha önce vaka-kontrol çalışmaları ve kohort çalışmalarını ayrı ayrı toplantılarda tartışmıştık. Yuvalandırılmış vaka-kontrol çalışmaları her iki tasarımın da avantajlarından yararlanabiliyor; örneğin yapılabirlik açısından bir vaka kontrol çalışmasına benzer ise de, bir kohort çalışma içine yuvalandırılmış olmasından dolayı, insidans hızlarını ve oranlarını elde edebiliyorsunuz. Bu gibi durumlarda, vakalar çalışmada etkene maruz kaldıktan bir süre sonra görünür hale geldiğinden ve burada maruziyet ilaç (yani bisfosfonat) değil de, ilaca maruz kalma süresi ve

DERGİ KULÜBÜ

Banu Çakır

*Halk Sağlığı Anabilim Dalı
Hacettepe Üniversitesi Tıp Fakültesi*

dozajı olduğu için, etkeni değerlendirmede ilacın dozajına paralel kümülatif zaman hesaplanmış.

Burada dört ayrı grubumuz var. Referans grubu en az 3 ay ila 3 yıl arasında tedavi edilen ve tedaviye başlangıç yaşı 68 ve üzeri olan kişilerden oluşturulmuş. Yaş kesim noktası olarak 68 yaşı seçmelerinin nedeni, Ontario sakinlerinin ancak 65 yaşın üzerinde ‘public prescription benefits’ programından yararlanabiliyor olmaları. Bu program verilerinden yararlanarak, araştırmacılar sadece reçete verilerini kullanarak hastaların ilaç kullanıp kullanmadığını değerlendiriyor, ek veri toplanmıyor. Referans grubunun en az 3 aylık tedavi alan ve uygun yaşın en az 3 yıl üzerinde alan kişilerden oluşturulması gerektiğinden, yaş grubu 65 yaştan başlamıyor, 68 yaş alt sınırı konulmuş. Bu seçimi diğer birçok çalışma ile karşılaştırabiliriz. Çalışmalara ait verilerin karşılaştırılabilir olması için benzer kesim noktaları (örneğin 65 yaş) önemli, eğer bunun dışında bir kesim noktası belirliyorsak bu yaşı (örneğin 68) seçmelerinin anlamlı bir nedeni olmalı ve bu durum raporda açıklanmalı. Bu araştırmada bunun doğru yapıldığını görüyoruz.

Vaka kontrol (1:5); yaş için eşleştirilmiş (± 1 yıl) ve kohorta giriş periyodu (takvim yılı artı 1/4). Kontrollerin (kırksız) endeks tarihi eşleştirilmiş oldukları vakanın endeks tarihine eşitlendi.

En az 5 yıl bisfosfonat kullanan kadınlar arasında subtrokanterik veya femur shaft kırığı insidansını saptamak için 2 yıl daha izlenen kadın kohortunda, neden 1:4 değil de 1:5 seçilmiş? Vaka-kontrol çalışmalarından biliyoruz ki, vaka sayısının sabit olduğu durumda dilersek kontrol sayısını artırarak örnek hacmini büyütebiliriz, bu da çalışmanın gücünü artırmaya yarar. Ancak vaka başına düşen kontrol sayısını 4 seçmek ile daha çok seçmek durumunda çalışmanın gücüne olan olumlu etki istatistik açıdan anlamlı bir artış sağlamıyor, üstelik insan gücü ve finans açısından kayba neden oluyor. Yani, vaka başına seçilen kontrol sayısını 4’den daha fazla arttırmak gereksiz. Burada veriler kayıtlar üzerinden toplandığı için parasal maliyeti ve insan gücünü fazla etkilemediği için ve sanırım veri setinde yeterince

kontrol olabilecek hasta verisi de olduğu için 1:5 seçilmiş. Ancak kontrol sayısı vaka başına 4 de seçebilirlerdi, fark olmazdı.

Vakaların seçilmesine paralel şekilde kontrollerin seçilme tarihleri de benzer seçilmiş. Her bir vakanın çalışmaya başlangıç noktası subtrokanterik veya femur shaft kırığı gibi bir kırık nedeni ile yapılan ilk hastane yatışıdır; yani vaka olduğu tarih. Vaka için eşleştirilmiş kontrol seçilirken, yaş için olduğu gibi, bu durum için de benzerlik aranmış: Kontrollerin endeksi, ilgili vakaların endeksiyle aynı seçilmiş.

Koray bu konular hakkında daha önce konuşmuş olsa da, benim de bir sorum var.

Çalışmada hangi tasarım kullanılmamıştır?

1. Yuvalandırılmış vaka kontrol
2. Kohort
3. Geçerlilik çalışması
4. Eşleştirilmiş vaka kontrol
5. Girişimsel (deneysel) çalışma.

Girişimsel çalışma yapılmamış. Ama çalışma grubundaki tüm bireylerin ilaç kullanıyor olması bazı okuyucularda önyargı oluşturabilir. Oysa tüm kohort üyelerinin bisfosfonat kullanması müdahaleye imkân vermiyor. Tüm bireyler aynı ilacı almış, tek fark tedavinin süresi. Bu süreye de karar veren araştırmacı değil. Yani gözlemsel bir durum söz konusu. Geçerlilik çalışması bir zorunluluk değil ama kırık tanısında ICD kodu kullanmış olmaları nedeni ile bir seçim önyargısı sorunu yaşayabileceklerini düşündüklerinden sayfa 785’deki doğrulama çalışmasını yapmışlar. Eğer ICD kodları yanlış girilmiş ise bu durumda seçime bağlı bir taraf tutma söz konusu olabilirdi: Analizlerde bu nedenle, 4 bağımsız hastaneye yatırılan femur kırıklı 2077 hastada subtrokanterik veya femur shaft kırığı ICD kodlarının doğru olup olmadığını kontrol etmişler.

Hasan Yazıcı: Hastaların nasıl seçildiğini anlatınız ve çalışmanın girişimsel olup olmadığından bahsettiniz. Çalışmanızda “Bisfosfonatlar subtrokanterik kırık gelişimini geciktirmektedir” diyor-sunuz, o halde son olarak ne söyleyebilirsiniz?

Banu Çakır: Bisfosfonat kullanım zamanlamasına göre mi, doza göre mi?

Hasan Yazıcı: Belki bu noktaya değinecektiniz, reçetesi olan herkes kohortu oluşturmuş. Subtrokanterik kırık ile hastaneye yatırılanlar vaka, diğerleri kontrol kabul edilmiş. Bisfosfonat başlanmasının subtrokanterik kırığa neden olup olmadığını nasıl anlarız? En basit yolu şu; bisfosfonat başlanmamış bir popülasyondan bir örnek alıp bisfosfonat kullanan bir gruba karşılaştırarak, bisfosfonatların subtrokanterik kırıkla alakasını tanımlarsınız. Sonra da, subtrokanterik kırığın zamanlamasına bakarsınız. Bu hususta siz ne dersiniz?

Banu Çakır: Her iki durumda çalışılan amaç farklı. Soru ilaç kullanımı ile kırık arasındaki ilişki olsa idi, sizin söylediğiniz tasarım uygun olurdu. Yapmaları gereken ilaçsız bir kontrol grubu oluşturmak olurdu. Ancak bunun için burada çalışılan kayıt sistemi uygun olmayabilir, başka bir şekilde veri toplanması gerekirdi. Burada ise hepsinin ilaç kullandığı baştan kabul ediliyor. Veri tabanındaki herkes zaten bir aşamada ilaç alan kişiler, dolayısı ile referans grubu daha az ilaç kullanan kişilerden oluşturulmuş. Burada ilaç kullanım süresi 3 yıla kadar uzuyor, böylece ilaç kullanmayan kalmamış oluyor. Başka bir ifade ile, etken ilaç kullanmak değil, ilaç kullanma doz ve süresi (veya kümülatif ilaç alım miktarı).

Hasan Yazıcı: Sormamın sebebi yaptıklarını tasarımlarının gücü olarak kullanmaları, çünkü bisfosfonatların tipik kırıkları önleyebildiğini veya geciktirebildiğini göstermişler. Ama bisfosfonatların ileride subtrokanterik kırıklara yol açabildiğini de farz etmişler. Yani, bu tasarım ile 2 ayrı hipotezi aynı anda test etmeye çalışmışlar. Birini ele alıp diğerini test etmişler.

Banu Çakır: Mantıkları sanırım şu: klinik çalışmalara bakarak, tipik kırıklarda bisfosfonatların osteoporozla bağlı kırıkları önlemede yeterli olabileceğini biliyorlar. Ama vaka bildirimlerinden ve küçük gözlemsel çalışmaların çelişkili sonuçlarından, uzun vadede bu ilaçların bazı atipik kırıklara da neden olabileceğini söylüyorlar. Muhtemelen

bu nedenle ilaçsız kontrolleri yok.

Koray Taşçılar: Bisfosfonatların bu tarz kırıkları sadece geciktirdiğini kabul edersek, kimsenin bisfosfonat almadığı bir gözlemsiz süreye de ihtiyaç var demektir.

Hasan Yazıcı: Reçetesi olanlar ve olmayanlar. Bir hipotezi oluşturmada önemli olacak böyle bir veriye sahip değiliz. Tipik kırıklara yönelik bir çalışmada da, ki klinik çalışmalardan elde edilen bulgular doğrulanmış, bir ilaçsız kontrol grubu olmadığını ve daha fazla ve daha az bisfosfonat kullananlar olmak üzere iki grup ile çalışıldığını unutmamak gerek. Bence ilaçsız kontrol grubu olmalıydı.

Koray Taşçılar: Bisfosfonat kullanmaya başladığınızda, kırılma riski genellikle 6 aydan sonra düşmeye başlar. İlk 6 ayda kırık hızında bir düşüş görmezsiniz. Belki de bu nedenle, 100 günden az bisfosfonat kullanan hastaların kırık riski kullananlarla benzer çıkmıştır. Aynı gruptan başka bir referans grup da oluşturabilirlerdi ama bunu da yapmamışlar. Bu iyi bir ‘Editöre Mektup’ nedeni olabilir. Neden verileri yeterince kullanmazsınız ki? Oysa böyle bir grupları da var. Bisfosfonat kullanımı o kadar yaygın olmadığına göre, bir sürü insanı bu gruba dâhil edebilirlerdi.

Biz bunları zaten daha önce tartışmıştık ama sadece bir soru sormak için:

Bu makaleye dâhil çalışmalarda aşağıdaki hangi epidemiyolojik ölçümler hesaplanamaz?

1. İnsidans hızı oranı
2. Risk
3. Atfedilebilir risk
4. Atfedilebilir risk fraksiyonu
5. Popülasyona atfedilebilir risk fraksiyonu

Soru şu, bunlardan hangisini bu çalışmadan elde edemeyiz.

İlk 4 ölçüt hesaplanabilir.

Bugün sizinle popülasyona atfedilebilir risk fraksiyonu hakkında kısaca konuşmak isterim.

Çünkü bu konu çoğu zaman gözden kaçıyor. Klinisyenler için öncelikli bir mesele olmasa da halk sağlığı çalışanları ve politika yapıcılar için popülasyona atfedilebilir risk hesaplamaları önemli bir konu. Koray'ın bahsettiği atfedilebilir risk, ki bu çalışmada da hesaplanmış ve popülasyona atfedilebilir risk farklı iki kavram. Genel popülasyonda bisfosfonat kullanmaksızın karşılaşılabilecek kırıklara karşılaştırılarak hesaplanacak oran demektir. Popülasyona atfedilebilir risk de atfedilebilir risk de 'insidans' ihtiyacı gösterir. Fraksiyon formülü, ki burada fraksiyon kullanılmış; bu bir klinik çalışmadan elde edilse idi, yeni ilaç başlananların riskinden plasebo veya diğer ilacı alanların riskini çıkararak elde edeceğimiz değer atfedilebilir risk olurdu. Eğer bu riski yeni ilaç alan birey sayısına bölerseniz, müdahaleye (yani ilaç kullanımına) atfedilebilir risk fraksiyonunu elde edersiniz. Aynı durum burada da geçerli ama şunu unutmayalım, hangi grupta çalışırsanız çalışın, mutlaka bisfosfonat alan da olacaktır almayan da ve mutlaka kırıkların bir kısmı subtrokanterik ya da femur shaft kırığı olacaktır. Başka bir ifade ile, insanların bisfosfonat kullanmasını engellesek de bir risk ve toplam risk yine de olacak. Bu toplam riskin yüzde kaçını sadece bisfosfonat kesilerek düşürülebilir? Bunu rate ratio (insidans hız oranları) ve insidans hızlarından elde edebiliriz. Şayet insidans hızı eksikse, yerine risk oranları da kullanılabilir.

Çok nadir bir hastalık durumunda odds oranı risk oranını tahmin etmemizi sağlayabildiğine göre, risk oranı yerine odds oranı kullanılabilir mi? Bu çalışma tasarım olarak klasik bir vaka-kontrol araştırması değil de yuvalandırılmış tipte bir vaka-kontrol çalışması olduğu için, lojistik regresyon modellemesinden risk hesaplaması yapılması mümkün. Bu yüzden burada bu formülü kullanıyoruz. Odds oranından, ki lojistik regresyon modellemesinde buna risk odds oranı denir, 1 çıkarıp, makaledeki risk odds oranına bölüyoruz. Risk oranı bu kohortta risk faktörüne yani bisfosfonata maruz kalanların subtrokanterik veya femur shaft kırığı geçirme boyutunu gösterir.

Önemli olan nokta şu ki, ilaç bu hastalığı önlemede ne kadar yararlı olursa olsun, bir popülasyonda bisfosfonat kullanımının yaygınlığını bilmedikçe topluma atfedilen risk oranının hesaplanmasının anlamı kalmaz. 785. sayfadaki hesaba göre, en az 5 yıl süre ile bisfosfonat kullanan kadınlardan oluşan kohortta subtrokanterik ya da femur shaft kırığı insidansına bakılarak, her 10 kişiden sadece 1'inde bisfosfonatları keserek bu tür kırık gelişiminin önlenebileceği iddia edilmiş. Çalışmada, bisfosfonat kullanımının halen yaygın olmadığı, gelecekte muhtemelen daha yaygınlaşacağı ve dolayısı ile popülasyona atfedilebilir riskin de daha yüksek olacağı yorumu yapılmış. Bunu doğru kabul edersek, 5 yıldan daha uzun süre ile bisfosfonat kullanan bireylerde daha sık subtrokanterik veya femur shaft kırığı göreceğiz demektir. Sonuç olarak, yazarlar bisfosfonat kullanımının tipik osteoporotik kırıklarda çok etkili olduğunu bildiklerinden ve geçerlilik çalışmalarında bunu doğruladıklarından, ilacın kesilmesini önermiyorlar, ama 5 yılın sonunda tedavinin hastaya göre planlanmasını, bisfosfonat kullanırken 'ilaç tatili' verilip verilmeyeceğinin hastanın durumu özelinde kararlaştırılmasını öneriyorlar.

Bir konuyu da hatırlatmalıyım. Klinik çalışmalarda çoğu zaman göreceli risk kullanırız ama atfedilebilir risk de kullanmalıyız, çünkü maliyetin hesaplamalarında mutlak riskin bilinmesi önem taşır. Ayrıca her bir vakada, yazar ister bir klinisyen veya politika yapıcı veya önleyici tıp uzmanı olsun, mümkünse, popülasyona atfedilebilir fraksiyonu da hesaplamalı. Bisfosfonat kullanımının maliyeti yüksek olduğuna göre, bu sorunun toplum açısından önemi nedir? Sosyal güvenlik sistemi için bir önceliği olmalı mıdır?

Bir başlık açmak için, çalışmaya yönelik bir soru daha.

Aşağıdakilerden hangisi çalışmanın potansiyel bir kaynak hatasıdır? Bunlardan bazıları çalışmada tartışılmış.

1. Seçime bağlı taraf tutma
2. Rezidü karıştırıcı
3. Tip 1 hata

4. Bilgi önyargısı

5. Hepsi

Hangisi bir potansiyel taraf tutma kaynağıdır? Sonraki konu da makalede bunu nerede ele almışlar?

Dünkü oturumda çalışmanın eksiklerini konuştuk ve güzel de bir tartışma oldu bence. Bu çalışmada yanlış ICD kodu kullanılması veya hekimlerin ICD kodlamasını unutmaları/boş bırakması nedeni ile seçime bağlı taraf tutma artmış olabilir. Yani, farklı tanılarda kişiler çalışmaya dâhil edilmiş olabilir ya da hatalı olarak grupta yer alması gereken hastalar çalışmaya dâhil edilmemiş olabilirler. Bunun için bir doğrulama çalışması yapılmış ve seçime bağlı taraf tutma konusunda nasıl tedbir alınacağı tartışılmış.

Rezidü karıştırıcı deyince, bu çalışmada veriler kayıtlardan elde edildiği için birçok olası karıştırıcı faktörün çalışma sonuçları üzerinde etkileri olabilecek iken, bunlar hakkında veri tabanından bilgi edinilememiş olabilir. Çalışmanın yorum bölümünde bu sorun çok güzel bir şekilde tartışılmış. Örneğin osteoporoz durumu önemli bir karıştırıcı faktör. Vakaların osteoporoza meyli açısından daha en başından kontrollerle karşılaştırılabilir olmadığını tartışmıştık. Çalışmada egzersiz, beslenme, sigara kullanımı vb'ye dair vakalara veya kontrollere ait herhangi bir veri olmadığı gibi, vücut ağırlıkları, OTC ilaç kullanımları ve benzeri gibi konularda da bilgisiziz.

Tip 1 hata biliyorsunuz rastgele bir hata, önlenemez ancak boyutunun ne olduğu rapor edilebilir ve okuyucu bunun boyutundan haberdar olur. Buna yönelik raporda hesaplanmış bol miktarda p değerleri mevcut.

Kayıt bilgileri kullanıldığından, farklı zamanlarda farklı kişiler tarafından toplanmış, standart olmayan bilgi toplama yöntemlerinden kaynaklanan bir bilgi toplamaya bağlı taraf tutma da söz konusu olabilir. O halde, sorunun doğru cevabı 5. şık.

Bir geçerlilik çalışması yapmışlar, 786. sayfadaki soru şu idi:

Geçerlilik çalışmasından elde edilen aşağıdaki ölçüm(ler)den hangileri çalışmaya dâhil edilmiştir?

1. Duyarlılık
2. Özgüllük (seçicilik)
3. Pozitif prediktif değer
4. 1 ve 3
5. 2 ve 3

Duyarlılık ve pozitif prediktif değer çalışılmış.

Soru şu ki; ICD-10 subtrokanterik veya femur shaft kırığı kodlamasına göre pozitif prediktif değer %90 ve duyarlılığın %81 olduğu belirtilmiş, çalışmada sunulan sonuçlar sizce bisfosfonatlara bağlı subtrokanterik veya femur shaft kırığı gerçek riskinin üstünde mi altında mı?

Birlikte düşünelim. %90'lık pozitif prediktif değer kırığı olduğuna inandığınız bireylerden %10'unun aslında yanlış pozitif olduğu anlamına gelir. Gerçek bir ilişki varsa ve bu çalışmada %10 yanlış pozitif mevcutsa, hesapladığınız değer gerçek değerden düşük demektir. Yani, 5 yıl veya daha fazla bisfosfonat kullanımı makalede yazılandan daha fazla zarara sebep olmuş olabilir.

Yüzde 81'lik duyarlılık, bu tür kırıklardan hemen hemen %20'sinin çalışmaya dâhil edilmediği anlamına gelir. Gerçek bir ilişki varsa, burada tanımlanan %10'luk topluma atfedilebilir risk de gerçek değerden altında demektir. Geçerlilik ile ilgili bir sorun olsa bile, burada bildirilen değerler gerçeğin altındadır.

Bu da son soru, bunu daha önce tartışmıştık:

Aşağıdaki istatistiksel işlemlerden hangisi sunulan çalışmada kullanılamaz?

1. Odds oranı
2. Güven aralığı
3. p değeri
4. Unconditional lojistik regresyon
5. Geçerlilik analizi

Unconditional lojistik regresyon doğru cevap. Bir araştırmada eşleştirilmiş vaka-kontrol tasarımı

kullandığınızda, tüm istatistiksel analizler buna uygun yapılmalıdır, çünkü aksi takdirde bazı geçersiz sonuçlar alabilirsiniz. Bu bir eşleştirilmiş vaka-kontrol çalışması olduğundan, unconditional değil conditional lojistik regresyon modellemesi kullanmışlar.

Teşekkürler.

Emmanuel Lesaffre: Conditional lojistiğin aşırı uyumu denince, lojistik regresyonda tahminler sonsuza çıkabilir. Bu bir neden olabilir; o halde ‘ayırım problemleriyle’ karşılaşabilirsiniz. Analizle de bir kolonun boş olduğu, diğer kolonlarda tüm vakaların bulunduğu birçok izole kategorik kolonunuz varsa buna aşırı uyum denir, ki bu temelde sayısal bir problemdir.

Bir Katılımcı: Bir soru sorabilir miyim? Aynı zamanda bir eklemem de olacak. Aşırı uyuma yönelik modelden birçok değişken gösterdiniz ve ilaç ve ko-morbidite gruplanmasına rağmen yaklaşık 30 değişken gördük. Bunun dışında, bu bir eşleştirilmiş vaka-kontrol çalışması ve eşleştirme bireyin yaşı +/-1 yıl formülü ile yapılır. Bu çalışmaya katılanlar 68 yaş üzeri olduğuna göre, birçok yaş grubu oluşturulmuş olmalıydı ve ikinci eşli değişken kohort girişi olmalı, ki yaklaşık 5 yıldır. Bu bir eşleştirilmiş vaka-kontrol çalışması olduğuna göre, tüm eşleştirilmiş olanlar modelde değişken olarak yer alacaktır. Conditional lojistik regresyonda, grupta toplam birey sayısı yaklaşık 4000 kişidir ve 100 değişken olması tahmin kalitesini düşürebilir, doğru mu? Uyum problemi de olabilir

mi, örnekteki birey sayısı modeldeki değişken sayısı ile ilişkili olabilir mi?

Emmanuel Lesaffre: Hiçbir istatistikçinin bir sihirli sayısı yoktur. Ama kategorik eş değişkenleriniz varsa, belirsiz tahmin riskiniz klasik lojistik regresyonda olduğu gibi yüksek. Klasik regresyon kullanarak bazı istatistiksel araştırmalar yapmıştım ve gördüm ki farklı eş değişkenleriniz varsa ve özellikle bir grupta hiçbir vaka yokken vakalar diğer bir grupta ise başınız belada. Bence işte bu bir aşırı uyum durumu. Prensip, temelde aşırı uyum olmaz. İstatistikte, aşırı uyum neredeyse imkânsız çünkü genelde rezidü karıştırılarda problem yaşayabilirsiniz.

Hasan Yazıcı: Anti-TNF lenfoma çalışmasında karşılaştığımız meseleler bunlar. Aslında daha basit bir tasarım geliştirilebilirdi. Amerika Birleşik Devletleri’nde ortopedi kliniklerine gidin ve bu tür kırıkları olan insanlara basit bir soru sorun; “Bisfosfonat kullanıyor musunuz ve ne kadar zamandır?”, ardından kırık geçirmemiş uygun bir kontrol grubu oluşturun, sonuçta bir cevaba varabilirsiniz. Yine birkaç yıl veya ay ilaç kullanmış olanları da bulursunuz. Geriye dönük bir çalışma bana çok ilginç geliyor, bence bu bizim reçete, veritabanları, kayıtlar ve çalışma tasarlamak dışındaki her şeyden bilgi edinme kolayına kaçmamızla ilgili olabilir. Bunu Ankara’da yaptık, insanları aradık ve bisfosfonat kullanıp kullanmadıklarını sorduk ve bir örnek oluşturduk ve onlara hiç bisfosfonat kullandınız mı, diye sorduk, burada bir fark var.

ROMATOLOJİDE KLİNİK ARAŞTIRMALAR VE NESNEL TİP VII

EDİTÖR
HASAN YAZICI

B. ÇAKIR
Ö. ERGÖNÜL
G. HATEMİ
S. SUISSA
Y. YAZICI

Öncelikle davetiniz için teşekkür ediyorum. Bu toplantılara katılmaktan çok mutlu oluyorum ve bugünkü konuşmaları da çok heyecanla bekliyorum.

Bugün birkaç farklı konuda konuşmalar olacak, biz de randomize kontrollü çalışmaları tekrar gündeme getirelim, önemli özelliklerini vurgulayalım istedik. Randomize kontrollü çalışmaları çoğu kez sadece avantajları ve kanıt kalitesine sağladıkları ile değerlendiriyoruz, kısıtlılıklarını görmezden geliyoruz. Randomize kontrollü çalışmalar sağlık alanında kanıt sağlama açısından mutlaka çok iyi özellikleri ve avantajları olan tasarımlar ama bugün biraz daha ağırlıklı olarak kısıtlılıklarını tartışalım, yöntem olarak sunarken dikkat etmemiz gereken noktalara dikkat çekelim istedik. Onun için başlığı, “Okuyucu, Dikkat Et!”

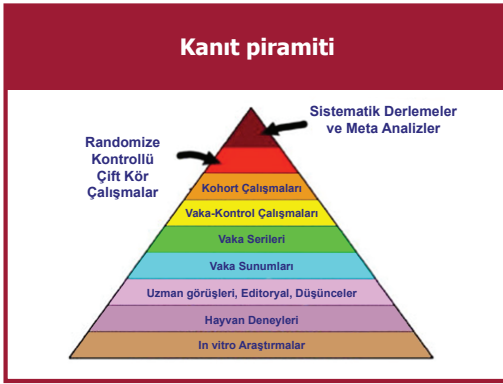
Hep birlikte tartışalım istiyorum ama benim aklımdaki kabaca plan şöyle: Randomize kontrollü çalışmaların kanıta dayalı tıptaki yeri ve güçlü yanlarını hızlıca geçtikten sonra, bir takım içsel sınırlılıkları var, bunları konuşalım. Bir de, “Hep randomize kontrollü çalışmalarda randomizasyonun ve kör tasarımların öneminden bahsediyoruz ama bu çalışmalarda ne tür hatalarla karşılaşabiliriz, hataların birbirlerine göre göreceli boyutları neler olabilir?” bunları konuşmayı arzu ediyorum. Hedefim, konuşmanın sonunda, randomize kontrollü çalışmaları okuduğumuzda ilgili çalışmanın sınırlılıkları konusunda zihnimizde hızlıca gelişirebileceğimiz kişisel kontrol listeleri için bir takım ipuçları oluşturmak. Sorularınızla ve katkılarınızla gelişecek bir konuşma.

Aramızda kanıta dayalı tıp çalışan arkadaşlarımız var, biliyorum. Kanıta dayalı tıp deyince ilk akla gelen genelde ‘eldeki en iyi tedaviyi uygulamak’ ve bunun için de ‘mevcut kanıtı en iyi şekilde değerlendirmek’. Ama en az bu kadar önemli olan kısım doktorun ve hastanın tercih ve deneyimlerini eş zamanlı olarak değerlendirmeye katmak. Kanıta dayalı tıpta çalışma adımlarına bakacak olursak ilk önce bir soru formüle ediyoruz (P), bu soruda hangi müdahalenin bizim için önemli olduğunu belirliyoruz (I), karşılaştırma yapılacak yöntem, yol seçiliyor (C) ve hangi son durumu inceleyeceğimize karar veriyoruz (O). Kısa ismi

RANDOMİZE KONTROLLÜ KLİNİK ÇALIŞMALAR: OKUYUCU, DİKKAT ET!

Banu Çakır

*Halk Sağlığı Anabilim Dalı
Hacettepe Üniversitesi Tıp Fakültesi*

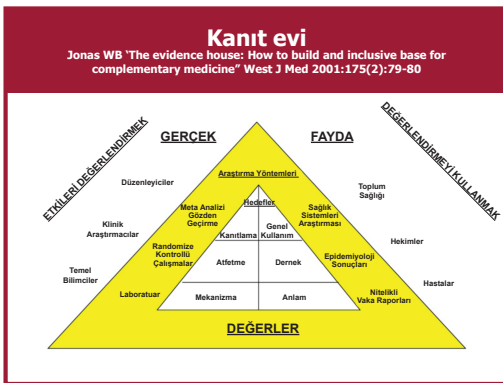


1. Slayt

PICO yaklaşımı. Bundan sonra mevcut en iyi kanıt bulmaya çalışıyoruz ve bulduğumuz kanıtı tekrar geri dönüp değerlendirmemiz gerekiyor: Hem geçerlilik açısından, hem hastamıza uygulanabilirlik açısından ve tabii uygulandığında ulaşılacak etki açısından. Ama burada, az önce söylediğim gibi klinisyenin rolü, deneyimi çok önemli. Ve yine hiç unutmamamız gereken, hastanın kişisel değerleri fizyolojik olsun, patolojik olsun ya da düşünsel, sosyal, davranışsal olarak, içinde bulunduğu konum olarak bu tedavinin hastaya uygunluğunu muhakkak tartmamız lazım. Bunu yerine getirene kadar tüm adımları tekrar tekrar yapıyoruz.

Kanıt dayalı tıp uygulamalarında yeri açısından bu üç adımdan biz bugün sadece “kanıt kalitesi”ni ve bu kapsamda randomize kontrollü çalışmaların yerini tartışacağız. Bu bizim sık olarak görmeye alışkın olduğumuz kanıt piramiti (1. slayt). Kanıt kalitesi göreceli olarak düşük olanlar en altta, ancak bunların sayıları çok, dolayısıyla piramidin en alt kısmı en geniş. Tanımlayıcı çalışmalar en altlarda. Elde edilebilecek kanıt kalitesi açısından daha yukarı çıktıkça vaka kontrol çalışmaları, kohort tipi araştırmaları, en yukarıda ise randomize kontrollü, özellikle de çift kör çalışmaları görüyoruz. Eğer en yukarıdaki meta analizleri bir yana bırakırsak, tek çalışmaya özel veriler incelendiğinde epidemiyolojik açıdan veri kalitesi en yüksek çalışma tasarımı randomize kontrollü çalışmalar. Ama bu önemi vurgularken ikinci piramiti de hiç unutmamamız gerek (2. slayt).

İkinci piramite baktığınız zaman, bizim çok değerli olduğunu düşündüğümüz ve hayatımızın önemli kısmını okumak/gerçekleştirmek için ayırdığımız randomize kontrollü çalışmaların bu kanıt evi içerisinde ufak bir odayı temsil ettiğini fark ediyoruz. Jonas WB’nin bu yazısı, okumak isterseniz çok güzel bir yazı. Okuması yormayan, ama birçok güzel mesajı olan bir yazı. Diyor ki “Evinizin mutfağının ya da salonunun çok geniş olmasının, yatak odanızın çok küçük olduğu durumda bir anlamı yoktur.” Yani kanıt evi içerisinde bütün odaların yaşamak için uygun genişlikte mekânlar olması lazım, bir evi sevebilmeniz için. Onun için bir gerçeğin “tümünü” bulmak ve başarıyla uygulayabilmek için, örneğin ‘bir ilacın gerçek uygulama



2. Slayt

madaki etkilerini değerlendirmek' istediğimizde bir yanda laboratuvar testlerinden tutun da meta analizlere kadar giden kanıt piramidi, öbür yanda hastaların, uzmanların, klinisyenlerin halk sağlığına genel olarak bakışı var. Özellikle klinisyen olarak çalışan hekimlere sadece kanıt piramidine odaklanmayıp, bu büyük pencereyi, kanıt piramitinin getirdiklerinin doldurduğu odanın evin tamamı içerisindeki yerini ve önemini hatırlatmakta yarar var diye düşünüyorum.

Randomize kontrollü çalışmalarda hep vurguladığımız çalışmanın iyi yanları, güçlü yanları nedir? Deneyisel çalışmalarda olan ve bu tasarımları gözlemsel çalışmalardan farklı kılan en önemli özellik nedir? Bu bir ilaç olabilir, cerrahi bir uygulama olabilir, önemli özellik 'etkenle karşılaşma durumuna araştırmacının karar veriyor' olmasıdır.

'Intervention' (müdahale) kelimesi ile 'invasiv' (girişimsel) kelimeleri karıştırıldığında önemli bir yanlış ortaya çıkıyor. Epidemiyolojide 'deneyisel' çalışma deyince bir 'müdahale' çalışmasından bahsediyoruz. 'Müdahale' yerine 'girişimsel' dendiğinde hastadan kan almak, biyopsi yapmak, hastayla ilgili bir girişim yapmak akla geliyor ve iki durum karıştırılıyor. Çalışmada bir etken ile sonuç arasındaki ilişki incelendiğinde 'etkenle karşılaşmaya' (yani ilacı almaya, cerrahi yöntemde ne kullanılacağına vb.) hasta veya doktoru karar verdiğinde, bu seçimi etkileyen çok sayıda faktör olabilir ve bu faktörler aslında hastanın akıbeti ile doğrudan ilişkili de olabilir. Öte yandan etkenle karşılaşmaya müdahale eden kişi 'dışarıdan biri' (yani, araştırmacı) olduğunda, hele de bu seçim rastgele yapıldığında, mümkünse de bundan hasta (tek kör) ve/veya doktoru (çift kör) habersiz kılındığında işte bu tür dışsal faktörler (karıştırıcı faktörler) çalışma sonuçlarını daha az etkileyecektir. Bunların tümü 'istenen' özellikler.

Randomizasyon ve körleme (maskeleye), bugün Önder Bey'in konuşmasında da göreceğiz, toplanan verinin epidemiyolojik açıdan kalitesini sağlamada değerli kavramlar.

Değişik çalışmalardan etki-sonuç ilişkisini çalışırken farklı ölçütler hesaplayabiliyoruz.

Bunların arasında odds ratio var, rölatif risk var. Rölatif risk, risk farkı değerlerini en iyi elde edebileceğimiz çalışma tipi yine randomize kontrollü çalışmalar. Bu tür çalışma sonuçlarının analizinde bulguların genellenebilirlik değerini kestirmeye yarayan bir de yöntem var 'intention-to-treat prensibi'. Önder ona bir isim bulmuştu, 'niyet ettim tedaviye' şeklinde... ki ben de bu çeviriyi değişik yerlerde kullanıyorum, çok severek.

Intention-to-treat prensibi kullanıldığında randomize kontrollü çalışmalardan elde edilen sonuçların, gerçek yaşamda elde edilenlere benzerlik gösterebilirliğini değerlendirmek kolaylaşıyor. Çünkü *efficacy* (yani laboratuvar koşullarında elde edilen etki) ile gerçek yaşam uygulamalarında gözlenen etki (*effectiveness*) farklı olabilir, hep konuşuyoruz, bu tür analitik yöntemler kullanıldıkça randomize kontrollü çalışmalardan elde edilen verilerin kullanılabilirliği de artıyor. Tüm bu bahsettiğimiz özellikler randomize kontrollü çalışma tasarımlarının hep iyi yanları.

Randomizasyonun en iyi özelliği de, hem seçime bağlı taraf tutmayı hem de karıştırıcı faktörleri gruplar arasında dengeleyebilmek. Benzer kılabilmek. Körleme nasıl işe yarıyor? Sonucun değerlendirilebilmesinde özellikle çok yarar sağlıyor. *Intention-to-treat* (niyet ettim tedaviye) analizi, özellikle izlemede kayıplar olduğu zamanlarda ya da gruplar arasında karşılıklı geçişler/yer değiştirmeler olduğu durumlarda bunun etkilerini dengelemeye, kontrol etmeye yarıyor. Standart protokoller kullanılması da çalışmadan elde edilen verilerin kalitesini artırıyor. Ama bunlar var, yapılıyor diye hiçbir zaman körü körüne sonuçlara güvenmememiz lazım. Örneğin yayımlanmış makalelerde randomizasyon yapıldığı belirtiliyor ise bunun nasıl yapıldığını da incelememiz gerekiyor.

1. Mesela, randomizasyon işlemi doğru şekilde yapıldı mı? Yazı tura atmak randomizasyon için yeterli mi? Ya da Pazartesi bir grubun hastasını alıyorum, Salı günü gelen hastayı öbür gruba alıyorum. Bunlar gerçek anlamda güvenilir randomizasyon yöntemleri değil. Bunu daha önceki toplantılarda da konuşmuştuk.

2. Ayrıca, randomizasyon yapıldıktan sonra hasta etken ile karşılaşınca dek bunun gizlenmesi, saklanması da önemli; çünkü öbür türlü randomizasyonun bozulma olasılığı var. Ve yine randomizasyon yapıldıktan sonra hastanın bu ilacı alıncaya kadar geçen süreler de çok önemli. Bunu ‘*allocation concealment*’ başlığı altında konuşmuştuk.
3. Ya da randomizasyonu yaptınız, hastaları iki gruba ayırdınız, sonrasında bazıları çalışmaya katılmaktan vazgeçti, katılamaz duruma geldiler, şehir değiştirdiler/hasta oldular. Bu gibi durumlar çalışmanın sonuçlarının güvenilirliğini bozabilir.

Dolayısıyla, çalışmanın randomize kontrollü yapılmış olması yeterli değil. İlgili özellikler ve işlem basamaklarının detaylı olarak açıklanması da gerekiyor.

Ayrıca bugün randomize kontrollü klinik çalışmalarda içsel olarak var olan kısıtlılıklardan bahsedelim istiyorum. Bu kısıtlılıklardan bir tanesi, randomize kontrollü çalışmaların genellenebilirliği ya da dış geçerliliği (external validitesi) düşük. Düşük olmasının bazı sebepleri var, ama genellenebilirlik derken hepimiz aynı şeyi mi anlıyoruz, ondan emin olalım.

Toplulun tümüne gidilmediği durumlarda çalışma grubuna alınan (örneğe çıkan) bir grup insan var, dolayısıyla bu kişilerin çalışma sonuçlarının genelleneceği insanları/hastalar grubunu ne kadar temsil ettiği önemli. Bunların geldiği kaynak grubun kim olduğu, diyelim romatoid artritlerin, tümünü ne kadar temsil ettiği önemli. Tabii çalışmanın amacına göre temsil edilecek grubun (hedef/referans grubun) kim olduğu değişebiliyor. Randomize kontrollü çalışmalarda dış temsiliye/geçerlilik bir kohort ya da kesitsel çalışmadaki kadar yüksek değil. Temsiliyetin bu randomize kontrollü çalışmalarda düşük olmasının farklı nedenleri olabilir. Sizler de kendi çalışmalarınızdan eminim biliyorsunuz. Doktorun muayenehanesine/hastaneye gelen hastayla deneysel bir çalışmaya girmeye ‘gönüllü olan kişilerin’ özellikleri farklılık gösterebilir. Özellikle de komorbidite açısından. Eğitim, hasta bakımına erişilebilirlik,

sosyal yapı, gelir, ilaç kullanımını kabul vb. özellikler açısından randomize kontrollü çalışmayı kabul eden kişiler ile ilgili hastalık grubu genelinin özellikleri tam olarak örtüşmeyebilir. Benzer şekilde ‘gerçek yaşamda’ hastalar kendilerine reçete edilen dışında geleneksel tedavi yöntemlerinden de yararlanıyor olabilir. Bu gibi durumlarda randomize kontrollü çalışmadan alınan sonuçlar ile ilacın toplumda uygulanması durumunda elde edilen sonuçlar farklı tür ve etkide olabilir.

Bir klinik çalışmada çalışılması hedeflenen son durumu (outcome) her zaman sizin bir hekim olarak hastada elde etmek istediğiniz son durum ile aynı olmayabilir. Örneğin, randomize klinik çalışmada ilgilenilen son durum ‘düzeltme ya da ölümün engellenmesi’ gibi bir sonuç olabilir ama siz kendi hastanızda düzeltme yerine ‘daha iyi bir yaşam kalitesi’ni hedefliyor olabilirsiniz. Bu durumda randomize kontrollü çalışmadan elde edilen kanıt ne kadar yüksek düzeyde olsa da doğrudan sizin ölçmek istediğiniz duruma uyarlanamayabilir.

Diğer bir kısıtlılık, çalışmalara alınan kişi sayısı ile ilişkili. Deneysel çalışmalara kabul edilebilir kişi sayısı yapılacak test/televi/sigorta bedelleri nedeniyle kısıtlı olmak zorunda. Ayrıca, çalışma her tür hastaya uygulanamıyor, gönüllülük esas. İzlem süresi de sonuca ulaşmak için yeterli süre ile sınırlanıyor. Onun ötesinde yıllarca izlem yapılmıyor. Örneğin, izlemler 3 yıl bilemediniz 5 yıl süreden sonra devam edilmiyor. Dolayısıyla çalışmanın örnek hacminin küçük olmasından gelen bir Tip 2 (beta) hataya bağlı sorunlar yaşanabilir. Çalışmanın gücü düşük olabilir, ‘gerçekte var olan bazı ilişkiler’ gösterilemeyebilir. Ya da, özellikle ortaya çıkma olasılığı nadir sonuçları, etkileri kısıtlı sayıda hastalar ile yapılan deneysel araştırmalarda tayin edemeyebiliriz. Sık gözlenen bir durum da, uzun süre izlemde (kümülatif kullanım sonrası) ortaya çıkması olası durumlar, örneğin bazı advers etkiler, çalışma için yeterli izlem süresi dahilinde gözlenemeyebilir.

Türkçe sözlüklerde İngilizce’de *efficacy* ve *effectiveness* kelimeleri ortak olarak ‘etkililik’ şeklinde yer aldığı için bu ayrımı yapmak kolay olmuyor.

Öte yandan, laboratuvar koşullarında bir randomize klinik çalışmada bulduğumuz etkililik (*efficacy*) gerçek topluma, rutin uygulamada elde ettiğimiz etkililik (*effectiveness*) her zaman örtüşmüyor. Bunun için hep diyoruz ki, “Randomize klinik çalışmalar yapıldıktan sonra ilaç ruhsatlanıp piyasaya çıktıktan sonra mutlaka gözlemsel çalışmalarla, kohortlarla desteklensin”. Özellikle, nadir görülen ya da geniş, heterojen popülasyonlarda var olabilen, farklı genetik havuzlarda, heterojen uygulamalarda vb. ortaya çıkabilecek etkiler, uzun izlemlerde ortaya çıkabilecek, kümülatif olayların getirebileceği güvenilirlik durumlarını (advers etkiler gibi) çalışmak için mutlaka gözlemsel araştırmalara ihtiyaç var. Ruhsatlandırma sonrası yapılacak kohort araştırmalar, randomize kontrollü çalışma sonuçlarını desteklemek açısından çok değerli. Bu tür gözlemsel araştırmalar ile desteklensin istiyoruz. Yani randomize kontrollü çalışmayı yapmakla ne yazık ki orada kalamıyoruz. Bu kısıtlılık değil belki ama sorunlar, zorlandığımız şeyler. Hepimizin bildiği şeyler, finans sorunları, izlemede gereken zaman, insan gücü sorunları gibi tartışılması gereken çok fazla konu var.

Randomize kontrollü çalışmalardan bahsederken mutlaka etik konulara dikkat çekmek gerekir. Gözlemsel araştırmalar ile deneysel araştırmalar karşılaştırılırken etik konular ve etik onay süreçlerindeki farklılıklar da vurgulanmalı. Randomize kontrollü çalışmalarda aydınlatılmış onam alınması, hastanın sigortalanması, araştırmanın gerektirdiği tüm masrafların araştırmacı tarafından karşılanması vb. gibi durumlar etik konuların özenle ele alınmasını gerektiriyor.

Randomize kontrollü çalışmalarda başarılı olabilmek için hem araştırmacılar hem de çalışmaya katılan hastaların katılım ve izlemler konusunda gönüllü olması, çok ciddi bir emek sarf etmesi gerekiyor. Bu konuda kısa bir makale paylaşmak, tartışmak istiyorum, size de okumanızı öneririm. Konusu, randomize kontrollü bir klinik çalışmada kanıtı nasıl değerlendirelim? Bunun için biliyorsanız, Jadad skorlama sistemi gibi, farklı skorlama sistemleri var, bu skorlama sistemlerini kullansak yeterli olur mu? Randomizasyon varsa sonuçlar güvenilir mi? Ya da randomizasyonla körlemenin

hatayı önleme açısından gücü göreceli olarak aynı mı, bunları konuşmak istiyorum şimdi. Çünkü hepimiz ip üstünde yürüyoruz, ne kadar kalın ve güvenilir olursa o kadar yararlı olacaktır.

Bu da (3. slayt) sizinle ikinci yarıda konuşmak istediğim makale. Bajard ve arkadaşları Monte Carlo simulasyon yöntemi ile oluşturulan sanal bir hasta havuzu üzerinden randomizasyon yapılmasını ve etkilerini çalışmışlar. Eğer randomizasyon yeterince iyi yapılmazsa, körleme iyi yapılmazsa, randomizasyondan sonra hastaya ilaç verilinceye kadar geçen dönemde bu durum iyi gizlenmezse, bu dönemde tedavinin etkililiğini hatalı olarak gerçek değere kıyasla %30-%50 daha fazla gösterebildiği konusunda bilimsel kanıtlar var.

Bu gibi olası hatalar bizi değişik randomize kontrollü çalışmaları kalite açısından değerlendirilmede skorlama sistemlerin kullanımına yönlendiriyor. “Randomizasyon yapılmış mı, yapıldıysa iyi tariflenmiş mi, kör mü, çift kör mü, körse iyi tariflenmiş mi, izlemede kayıplar var mı?” Skorlama sistemlerini incelediğinizde bu tür durumlar için ya 0 ya da 1 şeklinde bir puanlama yapıldığını görüyoruz. Yani göreceli olarak bunların ağırlıkları aynı. Oysa randomizasyon yapılmaması 0 puansa, çalışmada maskeleme yapılmamış olması da sonucu aynı güçle etkileyecektir diyemezsiniz. Bajard ve arkadaşlarının makalesinden araştırmacılar bu tür bir “Eş ağırlıklandırmanın etkileri gerçekten benzer mi?” diye de inceleme yapmışlar. Diyorlar ki “Farklı randomize kontrollü çalışmalarda farklı hata kaynakları var”. Bunlar hiyerarşik olarak nasıl? diye değerlendirildiğinde: Çalışmaya alınan kişilerin toplanması/bir araya getirilmesinde, randomizasyonunda, bu randomizasyonun gizlenmesinde, plasebo ya da nosebo kullanılmasında etkiler neler olabilir? “ “Kişinin aldığı ilacı bilmesi ya da o ilaçtan ‘iyi ya da kötü beklentileri’ne göre alınacak sonuç değişebilir mi?”.

Bu makalede çok dikkatle ele alınan konulardan bir de ölçüm hataları. Bu ölçüm hataları ilacın tedavi edici etkilerini ölçerken yapılabilir ya da ilacın advers etkilerini ölçerken yapılabilir. Hasta çalışmada verilene ek olarak başka tedaviler de alıyorsa bunun getireceği tedavi edici ve advers

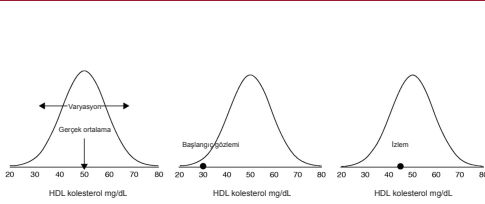
Revisiting the level evidence in randomized controlled clinical trials: A simulation approach

Bajard A, Chabaud S, et al.

Contemporary Clinical Trials 30 (2009) 400-410

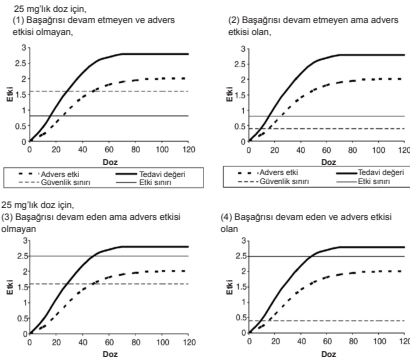
3. Slayt

Ortalamaya kayma toplumdaki asimetrik örnekler alırsan ortaya çıkar



Barnett A G et al. Int. J. Epidemiol. 2005;34:215-220

4. Slayt



5. Slayt

etkileri de olabilir. Son olarak, çalışmanın izlem süresindeki kayıplar da önemli. Bir de 'regression to the mean' denilen, tekrarlayan ölçümlerde elde edilen değerlerin hasta değerlerinden hastalardan elde edilen ortalama değerlere kayan bir yapı söz konusu olabilir.

Bu konuların tümünü değişik toplantılarda konuşmuştuk, bunlar arasında bugün 'regression to the mean' (ortalamaya kayma) kavramını tekrar tartışmaya açmak istiyorum.

Randomize kontrollü çalışmalarda çalışmaya dâhil edilen kişilere ait değerler toplumu doğru/tam olarak yansıtmıyor olabilir. Çalışma grubunun toplumda rastgele seçilmediği durumlarda çalışmaya dâhil edilen kişilerin değeri toplum genelindeki (normal) dağılımın alt ya da üst ucundan geliyor olabilir (4. slayt) Bu bir normal dağılım. Toplumda serum HDL düzeyi ortalaması 50 ve standart sapması 9 olan normal bir dağılım gösteriyor. Bu toplumdan seçilen ve çalışmaya dâhil edilen bir kişinin ilk ölçümünde HDL değerinin 30 olduğunu varsayalım yani bu kişi dağılım eğrisinde alt sınıra yakın bir yerden geliyor. Bu kişinin HDL değerleri tekrarlayan defalarda ölçüldüğünde elde edilen değer farklılık gösterecektir. Bu durumda başlangıç düzeyi ortalamaları ile kıyaslandığında izlemlerden elde edilen ortalamaların toplum ortalaması olan 50'ye doğru yaklaşması söz konusu oluyor. Dolayısıyla, istatistiksel bir durum, tekrarlayan durumlarda grup ortalamalarının göreceli olarak toplumun ortalamasına doğru yer değiştirmiş (yükselmiş) gibi görünebilir. Bu durum çalışılan durumlar arasında (ön test, son test gibi) elde edilen ölçümler korelasyonla da ilişkili tabii.

Neden? Bunun açıklaması ortalamaya olan eğilim ('regression to the mean' veya 'regression towards the mean' denen durum). Bunun istatistiksel bir durum olduğunu bilmemiz lazım. Tekrarlayan ölçümlerde ilk ölçümde çok düşük ya da yüksek değer alan kişilerin değerleri ortalamaya doğru yer değiştirdiği için grubun ortalaması, toplum ortalamasına doğru (aşağı ya da yukarıya doğru olabilir) gidecektir, bunun olasılığının farkında olmamız ve bu tür sanal bir değişikliğin

grupların zaman içindeki karşılaştırmalarında bizi yanıltmasına izin vermememiz lazım.

Ortalamaya regresyon hangi durumlarda olur, kısaca özetleyelim: 1) Seçilmiş örnek toplumu temsil eden rastgele bir örnek değilse (dağılımın alt ya da üst ucunda kişiler örneğe çıkmış ise), 2) Başlangıçta yapılan ölçüm değerleri ile tekrar edildiğinde yapılan değerler kişi özelinde yüksek korelasyon göstermiyorsa (örneğin, kişinin başlangıç düzeyi kan basıncı ile test sonrası kan basıncı değeri sıkı olarak korele değilse), 3) Seçilmiş örnekteki kişilerin değeri topluma ait normal dağılım incelendiğinde ne kadar uç bir değer ise (alt ya da üst uçta) bu durumlarda ortalama regresyon daha belirgin olacaktır. Ancak, bunun grup düzeyinde bir fark yarattığı (örneğin grubun başlangıç kan basıncı düzeyi ikinci ölçüm grup ortalamasından farklı olacaktır) unutulmalıdır; kişiler özelinde bu farkın farklı doğrultuda olabilmesi mümkündür. Bu farkın rölatif bir fark olduğunu, azalma ya da artma yönünde olabileceğini, ya da iki ölçüm yapıldığında elde edilen değerlerin normal dağılım içindeki yerine göre bazen pretest bazen posttest ortalama değerlerin bu tür bir fark yaratabileceğini de hatırlatmalıyım. Bu tür istatistiksel yanılgıdan kurtulmak için kişiye ait ölçümlerde tek bir değer almak yerine, örneğin kan basıncı değerinde olduğu gibi 3 ölçüm yapıp, son ikisinin ortalaması alınabilir. Bu durumda kişinin bir değeri çok yüksek veya düşük de olsa, bunun grup karşılaştırmaları üzerindeki etkisi daha az belirgin olacaktır.

Şimdi bu makalede kısaca yöntem olarak ne yapmış ve neler bulmuşlar onları tartışalım. 10.000 hastadan oluşan sanal bir hasta topluluğu yaratılmış. Seçilen hastalık migren ve 25 gr sumatriptan tedavisinin plaseboyla karşılaştırmasını yapıyorlar. Her hastaya rastgele belirli değerler atamışlar ve bu gruptan 432 kişi ve bu analiz için %90'lık güce denk gelen yeterli bir örnek büyüklüğü. Elli kez bu ana grup içinden 432 kişilik örneklem seçiyorlar. Bu 432 kişiyi rastgele olarak ikiye ayırıyorlar, 216'sına sumatriptan, 216'sına plasebo verdiklerini varsayıyorlar ve bir tedavi etkililiği bakıyorlar. Bu ilacın içerisinde ilacın farmakokinetik ve farmakodinamik etkilerinin yanı

sıra hastaya özel ve araştırmacıya ait bir takım değerlerin çalışmaya eklenmesi ve bunların yanlı olması durumunda ne tür hatalar olabileceğini incelemişler.

5. slayt çok önemli değil ama hızlıca incelemek uygun olur; 25 mg'lık bir doz veriyorlar, farklı modellerden 25 mg için elde edilen tedavi ve advers etki profillerine bakıyorlar ve sonuç olarak tedavi değerini ve advers etkiyi 'kabul edilebilir sınır olarak belirledikleri' bir değerden bölüyorlar. Eğer advers etki güvenlik sınırının altında bir yerdeyse "Bu ilaç güvenlidir, advers etkisi yoktur" diye karar veriyorlar. Sonuçta, bu modellerden sonra sanal havuzda başağrısı devam etmeyen ve advers etkisi olmayan, başağrısı devam etmeyen ama advers etkisi olan, başağrısı devam eden ama advers etkisi olmayan ve başağrısı devam eden ve advers etkisi olan şeklinde 4 grup hasta ortaya çıkıyor. Önce Bayesian bir değerlendirme olmaksızın bir sonuç elde etmişler, ki bunu 50 kez yapıyorlar, sonra da 200 tane ayrı çalışmada da kör olma durumunu, open trial, double blinded ya da single blinded diye farklı farklı inceliyorlar. Randomizasyon şemasını değiştirerek ya da kişilerin izlemdeki kaybolma yüzdelarını değiştirerek ki bu izlemde rastgele kayıp da olabilir, advers etkisi nedenli hastanın ilacı bırakması olabilir, hasta iyileştiği için erken dönemde ilacı bırakmış olabilir. Farklı farklı yanlılık modelleri oluşturmuşlar. Ayrıca, plasebo ya da noseboya bağlı etkileri incelemişler, onun dışında eş zamanlı bir tedavi alıyorsa o tedavinin etkisi ve advers etkilerini farklılaştırmışlar modelde. 'Regression to the mean' için değişik öneriler getirmişler, en sonunda ölçüme bağlı hataların tedavinin etkinliğini ölçmede ya da advers etkinin boyutunu ölçmekteki hata düzeyleri ile oynamışlar. Bunu 50 örnek üzerinde 200 kez yapıp toplam 10.000 sonuç almışlar. Kısaca, aynı toplumdan seçilmiş 50 farklı örnek üzerinde farklı uygulamaları tekrar tekrar yapıp, topluma ait verilecek tahmini sonuçların üzerindeki etkilerini simule bir veri tabanında incelemişler. Bu kapsamda, 50 farklı örneğin her birinde hangi değişkeni değiştirmek istatistiksel olarak anlamlı bir fark yaratıyor diye bakmışlar. Buna göre, 50 örnekte de körleme değiştiği anda mutlaka çalışmaya bir hata girmiş. Eğer tedaviyi

ölçmede hata yaparsanız (yani, bilgi toplamaya bağlı taraf tutulursa) 49’unda anlamlı bir şekilde sonucunuzu etkilemiş. Plasebo sadece %36’sında elde edilen sonucu anlamlı şekilde etkilemiş. Advers etkileri yanlış ölçüyorsanız sonucu %66 sıklıkla etkilemiş. Randomizasyonun tek olarak iyi yapılmadığı durumlarda sonucun %26’sında etkilemiş. Yani, bu örneğe göre, randomizasyona çok önem veriyoruz ama körleme randomizasyondan daha çok etkiliyor randomize kontrollü çalışmaları, hiyerarşik olarak. Eğer böyleyse ve körleme etkiliyorsa değişik körleme durumlarında acaba sonuç farkediyor mu diye bakmışlar.

Hasan Yazıcı: Plasebo etkisi %36, regression to the mean 0. Peki regression to the mean etkisinden plasebo etkisini nasıl ayırt ediyorlar? Plasebonun iyi bir kısmı regression to the mean’de de var aslında, çok iyi bir kısmı. Plasebo etkisi diye tanımlanan hikâyenin bir kısmı regression to the mean. Onu ayırt etmek çok zor.

Banu Çakır: Tabii ayırt etmek çok zor. Lojistik modellemede tüm diğer faktörlerin etkisi kontrol edildikten sonra regression to the mean’e ait etki 0 çıkmış, belki daha bariz bir hata olurken, o gerilemiş olabilir. Bir de bu modeli 200 kere deniyorlar ve istatistiksel olarak anlamlı mı, değil mi diye bir tek modelde bakıyorlar. Bu tek modelde fark çıkmamış olabilir. Grupların başlangıç değerleri normal dağılıma uygun olabilir. Plasebo grubunda da regression to the mean söz konusu olabilir tabii. Bilemiyorum.

Çalışma çift kör olursa, hataya neden olabilecek en önemli basamak randomizasyon. Randomizasyon diğer tedavilerin etkilerinden ya da ölçüm hatalarından ciddi boyutta sapmaya neden oluyor sonuçlarda. O zaman çift kör bir çalışmada randomizasyonun doğru yapıldığını düşünüyorsanız diğer olası hata kaynaklarından çok çekinmemek lazım belki. Şu da ilginç, tek kör bir çalışmada sonucu etkilemede en önemli faktör plasebo. Randomizasyon 50 örnek içinde %14’lere düşmüş. Eğer açık bir çalışma ise ölçüme bağlı hatalar daha önemli. Ve plasebo etkisi önemli, %72’lerde. Ölçüme dayalı hatalara ait yüzdeleri toplarsanız, en önemli hata kaynağı gibi. Dolayısıyla tek kör ya

da açık çalışma söz konusu olduğunda randomizasyona bağlı hatalar göreceli olarak düşüyor.

Hasan Yazıcı: Tek kör çalışmada en önemli plasebo etkisi ama kimin kör olduğu çalışmada? Gözlemcinin mi, deneğin mi?

Sonuçta bu makale bize ne diyor, getiriyor? Randomize kontrollü çalışmada 2 tür önemli hata kaynağı var. Bir tanesini araştırmacı olarak bizim çok da engellememiz mümkün değil. Bunlardan biri plasebo etkisi, nosebo etkisi ya da hastanın bizim bilgimiz dışında ek tedavi alıyor olması. Ama bizim engelleyebileceklerimiz; randomizasyonun gizlenmesi var, körleme var ya da ölçüme bağlı hatalarımız var. Bugün itibarıyla randomizasyon önemli hata kaynağı. Ama çalışma çift kör değilse açık ya da tek kör bir çalışmada randomizasyonun önüne geçen başka hata kaynakları da var. Bunlara dikkat etmeliyiz. Dolayısıyla, yazarların geldiği nokta, bir çalışmada randomizasyon yapılmış ve körleme var” olması randomize kontrollü çalışmaların bilimsel olarak kısıtlılıklarını göstermek için yeterli değil. Belki bu ileride skorum sistemlerinde değişikliğe de neden olabilir ama her şeyden önce biz çalışmayı okurken bu basamaklara da dikkat etmeliyiz.

Randomize kontrollü çalışmalar için, en önemli getirisi elde edilen kanıtın kalitesi (iç geçerliliği), dolayısıyla klinik karar sürecine çok ciddi bir yararı var. Hâlâ altın standart, hem nedenselliği çalışmakta hem de farklı ilaç ya da yöntemlerin karşılaştırılmasında. Öte yandan, unutmamamız gereken birçok faktörün çalışmanın geçerliliğini etkileyebileceği, bunların içerisinde de etkileme açısından bir hiyerarşi olduğu. Ve her zaman vurgulamaya çalışıyorum, randomize kontrollü çalışmalar bittikten sonra mutlaka bu çalışmalardan elde edilen etkililik değerlerinin, gözlemsel çalışmalarla, büyük popülasyonlarla gerçek yaşam ortamlarında desteklenmesi lazım.

Bir Katılımcı: Arthritis&Rheumatism’de bir yazı vardı. 2x2 faktöriyel dizayn olarak tasarlanmış ve erken artritlerde 4 farklı grubu karşılaştıran bir çalışma. Şimdi burada başlangıçta, gruptaki kay-
bın %10 olacağı hesaplanmış bir randomize kont-

rollü çalışma ama çalışmanın sonunda baktığınız zaman %30'luk bir kayıp var. Şimdi ben %30'luk kaybı gördüğüm zaman çalışmanın sonunda aslında primary outcome karşılanmamış gibi gözüküyor, bir fark yok gruplar arasında. Secondary outcome'larda da sınırlı bir anlamlılık var. Şimdi sizin listeye baktığımız zaman bu izlemlerdeki kayıplar sonucu çok fazla etkiliyor gibi gözüküyor, %6-%8 arasında farklı farklı gruplandırmışsınız. Peki, gerçek hayatta mesela bahsettiğiniz kontrol hatalarından bir tanesi olarak kabul etmeden bu çalışmayı yapabilir miyiz? Randomizasyon sorunlu gözüküyor, kontrolde de sorun yok.

Banu Çakır: Yanlış yönlendirmek istemem. Bu çalışmada yaptıkları analizde, göreceli olarak birbirlerine etkilerine bakmışlar. İzlemdaki kayıp %20 ama rastgele olduğunda sonuç üzerindeki hata, %10'luk ama rastgele değil de sistematik bir kayıp olduğundan daha az sıkıntı yaratabilir. Örneğin izlemdaki kayıplar kullanılan deney ilacına ait advers etkiler nedenliyse, kayıp az olsa dahi sonuçlar ilacın etkili ve zararsız olmadığı sonucuna doğru hataya neden olabilir. O yüzden çalışmalarda izlemda kaybolanların hem kayıp nedenlerini hem de akıbetini öğrenmeye çalışıyoruz. 'Intention-to-treat' analizleri bu tür hataların etkilerini kontrol etmeye yönelik yapılıyor. Ama hiçbir zaman sabit bir yüzde verip, bunun altındaki kayıplar sorun yaratmaz demek mümkün değil. Dolayısıyla tek başına kaybın etkisi sizin çalışmanızda çok daha fazla çıkmış olabilir. %30'luk kayıp çok ciddi bir kayıp olabilir. Ama en önemlisi özellikle bu randomize kontrollü çalışmalarda zaten insanlar gönüllü, ulaşım bilgileri var. En azından akıbetlerini bilmek ve çalışmayı neden bıraktıklarını öğrenmek değerli.

Ahmet Gül: İki sorum var, birisi aslında tam soru değil. Bu modelleme yöntemiyle yapılan analizde regression to the mean etkisinin O çıkması aslında modellerken, iki grubu da karşılaştırırken büyük oranda çalışma grubunu seçerken dâhil etme ve hariç tutma kriterleri ile alakalı bir durum. Yani eğer siz protokolü uygun planladıysanız regression to the mean etkisini minimuma indirecek bir popülasyonu hedefleyerek de bunu kontrol edebilirsiniz. İki grubu karşılaştırdığında herhalde

O çıkmasına en azından bu modelin de katkısı var, diye düşünüyorum.

Asıl sorum şu. Bundan sonra maliyeti düşürmek için alternatif klinik çalışma modelleri gündemde. Bunların içinde adaptif yöntemler ve Bayesian yöntemlerle hasta alımının kontrolü ve randomizasyonun bunlarla kontrolü söz konusu. Yani baştan bir hedef koyuyorsunuz ve az evvel Şule'nin dediği gibi çalışma yürürken eğer beklediğinizin dışında kayıplarınız varsa yolda yürürken de hasta sayınızı adapte edebiliyorsunuz. Bu adapte çalışma yöntemi maliyeti düşürmede çözüm olarak tartışılan yöntem. Bu randomizasyonu ve sonuçları ne kadar etkiler?

Banu Çakır: Ahmet hocanın dediği şey aslında makalede tartışılıyor. Makalenin appendix kısmında da ciddi şekilde detaylı olarak neyi ne olarak modele koyduklarını söylemişler. Ben oraya çok girmedim ama tam da sizin dediğiniz gibi o konu vurgulanıyor. Biz regression to the mean'in zaten var olan bir şey olduğunu bildiğimiz için çalışmalarda sık olarak bunu da kontrol edecek tarzda bir ortalama değer üzerinden çalışmaya başlıyoruz. Bundan etkilenmemeye çalışıyoruz.

Randomizasyona gelince, randomizasyon yapıldığı an itibarı ile grupları benzer kılıyor. Ondan sonraki kısmı tabii ki benzer kılmıyor. Yani randomizasyon sonrası olan izleme bağlı kayıplar iki grubun karşılaştırılabilirliğini bozabilir. Dolayısıyla randomizasyon bu tür kayıplara bağlı olarak grupların karıştırıcı faktörler açısından karşılaştırılabilirliğini garantiye alamaz. Randomizasyon öncesi yapılan, olası kayıplara göre örnek hacminin minimum sayının üzerinde tutulmasını artırmaya yönelik düzenlemeler ise çalışmanın gücü açısından önemli.

Emmanuel Lesaffre: Bence burada iki konu var. Bir tanesi çalışmanın gücü ile Bayesin hassaslığı ve sorunları arasındaki ayrımı yapmak. Adaptif dizayn ya da sadece örnekleme büyütme gücü arttıracaktır ama tüm bu taraf tutmaları etkileyecektir. Regression to the mean hatası da öyle. Yani örnek artışının bunun üzerinde etkisi olmaz. İkincisi, regression to the mean, hastaları kendi-

leriyle karşılaştırdığınızda büyük sorun. O yüzden plasebo grubunda bu plasebo etkisini görüyoruz. Ama iki randomize grubu karşılaştırdığınızda birbirlerini sıfırlıyorlar. O yüzden 0 (sıfır) etki görmek sürpriz değil. Measurement ve plaseboyu hesaba kattığınızda etki yok olacak. Bir zaman sonra ortaya çıktığını görüyoruz ama çok sık değil. Çünkü bir grup başka bir grupta karşılaştırılıyor, ama gruplar içinde büyük bir mesele.

Bir Katılımcı: Randomize kontrollü klinik çalışmalarda randomizasyon ve körlemenin çok önemli olduğunu söylediniz, bunların kalitesiyle ilgili etik olarak, sonuçta herkes kendi çalışmasına randomize ve kördür diyor, randomizasyonla ya da körlemeyle ilgili, bunların kalitesiyle ilgili literatürde veri var mı? Her çalışma randomize kontrollü, gerçekten hepsi aynı kalitede mi, bunu ölçen bir sistem var mıdır, yani acaba gerçekten kör mü? Kör dediğiniz çalışmalarda aslında bir şekilde araştırmacı bazen kör olmayabiliyor, bunun derecesini tespit eden ya da randomizasyonda da bunları değerlendirme sistemi var mı?

Banu Çakır: Daha önceki toplantılardan birinde farklı randomizasyon tekniklerinin kullanıldığını tartışmıştık. Farklı randomizasyon tekniklerinin karıştırıcı faktörler açısından grupları benzer kılma yönünde göreceli olarak üstünlükleri var. Ama buradaki en önemli ayrıntı, okuduğunuz makalede bu randomizasyonun ne kadar doğru yapıldığı, körlemenin nasıl yapıldığı çok açık değil, ya da ikinci bir şey, randomizasyon yapılsa bile biz bunu Türkiye'deki çalışmalarda da çok görüyoruz, randomizasyonu yapan kişi randomize olarak grupları bölüyor ama bu çalışma aynı şekilde hastaya uygulanmıyor. Randomizasyonun hastaya uygulanana kadarki kısmında randomizasyondan bihaber olması ve kişilerin doğru tedavi gruplarına dâhil edilmeleri gerekiyor. Tabii bu konudaki teorik karşılaştırmalar var ama bunu makaleden anlamak çok kolay değil. Bilmiyorum siz ne düşünüyorsunuz ama özellikle kör konusunda. Yani onu nasıl değerlendiriyoruz, onun için raporların bu konuda mutlaka belki editörlerin de duyarlı olup raporda bu bilginin sağlanması lazım. Sorulabilir: "Randomizasyon nasıl yapıldı, kör olma nasıl sağlandı?" diye. Yoksa sadece çift kör

dendiğinde gerçekten hem hasta hem de hekim durumdan habersiz miydi, bir çalışmayı okuyarak bunu değerlendirmek mümkün değil.

Emmanuel Lesaffre: Randomizasyon lehine yorum yapmak isterim. Bence bunu gayet iyi açıklayabiliriz ama körleme sık sık problem olur. Örneğin, bizim Women's Health Initiative, östrojenin plaseboya karşı çift kör çalışmamız var. Ama birçok kadın bir süre sonra östrojen aldığını öğrendi. Eğer kadınlar bilmeseydi, hekimler bilecekti. Östrojenin kanama gibi yan etkileri yüzünden anında anlayacaklar. Bu çift kör çalışma diye adlandırılıyor ama 3 ay sonra artık çift kör değil. Sonra bir grupta bazı kardiyovasküler etkiler olur, diğerinde olmaz, körleme bu yüzden ortadan kalkar. Çift kör çalışma artık kör değil. Bu konuda dikkatli olmalısınız. Hekimler bu tür etkileri biliyor ve bu çalışmanın artık çift kör olmadığını söyleyebilirler.

Gülen Hatemi: Bu konu konuşuldu mu bilmiyorum ama randomize kontrollü çalışmalarda körlemeyi bozan şeylerden bir tanesi de daha önceki vizitlerin datasının sizin önünüzde açık olabilmesi. Çünkü bakarken önceki gözlemlerinizi sizde bunun ne olduğuna dair bir fikir oluşuyor ve o günlük gözleminizi çok etkiliyor.

Banu Çakır: Çok haklısınız, bir de bu çalışmayla ilgili, bu bir örnek çalışma, nispeten bilimsel olarak göreceli bir hiyerarşi getirdiği için ben sizinle paylaşmak istedim ama bu çalışmanın önemli sıkıntılarından biri de mesela buradaki etkiler hep birbirinden bağımsız değerlendirilmiş. Bunlar arasında etkileşimler olabilir. Mesela körleme ile randomizasyon arasında etkileşim olabilir. Çünkü randomizasyonun gizlenmesi ya da ölçmeye dair bir taraf tutmayı önlemek açısından körlemenin özel bir etkisi olabilir ya da dediğiniz gibi bunlar arasında etkileşimler de olabileceği için bu model varsayımlarıyla hiyerarşi tam çözülmemiş de olabilir. Ama benim bugün vurgulamak istediğim şey sadece randomizasyon var diye çalışma sonuçlarına tümüyle güvenmek yanlış. Yapılması gereken hem araştırmacının yaptıklarını açık ve detaylı olarak rapor etmesi hem de okuyucunun okuduklarını eleştirel açıdan değerlendirmesi. Bunu öğrenmemiz önemli diye düşünüyorum.

ROMATOİD ARTRİTE YAPILAN RANDOMİZE KONTROLLÜ ÇALIŞMALAR: BİZE SÖYLEDİKLERİ VE SÖYLEMEDİKLERİ

Yusuf Yazıcı

Romatoloji Bölümü
New York Üniversitesi
Amerika Birleşik Devletleri

Bu konuşmada randomize kontrollü çalışmaların bize anlatabilecekleri ya da anlatamayacaklarını örneklerle göstermek istiyorum. Ana hatlarıyla bakacak olursak, karşılaştırmalı etkinlik, bu etkinlik verilerinin anlatılmasındaki sorunlar, radyografik sonuçların analiz ve yayımlanış şekilleri, aynı şekilde güvenlik bilgilerinin analizi ve son olarak da bu güvenlik bilgilerinin aktarılışındaki sorunları özetlemeye çalışacağım. Bütün bunları anlatırken örnek olarak romatoid artrit (RA) çalışmalarını kullanacağım.

Karşılaştırmalı etkinlik günümüzdeki RA tedavisinde çokça söz edilen bir kavram. Birçok etkili biyolojik ve biyolojik olmayan ilacın kullanıldığı RA'da hangi ilacın ne zaman seçileceğini saptayabilmek, özellikle tedavi seçeneklerinin çeşidi arttıkça zor olabilmektedir. Bu zorluğu biraz olsun azaltabilmek için 'head to head' (kafa kafaya) karşılaştırmaların yapıldığı çalışmalara bakılabilir. Ancak bu tip çalışmalar genelde zaten birbirine yakın etki gösteren iki ilacı karşılaştırdığı için, anlamlı bir fark görebilmek ancak çok sayıda hastanın katıldığı çalışmalarda mümkün olabilmektedir. Böyle çok denekli çalışmalar gerek zaman gerekse de para açısından kolay değildir. Bir başka yöntem üstünlük değil benzerlik çalışması yapmaktır ama bu sefer de zaten birbirine benzer olduğunu bildiğimiz ilaçları bir RCT içinde denemek çok mantıklı gelmemektedir. Başka bir yöntem, yayımlanmış çalışmalara dikkatli ve doğru bir şekilde bakıp, oradaki verileri yeniden analiz etmek olabilir.

Yaptığımız bir çalışmada biz de bunu denedik. Birbirine olabildiğince benzer hastaların, yani hiç metotreksat (MTX) kullanmamış, erken RA hastalarının çalışıldığı çalışmaları topladık. Bunun sebebi aynı tip ve özellikle tedavi geçmişli olmayan hastalar kullanılmaz ise çalışmaların karşılaştırılması da daha zor ve bilimsel olarak yanlış olacağı içindi. Çalışmayı yaptığımız 2010 senesinde 4 ilacın bu tip çalışması vardı: etanersept (ETA), infliksimab (IFX), adalimumab (ADA) ve abatasept (ABA). Bu ilaçları ACR20, 50 ve 70 cevaplarına göre karşılaştırdığımızda NNT sonuçlara, yani remisyona ve 6 ay süreyle devamlı cevap oranlarına baktığımızda, ETA, ADA ve ABA halen birbirine

benzer ancak IFX kullanan hastaların diğerlerine oranla daha az cevap verdiğini gösterdik. Bu aynı zamanda ABD reçete veri tabanlarında da görülen sonuçlara benziyordu, %70 IFX kullanan hasta bir sene içinde dozu yükseltmek zorunda kalıyordu. Bunun yanında yan etki profilleri karşılaştırıldığında da IFX'in diğer 3 biyolojikten daha fazla yan etkiye sebep olduğu görüldü ki bu da başka yayınlarda gösterilmişti, hem bizim çalışmamızdan önce hem de daha sonra.

Bunun yanında ABA ile ADA arasında tedaviye yanıt açısından hangi karşılaştırılmaya bakılırsa bakılsın çok bir farklılık olmadığını gördük. Bizim çalışmamızdan 3 sene sonra tamamlanıp yayımlanan AMPLE çalışması MTX'a cevap vermeyen hastaları randomize olarak ADA ya da ABA'ya başlatıp iki ilacı karşılaştıran bir çalışmaydı. Bu çalışma da, 55 milyon dolar sonra, bizim yayımlanmış olan çalışmalardan çıkardığımız aynı sonuca ulaştı; ABA ile ADA arasında klinik yanıt olarak bir farklılık yoktu.

Başka bir yöntem de gerçek dünyada, yani rutin hasta takibi sırasında toplanan verilerin karşılaştırılması olabilir. 2011 senesinde New York Üniversitesi RA hasta veritabanında ilk defa bir biyolojik ajana başlayan hastaları çalıştık ve bu hastaların kullandıkları biyolojik ajana olan cevaplarına baktık. Yine 4 tane biyolojik ajan, ETA, IFX, ADA ve ABA en sık kullanılan ilaçlardı ve bir evvelki çalışmamızda bulduğumuza benzer olarak, ETA, ADA ve ABA arasında ilaca yanıt olarak bir fark görülmezken, IFX tekrar diğerlerine göre daha az yanıt gösteren tedavi opsiyonu olarak karşımıza çıktı. NYU'da rutin hasta takibinde kullandığımız RAPID3 ile yapılan analiz sonuçları ise, diğer çalışmada yaptığımız DAS28 ve ACR 20, 50 ve 70 sonuçlarından farklı çıkmadı. Bu da hangi outcome ölçeği kullanılırsa kullanılsın birbirlerine benzer sonuçlar verdiklerini gösterdi bize.

Bu değişik yöntemlerin hepsi de bize aynı sonucu verdi: ETA, ABA ve ADA arasında çok bir fark yoktu, IFX, özellikle başlangıç dozu olan 3 mg/kg dozunda, bu üçüne göre daha az etkiliydi. Bu çalışmalara benzer çalışmalar ile 'head to head' çalışma yapmaya gerek kalmadan karşılaştırmalı

etkinlik ile ilgili sorulara cevap bulmak mümkün ve çok da masraf gerektirmeyen çalışma şekilleri.

Yukarıda bahsettiğimiz benzer hastaların, özellikle MTX kullanmamış ve erken hastaların karşılaştırmada kullanılmasının önemi başka bir örnekle de gösterilebilir. Fostamatinib, 2010 yılında NEJM'da yayımlanan çalışmada MTX'a yeterli yanıt vermemiş hastalarda kullanıldığında plaseboya göre daha iyi olduğu gösterilmiş bir ilaç. Ancak yazıya dikkatli bakıldığında çalışmaya girmek için altı ülkeden (Bulgaristan, Kolombiya, Meksika, Polonya, Romanya ve ABD) RA hastaların en azından 3 ay boyunca, MTX'i 7,5-25 mg dozları arasında alıp, cevap vermedikleri zaman bu ilaç eklenmiş ve sonuçlara bakılmış. Hastaların Latin Amerika ve doğu Avrupa'dan gelenleri, ABD'den gelen hastalara göre daha iyi cevap vermişler. Bütün grupta zaten ilacı alan plaseboya göre anlamlı olarak çıkıyor ama hastaların nereden geldiklerine bakıldığı zaman, ABD dışından gelenler daha iyi cevap veriyor. Ancak yazıya bakıldığı zaman, ortalama MTX dozu hiçbir yerde verilmemiş. Yani birisi 7,5 mg MTX'i 3 ay alıp buna cevap vermiyor denip çalışmaya girmiş olabiliyor. ABD'de genelde MTX 20-25 mg'a cevap vermemişler çalışmalara girerken, başka yerlerde, belli sebeplerle, çok daha az MTX denemiş ve cevap vermedi denilen ama aslında gerçek cevabın bilinmediği hastalar çalışmalara girebiliyor. Tabii ikisi aynı tip hasta değil. Onun için MTX'a cevap vermemiş hastaların katıldığı çalışmalar karşılaştırılırsa fostamatinib de ETA ya da ADA kadar etkili diye düşünülebilir. Ancak daha sonra TNF'ye cevap vermeyen hastalar fostamatinib ile çalışıldığında, plasebodan farklı olmayan bir cevap görülüyor ki bildiğimiz gibi ETA ve ADA gibi biyolojik ajanlar TNF'ye cevap vermemiş hastalarda da etkili oluyor. Bütün çalışmalara bakıldığında fostamatinib, MTX'a cevap vermeyenlerde aynı gibi görülse de, genelde piyasada bulunan biyolojikler gibi etkili değil. Zaten daha sonra klinik çalışmalarına son verildi ve halen piyasaya çıkmış değil. Bu da ilaçları karşılaştırırken doğru hasta gruplarını kullanan çalışmaları karşılaştırmanın ne kadar önemli olduğunu bize gösteriyor, yani RA'da da bu doğru grup hasta MTX kullanmamış, erken dönemdeki hastalar.

İkinci üzerinde durmak istediğim radyolojik verilerin yayımlanış ve yorumlanış şekilleri.

Biyolojik ajanların kendi başlarına, MTX ile beraber ve sadece MTX alan hastalarla karşılaştırıldığı çalışmalar, ilk önce ETA ve IFX ile başlayıp daha sonra bütün biyolojik ajanlarla yapıldı. Bütün bu çalışmalarda benzer sonuçlar görülüyor, şöyle ki biyolojik ve MTX kombinasyonu, biyolojik ve MTX kendi başına alan hastalardan daha etkili, ancak klinik cevap olarak sadece biyolojik ve sadece MTX alanlar arasında büyük bir fark yok. Bütün çalışmalarda bu görülürken, yine bütün çalışmalarda radyolojik zararın durdurulması da yine en iyi kombinasyon alanlarda, ancak sadece biyolojik alanlar sadece MTX alanlara göre istatistiksel anlamlı olarak daha az radyografik zarar gösteriyorlar.

Bu çalışmalarda verilen radyolojik hasar skoruna baktığımızda bütün bu çalışmalarda Sharp van der Heijde skorunun kullanıldığını görüyoruz. Bu skor küçük eklemlerdeki hem eklem mesafesi azalmasını hem de erozyon sayısını ölçen ve skalası 0'dan 448'e kadar giden bir skor. Doktor tarafından görülebilen en küçük zarar skoru 5 ünite, klinik olarak anlamlı skor seviyesi de 15 ile 20 ünite arası olarak yayımlanmış. Bunları göz önünde tutarsak, bu çalışmalara baktığımızda neredeyse bütün çalışmalarda herhangi bir ilacı alan hasta grubunda ortalama skor hep 5 ünitenin altında, hatta çoğu zaman 1-2 üniteyi bile geçmemektedir. Hasta sayısının çok olmasından dolayı klinik olarak anlamlı olamayan bu ünite farklılıkları istatistiksel anlamlılık kazanıp sanki gerçek bir fark varmış gibi gösterilmektedir.

Bu çalışmaların radyolojik hasar için olasılık eğrilerine baktığımızda niçin bunun böyle olduğunu görebiliyoruz. Ortalama %80 hastada ne alırlarsa alsınlar hiçbir radyolojik hasar ilerlemesi görülüyor. Hastaların küçük bir kısmında ise az miktarda bir artma oluyor, eğer sadece MTX alıyorlarsa. Ancak tek tek hasta bakarken bu çok küçük sayıda hastanın kim olduğunu bilmediğimiz ve bilemeyeceğimiz için, bu radyolojik veriler de pratikte işimize yaramaz oluyorlar. Tedavi hakkında karar verirken sadece klinik cevap sonuçlarına bakarak ilaç seçmenin de böylece doğru bir hamle

olduğunu anlamış oluyoruz. İlaç alan hastalarda, bu ilaç ne olursa olsun radyolojik hasar görülme şansı çok düşük.

Son olarak da güvenilirlik ve yan etkiler hakkında konuşmak istiyorum. Yan etkiler randomize klinik çalışmalarla tam olarak ortaya çıkmazlar. RKÇ'lar yan etkilerin ortaya konmasında çok etkili olmazlar. Sadece genel bir fikir verirler ancak yan etki profili ilaç piyasaya çıktıktan ve çok sayıda hasta tarafından kullanıldıktan sonra belli olacaktır.

Yan etkilerden bahsederken ise iki tane olayı unutmamamız gerekiyor. Bir tanesi yan etkinin ilaca başladıktan sonra ne zaman ortaya çıktığı. Bir yan etkinin ilaca bağlı olup olmadığını en iyi tanımlayan bilgilerden biri ilaç başladıktan ne kadar sonra yan etkinin görüldüğüdür. Onun için örneğin bir sene sürececek bir çalışmada sadece en sonda görülen yan etkilerin, sadece sayıları yüzde olarak verilir ise tam bilgi verilmemiş olur. Klinisyenin görülen yan etkinin ilaca bağlı olup olmadığını anlayabilmesi için ilacın verilişinden ne kadar sonra bahsedilen yan etkinin görüldüğünü bilmesi gerekir. İkinci bir sorun da eğer bazı yan etkiler erken ya da geç görülüyorlarsa, yani rastgele görülüyorlarsa, o zaman uzun bir süre üzerinden verilen ortalama sayılar erken dönemde artmış olabilecek yan etki oranlarını da azaltmış olup sanki yan etki oranı olduğundan daha düşükmüş gibi gösterebilmektedir. Bunun en önemli örneklerinden biri TNF kullanan hastalarda kanser riskidir. Randomize çalışmalara bakıldığında bir senede görülen kanser riski genel toplumda beklenenden farklı çıkmamaktadır. Ancak TNF çalışmalarında görülen kanserlerin ne zaman görüldüğüne bakıldığında çoğunun ilk 3-4 ay içinde ortaya çıktığı görülmektedir. Eğer genel toplumda 4 ayda beklenen kanser sayısına bakılırsa o zaman bu erken dönemde TNF'lerin beklenenden daha fazla kansere sebep olma şanslarının arttığı görülebiliyor.

İkinci dikkat edilmesi gereken unsur 'patient years' (hasta yılı) kavramı. Bu da yine ancak bahsedilen yan etkilerin rastgele olduğunu biliyorsak kullanabileceğimiz bir yöntem. Onun da örneğini vermek gerekirse, diyelim ki bir yan etki

var ve sadece %20 hastada görülüyor ve sadece birinci ay görülüyor, sonra görülmüyor, kimse de ölmüyor, ilacı da bırakmıyorlar. Bu 100 hastayı bir sene takip ediyoruz. Bir sene sonra %80'inde bir şey olmuyor, %20'sinde oluyor. Yani yan etki riski 20/100 hasta yılı. Bu hastaları bir sene daha takip ediyoruz. Şimdi yan etki sadece 1. ayda ve sadece %20'de görüldüğü için, 2. senede kimse yan etki olmuyor, ama o hastaların 100 tane yılı daha toplanıyor, 2. senenin sonunda 20 /200 hasta yılı, yani 10, risk yarı yarıya indi, ama ilk ay için risk hiç değişmedi. O yüzden yan etkiler rastgele olmadığı zaman hasta yılı kavramını da kullanmamak lazım. En azından yan etkileri yorumlarken tek kullandığımız ölçüt olmaması lazım.

Özet olarak hem etkinlik hem de yan etkiler olsun, randomize çalışmalar hakkında düşünürken hangi hastaların analiz edildiği, sonuçların nasıl anlatıldığına ve söylenmeyenlerin neler olduklarına dikkat etmemiz gerekiyor. Ancak bu şekilde doğru ve güvenilir sonuçlara varabiliriz.

Kaynaklar

1. Yazici Y, et al. ACR 2011 #2233.
2. Klareskog L, van der Heijde D, de Jager JP, et

al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 2004;363:675-81.

3. Yazici Y, Yazici H. Tumor necrosis factor alpha inhibitors, methotrexate or both? An inquiry into the formal evidence for when they are to be used in rheumatoid arthritis. *Clin Exp Rheumatol* 2008;26:449-452.
4. Yazici H. Lymphomas after tumor necrosis factor antagonist therapy: comment on the article by Brown et al. *Arthritis Rheum* 2003;48:2389.
5. Bongartz T, Sutton AJ, Sweeting MJ, et al. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. *JAMA* 2006;295:2275-2285.
6. Paterson KR. Events per person year. Can obscure the true risk of certain adverse drug reactions. *BMJ* 1995;310:1470.
7. Yazici Y, Adler NM, Yazici H. Most tumour necrosis factor inhibitor trials in rheumatology are undeservedly called 'efficacy and safety' trials: a survey of power considerations. *Rheumatology* 2008;47:1054-1057.

TEDAVİ İÇİN GEREKLİ SAYI VE ZARAR ÇIKMASI İÇİN GEREKLİ SAYI İLE İLGİLİ SORUNLAR

Samy Suissa

*Klinik Epidemiyoloji
McGill Üniversitesi
Kanada*

Bana iki görev verildi. Biri “tedavi için gerekli sayı”, ikincisi “ölümsüz (immortal) zaman yanlılığı”. Önümüzdeki bir saat boyunca bunları konuşacağız.

Kalsiyum pirofosfat birikimi (yalancı gut) hastalığı ve bu hastalığın tedavisiyle ilgili EULAR kriterlerinin amacı, söz konusu tedavinin kanıta dayalı tıp kurallarına uygun gerçekleştirilmesidir. Bu bağlamda etkinliğin irdelenmesi tedavi için gerekli sayı (number needed to treat / NNT) ile etkiler ise RR (relatif risk) veya OR (odds ratio) ile gösterilir. Peki, NNT nedir?

Önce size, NNT’nin çok kısa bir tanımını yapıp hesaplamasını göstereceğim. Örneğin romatoid artrit hastalarında yapılan randomize kontrollü bir çalışmada yeni bir DMARD’ın kardiyovasküler ölüme etkisini sınıyoruz. Bunun için de hastalar ya söz konusu yeni DMARD’a ya da plaseboya randomize ediliyorlar. Diyelim ki bir yıl sonunda 100 hastadan 5’i, plasebo grubunda da 100 hastadan 9’u ölmüş. Peki, bu 5/100 ve 9/100 oranlarını nasıl karşılaştırırız?

Söz konusu oranları karşılaştırmada sık kullanılan bir ölçüt sıklık oranı (rate ratio) dır. Bu oran elimizdeki örnekte 5/100’ü, 9/100’e bölerek edilir ve 0.56’dır. Bu değer 1’den çıkartılırsa 0.44 kalır. Sıklık oranında azalmaya (rate reduction) deriz ki bunun anlamı yeni DMARD’ımızın örnek diye verdiğimiz çalışmada kardiyovasküler ölümcüllüğü %44 azalttığıdır.

İki oranı kıyaslamada sık kullanılan ikinci ölçüt ise sıklık farkı (rate difference) dır. Çalışma örneğine dönersek 5/100 ile 9/100 arasındaki farkı göstermek amacıyla bu kez 5/100’ü 9/100’e bölmek yerine ondan çıkartırız. Yani $9/100 - 5/100 = 4/100$ olur. Diğer bir deyişle klinik çalışmasını yaptığımız DMARD alanlar arasında, plaseboya kıyasla, kardiyovasküler nedenlerden 4 kişi daha az ölmüştür. İşte sıklık oranı ve sıklık farkı ölçütleri, gerek randomize kontrollü gerekse gözlemsel çalışmalarda etkiyi belirtmek amacıyla en sık kullanılan iki ölçüttür.

Esas konumuz olan NNT ise, ki kullanımı giderek yaygınlaşmaktadır, esasında sıklık farkına

dayanır ve sıklık farkının resiprokalinden farklı değildir. Örnek çalışmamızda sınıadığımız DMARD'ın NNT değeri $1/0.44 = 25$ 'tir. Peki, bu rakam bize neyi anlatıyor? Bir yıl boyunca bu DMARD ile tedavi edilen her 25 hastadan 1'inde ölümün engellenebileceğini gösteriyor. Yan etkileri değerlendirmek için kullandığımız zarar vermek için gerekli sayı (number needed to harm / NNH) tedavi için gerekli sayı ile aynı, ancak aradaki önemli nokta NNH'de sıklık farkının NNT'de olduğu gibi pozitif değil, negatif olması. Diğer bir deyişle NNH'de bir risk ölçüyoruz. Burada duralım ve NNT hakkında bilmeniz gereken her şeyin bundan ibaret olduğunu söyleyelim. Hesap çok açık ve literatürü tararken bu basit hesabın ne kadar kolay yapıldığını göreceksiniz. Ancak bu basit hesabı, bazen çok üst düzey dergilerimizde kullanırken, bazı sorunlar oluyor. Bu sorunların da temel nedeni NNT ve NNH hesaplarırken dikkat etmemiz gereken zaman faktörü. Yukarıda verdiğimiz örnekte DMARD grubunda 100 ve plasebo grubunda 100 hasta tam 1 yıl boyunca takip edilmiş. Ama gerçek hayatta yapılan randomize kontrollü çalışmalarda hastaların bazen eşit sürelerle takip edilmediklerini biliyoruz. Bazı hastalar çalışmaya çalışmanın sonuna doğru, bazıları ise daha erken girer, dolayısı ile takipler çoğunlukla farklılık gösterir. Ayrıca, birçok çalışmanın tekrarlayan olaylarla tanımlanmış sonuçları vardır. Örneğin romatoloji çalışmalarında sık olarak kullanılan atak sayısı böyle, tekrarlayan bir çalışma sonucu (outcome measure) dur. Üçüncü ve diğer önemli bir sorun çalışmayla sınıadığımız ilaç etkisinin zaman içinde değişebilmesidir. Şimdi NNT'nin bu 3 ayrı faktörden nasıl etkilenebileceğini görelim.

İlk olarak, kimi çalışmada hastalar eşit sürelerle takip edilmiyor. Grubumuz bundan evvel NNT hesaplamasında, süreyle ilgili sorunlar göz ardı edildiğinde nasıl yanlış yapılabileceğiyle ilgili birkaç yazı yayımladı. Kasım 2000'de yayımlanmış bir çalışma ile başlayalım; çalışmanın adı Kalça Koruyucu Kullanımıyla Yaşlı İnsanlarda Kalça Kırığının Önlenmesi (1). Bu çalışma 1:2 randomizasyon yöntemiyle yapılmış klasik bir randomize kontrollü çalışma. 650 hasta kalça koruyucu grubuna ve 1075 hasta ise kontrol grubuna randomize edilmiş. Çalışmanın sonlanım noktası kalça

kırığı ve makalede bir kalça kırığının önlenmesi için NNT hesaplanmış.

Sonuçlar 1000 kişi yılı (person years) bazında hesaplanmış. Özetle kalça koruyucu grubunda 13 kalça kırığı olmuş, kontrol grubunda (kalça koruyucu kullanmayanlar) ise 67 kalça kırığı görülmüş. Bu iki sıklığı 1000 hasta yılı üzerinden ifade ettiğimizde ise kalça koruyucu grubunda sıklık hızı 21.3/1000 hasta yılı, kontrol grubunda ise 46.0/1000 hasta yılı olarak hesaplanmış. Hesapların 1000 hasta yılı üzerinden yapılmasının nedeni açık. Bu çalışmada hastalar eşit sürelerle izlenmemiş. Çalışmada 6 ay süreyle izlenen hastalar olduğu gibi 2 yıl süreyle izlenenler de var. Yazarlar bu nedenle çalışma sonuçlarını belirli bir zaman dilimi bazında standardize etme gereğini duymuşlar. The New England Journal of Medicine'da yayımlanan bu makalede hastaların gerçek takip sürelerinin ne denli önemli olduğunu anlamak için yine aynı makalede verilmiş Kaplan Meier eğrisine bakmak gerek. Başlangıçta kalça koruyucu grubunda 653 hasta, kontrol grubunda 1148 hasta varken, 1 yıl içinde bu sayı anlamlı ölçüde düşmüş. Söz konusu eğride birinci yılda hastaların yaklaşık yarısı kalmış iken 2. yılda çalışmada neredeyse kimse kalmamış. Demek söz konusu çalışmada izleme süresi hastadan hastaya çok değişkenmiş ve NNT ile ilgili tartışmada bu çok önemli bir nokta.

Bu yazının 'Tedavi için Gerekli Sayı' adlı bir kısmı var. Burada 1 yılda 1 kalça kırığını önleyecek NNT'nin 41 ve 5 yıllık NNT'nin 8 hasta olduğu belirtilmekte. Bu da yazıda verilen sonuç. Şimdi bu rakamlarda olan yanlışla bakalım.

Bizim bu yazıyla ilgili yayımladığımız çalışmada (2) yaptığımız hesaba göre kalça koruyucu çalışmasına katılan hastaların ortalama izlenme süresi 1.1 yıl ve bir kalça kırığını önlemek için bildirilen bir yıllık NNT bu 1.1 yıllık ortalamadan geliyor. Soru şu; NNT gibi bir ölçütten bahsederken ortalama kullanabilir miyiz? NNT ortalamalara uygulanabilir mi? Cevap hayır; nedenine bakalım. Söz konusu makalede hesaplamalar şöyle yapılmış: kalça koruyucu grubunda 1000 kişi yıl başına kalça kırığı sıklık hızı 21.3, kontrol

grubunda 46.0 olduğunu hatırlarsak oran farkı 1000 kişi yıl başına -24.7 çıkar ve bu doğal olarak koruyucu bir değerdir. İşte yazarlar NNT hesabı için $1000/24.7 = 41$ bulmuşlar. Bununla da kalmamışlar, aynı NNT'yi bu kez 5'e bölmüşler ve bunun sonucunda da 5 yıllık NNT = 8'dir demişler.

Bu hesaplar neden yanlış? İlk olarak hesaplamada payda olarak tedavi edilmiş 1000 hasta ile 1000 hasta yılı aynı şeyler olarak kabul edilmiş. Peki, bunlar aynı mı? Birinde 1000 hasta alıyorsunuz ve bunların tamamını 1 yıl boyunca tedavi ediyorsunuz. Diğer taraftan 1000 hasta yılına 200 hastayı 5 yıl takip ederek veya 12.000 hastayı 1 ay tedavi ederek de ulaşabilirsiniz. Ancak böyle bir hesap NNT için kabul edilemez. NNT hasta yılına değil hastaya dayalı bir ölçüttür. Burada kullanılması gereken, irdelediğimiz örnekte de kalça kırığı kümülatif hızını (oranını) veren Kaplan-Meier eğrisidir. New England Journal'daki makale ilginç olarak bu Kaplan Meier eğrisini de vermiş. Bu eğride görüyorsunuz ki kontrol grubunda 1 yıllık kümülatif kalça kırığı hızı %5 iken, kalça koruyucu grubunda %2.1. %2.9'luk fark 1 yıllık NNT'nin 41 değil, $1/2.9 = 35$ olması sonucunu çıkarır. Şöyle diyebiliriz. Küçük bir hasta grubunda bu büyük bir fark değil ama maliyet açısından nüfus bazında politikayı oluşturacağınız nüfus seviyesinde, 1000 kalça kırığını önlemek için tedavi edilmesi gereken 35.000 hasta ile 41.000 hasta arasındaki fark tedavi ettiğiniz 6000 ekstra hasta olup, bu ekstra hastaları sadece bu sayının yanlış hesaplanmasından dolayı tedavi edeceksiniz. Bu eğrilere bakarak 2 yıllık tedaviyi sorabilir ve 2 yıllık NNT 19 olmak üzere %10 ve %4.5'lik tahminlerle hesaplama yapabilirsiniz.

Yazarlar 5 yıllık NNT'yi 8 olarak hesaplamış. Ancak bu çalışma 2 yıl sürmüştü. Yani 5 yıllık NNT nasıl olur da 2 yıllık çalışma ile hesaplanır? 5 yıllık NNT'yi hesaplamanız için 5 yıllık takip gerekir, dolayısı ile bu çalışmada elde edilen NNT 8 doğru değil. Beş yıllık sürede Kaplan-Meier eğrisinin nasıl olacağı hakkında bir fikrimiz olamaz, dolayısı ile bu makalede NNT hesaplanmamalıdır.

Bu önemli bir mesele ve bu yanlış bir çok yazıda görebiliyorsunuz. The Journal of Clinical Epi-

demiology dergisinde yayımladığımız makalede (3), New England Journal of Medicine, JAMA, The Lancet, BMJ, Annals of Internal Medicine ve diğer seçkin tıp dergilerinden örnekler sunuldu.

İkinci sorun ise tekrarlayan olayların tedavi sonucu olarak irdelendiği, hasta yılı ölçütünün kullanıldığı durumlarda görülüyor ve çalışma sonuçlarını değerlendirmekte çok yanıltıcı olabiliyor. Bu da, yine New England Journal of Medicine'da yayımlanan bir yazıda karşımıza çıkmakta. Bu çalışma, KOAH'da kombinasyon tedavisi, salmeterol ve flutikasonu karşılaştıran çok büyük bir çalışma olup tüm nedenlere bağlı ölüm irdelenmiştir (4). Yazının bir bölümünde ilaçların KOAH alevlenmelerine etkisi tartışılıyor ve tedavinin etkisi "alevlenmelerde %25'lik azalma ve 1 yılda 1 atağı önlemek için NNT 4" olarak ifade ediliyor. Oranlar total alevlenme sayısını kişi yıla bölerek hesaplanmış ve NNT = 4 elde edilmiş. Ancak bu hesap, bir önceki örnekte olduğu gibi kişi yılına dayanarak değil de belirli bir zaman diliminde izlenen kişi sayısına dayanarak hesaplanmalı. İşte biz bunu böyle hesapladık ve 4 değil 9 bulduk (3). Özetle 1 yıl boyunca tedavi edilmesi gereken 4 hasta ile 1 yıl boyunca tedavi edilen 9 hasta arasında büyük fark var.

Makalemizde (3) başka bir örnek yine NEJM'den ve transfüzyonlarla ilgili (5). Yazarlar 28 günlük NNT'yi 2 olarak hesaplamış, bu sayı aslında 5 ve bu çok büyük bir fark! Bu hesaplanan NNT'nin 2.5 katıdır, bunun maliyet hesaplamasında çok büyük bir etkisi olacaktır.

Ele almak istediğim son nokta ise bir ilacın zaman içinde değişiklik gösterebilen etkisi. Yusuf'un bugünkü sunumunda çok iyi bir örnek var, ilacın bazı yararlı etkileri erken ortaya çıktığı halde 6 ay sonunda etki kayboluyor. Bunun NNT üzerinde büyük bir etkisi var.

Bu söylediğime bir örnek NEJM'de yayımlanan randomize kontrollü bir antiplatelet çalışması. 30. günde tirofiban ile abciximabın karşılaştırmasında kompozit kardiyovasküler sonuç farklılığı gözlenmiş, yani 30 günlük NNT'yi hesapladığınızda abciximabın bir miktar hastada yararlı olabileceği

düşünülebilir (6). Ancak aynı çalışma uzun vadeli bir takiple yayımlandığında ve 6. aya baktığınızda bu etkiyi gözleyemiyorsunuz (7), yani 30 günlük NNT ile 180 günlük NNT çok farklı. O halde, özellikle ilaç çalışmalarında NNT'nin zaman içinde değişebileceği unutulmamalı. NNT'yi sabit bir değer olarak kabul edemeyiz.

KOAH'da 48 ayda bile uzun etkili bir bronkodilatör olan tiotropiumun etkisini gözleyebiliyoruz. Yusuf'un bahsettiği sorunların farkında olmalıyız, enfeksiyonlar erken dönemde gelişebilir ve eğer bir etkiyi erken veya geç ölçersek, çok farklı NNT'ler elde ederiz. Mekanizma hakkındaki klinik ve farmakolojik bilginizi kullanmanız gerekir.

Tekrar özetlemek gerekirse, NNT veya NNH bir tedavinin etkisinin yararlı ölçütleridir ancak bunların hasta yıla değil hastalara, belirli bir süre takip edilen hastalara dayalı olduğu unutulmamalıdır. Değişken takip sürelerini hesaba katmalıyız. Aslında bunu yapmak kolay, Kaplan-Meier eğrisi gibi NNT'yi hesaplamak için gerekli değerleri sağlayan bir ölçüt var elimizde. Bu eğri aynı zamanda hesaplamayı ne zaman bırakacağımızı da söyler; eğer Kaplan-Meier eğrisi 2 yılda kesilirse, 5 yıllık NNT'yi hesaplamamız da gerekmez.

Tekrarlayan sonlanımlı olayları ele aldığınızda durum çetrefilleşebilir. Bazı çalışmalarda sadece olayın ilk meydana gelişine bakılırken siz birden fazla olayı da kullanabilirsiniz ve bunu yaptığınız takdirde bu kümülatif oranı hesaplamanızı sağlayacak bazı istatistik modeller de vardır. Bir ilacın etkisinin zaman içinde değişebileceğini de unutmayın. 3., 6. ay ve 1. yıl NNT değeri çok farklı olabilir.

Hasan Yazıcı: NNT güven aralıklarıyla ilgili sorun var mı?

Samy Suissa: Normalde hiçbir sorun olmamalı çünkü güven aralığını oran farkıyla hesaplayabilirsiniz, güven aralığının tersini almak çok kolay ve matematiksel olarak kabul edilebilir bir yaklaşımdır olacaktır.

Gülen Hatemi: NNT'yi hesaplarken son iki takibi

en iyi ele alma yöntemi nedir? Tedavi amaçlı yaklaşımı mı kullanmalısınız, yoksa bunları paydadan atmayı mı?

Samy Suissa: İdeal olarak, tedavi amaçlı verileri kullanacaksınız ama iyi bir tedavi amaçlı çalışma yaparken hastaların tedaviye devam etmese bile sonuç için takip edildiklerini varsayarsınız. Bu günümüzdeki çalışmalarda gördüğümüz başlıca hatalardan bir tanesi. Bir hasta tedaviye veya plaseboya devam etmediğinde kullanılan terminoloji "takip dışı kalmış" olur, ancak takip dışı kalmışlardır, sadece tedaviyi kesmişlerdir. Bu hastalara takip dışı kalmış denmemelidir çünkü kliniğe geri gelecekler ve siz de sonuçları ölçmeye devam edeceksiniz. Tedavinin kesilmesi sonrası sonuçları ölçmezseniz tedavi amaçlı analiz için önemli bir yanlılığa yol açmış olursunuz. Bu nedenle sorunlu olan ve yanlılığın geliştiği bazı göğüs hastalığı örneklerim var. Sorunuza cevabım ideal tedaviye atanmış hastalar ilacı kesse bile tedavi amaçlı analiz yapmanız ve NNT hesaplaması için bu tedavi amaçlı verileri kullanmanızdır.

Gülen Hatemi: Prof. Yazıcı'nın belirttiği güven aralığı bazen negatif değerlere ulaşıyor.

Samy Suissa: Evet, aslında hız farkının istatistiksel olarak anlamlı olmadığı anlamına gelmekte. Haklısınız, tedavi edilmesi gereken hasta NNT – 4 yazılmamalı, bu anlamsız.

Yusuf Yazıcı: İstenmeyen olay fark oranına ve NNH'ye baktığınızda, NNH'nin negatif olması plasebo alan gruba göre tedavi alan grupta daha az istenmeyen olay geliştiği anlamına gelir.

Samy Suissa: NNH pozitif dediğimizde, maruz kalan gruptaki oranın maruz kalmayan gruptaki orandan daha yüksek olduğu anlamına gelir.

Yusuf Yazıcı: Yani fark negatifse, maruz kalan gruptaki oran maruz kalmayan gruptaki orandan daha düşük. Bunu yazdığımızda, bazıları negatif bir NNH'yi yorumlayamayacağımızı söylediler, aldığımız cevap buydu ve bunun hiçbir açıklaması yoktu, yapabileceğiniz kısa bir açıklama var mı?

Samy Suissa: Aslında bu konuyu karmaşık yapan “NNT” ve “NNH” terimleri. Bence “NNH” terimi hiç olmamalı. NNT, korumamız gereken tek terim. Eğer tedaviniz 1 istenmeyen olaya neden olduysa ve 1 yararlı sonucu engellediyse, buna da NNT denmeli. Sadece kullandığınız kelimeler farklı olacak: önlemek için NNT veya neden olmak için NNT. NNH terimi literatürden atılmalı. Bir keresinde NNH diyen birine karmaşaya yol açtığı halde bu terimi neden kullandığını sorduğumda sadece gülmüştü.

Bir Katılımcı: Kalça koruyucu çalışması hakkında soru sormak istiyorum. Kaplan-Meier eğrisine baktığımızda hastaların plasebo tedavi oranının çalışmanın başlangıcında 2:1 iken sonunda 5:1’e düştüğünü görüyoruz. Hayatta kalma analizinin eksik gözlemi vb. hesaba katmak için kullanıldığını biliyoruz ama hayatta kalma analizinin işe yaramadığı veya sonuçlarında kesinlikle hata vereceği bu eksikliğin belirli bir limiti var mı?

Samy Suissa: Evet haklısınız, örneğin bu çalışmada farklı bir takipten çıkma sorunu var ve çalışma 2:1 ile başlayıp 5:1 ile bitiyor. Birçok iyi çalışmada, iki grupta Kaplan-Meier bırakma zamanı eğrileri gösterilir ve bu eğriler özellikle zamanın bir noktasında ayrıştığı takdirde, 2 grupta körleme veya yan etki sorunlarının var olduğunu anlayabilirsiniz. Bu tarz ayırıcı ayrılma enformatif sansür yanlılığına ve bu da sağkalım analizi yanlılığına yol açabilir.

Banu Çakır: Kafam karıştı. Aslında, siz insidans hızı yerine kümülatif insidans kullanılmasını öneriyorsunuz, değil mi? Sadece bunun karışıklığa neden olup olmayacağını merak ediyorum. Çünkü epidemiyolojide ‘oran’ ve ‘risk’ terimlerini ayırt ederken ‘hız farkı’ veya ‘ağırlık oranı’ dediğinizde kişi zaman verilerini kullanıyoruz, diğeri için de ‘risk oranı’ terimini kullanıyoruz.

Samy Suissa: Aslında terminoloji karışık. Ben kendi öğrencilerime “kümülatif insidans” kullanırım, bu NNT için gerekli oranı bize verir. Ama klinisyenlerle konuşurken tanıdık oldukları bir ölçüt olmadığından hız terimini kullanıyorum, aslında kullanmamalıyım.

Banu Çakır: Doğru anlamış mıyım, kümülatif insidansın kümülatif insidansa bölümü. Ama olgu oranlarını elde etmek için Kaplan-Meier eğrilerinin veya bir çeşit sağ kalım analizi olan Cox regresyonunun kullanımı önerildiğinde, kişi zaman verileri kullanılıyor. 1 ile 5 yıllık sürede belli bir zaman noktasında insidans tahmini yapmak için kişi zaman verileri, NNT hesaplanırken ise aynı grafikten kümülatif insidans kullanılmalı. New England Journal of Medicine dergisinin sorununun bu olduğunu düşünüyorum çünkü ağırlık oranı kullanıyorlar ve NNT için ağırlık oranını bölüyorlar. Belki bu hepimiz için dergilerde vurgulanacak bir açıklama olabilir çünkü bizler onları insidans oranı, ağırlık oranı kullanmaya itecek ve daha sonra NNT için kümülatif insidans kullanmalarını istemek durumundayız, değil mi?

Samy Suissa: Bu doğru, işte bu günün dersi: NNT kişi yıla değil tedavi edilen ‘kişiye’ dayalıdır.

Kaynaklar

1. Kannus P, Parkkari J, Niemi S, Pasanen M, Palvanen M, Järvinen M, Vuori I. Prevention of hip fracture in elderly people with use of a hip protector. *N Engl J Med* 2000; 343:1506-13.
2. Suissa S. Calculation of Number Needed to Treat. *N Engl J Med* 2009; 361:424-25.
3. Suissa D, Brassard P, Smiechowski B, Suissa S. Number needed to treat is incorrect without proper time-related considerations. *J Clin Epidemiol* 2012; 65(1):42-6.
4. Calverley PM1, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW, Yates JC, Vestbo J; TORCH investigators. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med* 2007;356:775-89.
5. Lacroix J, Hébert PC, Hutchison JS, Hume HA, Tucci M, Ducruet T, Gauvin F, Collet JP, Toledano BJ, Robillard P, Joffe A, Biarent D, Meert K, Peters MJ; TRIPICU Investigators; Canadian Critical Care Trials Group; Pediatric Acute Lung Injury and Sepsis Investigators Network. Transfusion strategies for patients in pediatric intensive care units. *N Engl J Med* 2007;356:1609-19.
6. Topol EJ, Moliterno DJ, Herrmann HC, Powers ER, Grines CL, Cohen DJ, Cohen EA, Bertrand

M, Neumann FJ, Stone GW, DiBattiste PM, Demopoulos L; TARGET Investigators. Do Tirofiban and ReoPro Give Similar Efficacy Trial. Comparison of two platelet glycoprotein IIb/IIIa inhibitors, tirofiban and abciximab, for the prevention of ischemic events with percutaneous coronary revascularization. N Engl J Med 2001;344:1888-94.

7. Moliterno DJ, Yakubov SJ, DiBattiste PM, Herrmann HC, Stone GW, Macaya C, Neumann FJ, Ardissino D, Bassand JP, Borzi L, Yeung AC, Harris KA, Demopoulos LA, Topol EJ; TARGET investigators. Outcomes at 6 months for the direct comparison of tirofiban and abciximab during percutaneous coronary revascularisation with stent placement: the TARGET follow-up study. Lancet 2002;360:355-60.

SONSUZ ZAMAN YANLILIĞI

Samy Suissa

*Klinik Epidemiyoloji
McGill Üniversitesi
Kanada*

11 Temmuz 2007, en sevdiğiniz dergiyi açıyorsunuz ve şöyle ilginç bir çalışma görüyorsunuz (1). Birçok klinisyenin yaptığı gibi önce özete, hatta özetin sonuç kısmına bakıyorsunuz. Çalışma romatoid artrit (RA) hastalarında hidroksiklorokin kullanımının diyabet riskini azalttığını söylüyor, ilginç değil mi? Bunun kendi hastalarınız için harika olacağını düşünüyor ve bundan sonra çalışmanın tasarımına bakıyorsunuz. Yaklaşık 5000 RA'li hasta arasında ileriye dönük çok merkezli bir gözlemsel çalışma yapılmış ve çalışmanın sonucunda hidroksiklorokin kullanan RA hastalarında yeni diyabet ortaya çıkması risk oranının 0.62 olduğu anlaşılmış. Bu hidroksiklorokinin koruyucu etkisi olduğunu gösteriyor. Ayrıca bu sonucun güven aralığı çok dar ve 0.42 ile 0.92 arasında değişiyor. Bu da bu farkın anlamlı olduğunu gösteriyor. Özetle hidroksiklorokin alan hastalarda diyabet oranı %38 daha az. Makalenin özeti oldukça açık ve çok prestijli bir araştırmacı grubu tarafından çok prestijli bir dergide yayımlanmış. Şimdi, özeti bırakıp, makaleye dönelim, bu gözlem çalışmasının yöntemini irdeleyelim.

Makalede verilen tabloya baktığınızda ve hiç hidroksiklorokin kullanmamış kontrol referans grubuyla karşılaştırıldığında, kullanım süresine baktığımızda, 4 yıldan daha uzun süreyle hidroksiklorokin kullanan hasta grubunda diyabet riskindeki düşüşün çarpıcı olduğunu görüyorsunuz. Diyabet oranında %77 düşüş var. Makalede şu denilmek isteniyor: Ne kadar uzun süre kullanırsanız diyabet riskiniz o kadar düşer. Yani, özetle hastalarınız için iyi bir şey yaptığınızı düşünüyorsunuz; 21. yüzyılın sorunu olan diyabeti önleyorsunuz, muhteşem bir sonuç.

Konuşmamda şimdi size RA tedavilerinden ve kanser görülme sıklığından bazı örnekler vereceğim ve randomize çalışma sonuçlarıyla gözlemsel verileri karşılaştıracam. Daha sonra da sonsuz zaman yanlılığı kavramından bahsedeceğim.

İlk olarak RA tedavileri ve kanser insidansından bahsedelim. RA'da anti-TNF tedavisinin malignite riski üzerindeki zararlı etkilerine dair randomize kontrollü çalışmalar ve metaanalizlerden elde edilen veriler mevcut. Maligniteler bakımından, inflik-

simab ve adalimumaba bakarken, tüm bu çalışmalar bir araya getirdiğimizde bu 2 anti-TNF ajandan birini kullanan hasta grubunda 29 olay ve plasebo kullananlarda 3 olay geliştiğini görüyoruz, bu da 3.29'luk risk oranı veriyor, güven aralığı da istatistiksel olarak anlamlı. Yani, bu çalışmalarda anti-TNF alfa kullanan hastalarda malignite eğiliminde bir artış var (2). Aynı yazar sadece etanerseptle bu metaanalizi farklı bir şekilde tekrarlamış ve risk oranının 3.29 değil 1.84 olduğu sonucuna varmış. Bu fark istatistiksel olarak anlamlı değil ama yine de TNF alfanın yüksek bir kanser eğilimi olduğunu gösteriyor (3). Yine anti-TNF grubundan infliksimabın kullanıldığı bir KOAH çalışmasında kanser riski; çalışmada infliksimab alan 157 hastada 9 malignite, infliksimab almayan, plasebo alan 77 hastada 1 malignite geliştiği görülmüş. Bu çalışmada da muhtemelen yüksek kanser riski gösterilmiş (4). Wegener hastalığına baktığınızda, bu çalışmada etanersept kullanan hasta grubunda 6 vaka varken, plasebo kullanan grupta yeni vaka yok. Bu çalışma da bazı etkilerin varlığını gösteriyor (5).

Ama bu soruna yönelik gözlem çalışmalarına baktığımızda TNF alfa kullanan RA hasta grubunda risk artışı olmadığı bulunmuş (insidans oranı 0.9) (6).

Fred Wolfe da ABD kohort çalışmasında bu sorunu inceledi. Kohortunda en sık kullanılan ilaçlar infliksimab ve etanerseptti. Bu ajanlarla kanser çıkma odds oranı 1.0 olarak bulundu, yani bu gözlem çalışmalarında da bu ajanlarla herhangi bir risk artışı göstergesi yoktu (7). Bir sonraki analizinde metotreksat etkilerini ayırarak odds oranını yine 1.0 buldu (8).

Özetle randomize çalışmalarda kanser riski artışına doğru bir eğilim gözlenirken, gözlem çalışmalarına baktığımızda böyle bir etki görülüyor. Ortada bir çelişki var ve çalışmalar RA dışındaki farklı hastalıklarda yürütülmüş olsa da bu çelişki görülüyor.

Şimdi gözlem çalışmalarındaki “sonsuz zaman yanlılığı”ndan bahsedeceğim. Bunun için, örnek olarak SLE’de antimalaryal kullanımı ve malignite riski çalışmasını irdeleyeceğiz. Söz edeceğim

çalışma, bu konuya yönelik 1975 yılından bu yana İspanya’da tek bir merkezde yürütülen bir çalışma (9), 235 SLE hastada yapılmış, ileriye dönük uzun süreli bir takip çalışması. Hastalar SLE tanısı konmasıyla kohorta girmişler ve 31 yıl izlenmişler. On üç hastada kanser gelişmiş. Kanser sıklığı 156 antimalaryal alan ve 79 antimalaryal almayan hasta arasında kıyaslanmış. Verileri analiz ederken, kanser riskini etkileyebilen tüm potansiyel değişkenler için istatistik ayarlamaların yapıldığı, sofistike Cox orantısal risk modeli kullanılmış ve bir Kaplan-Meier eğrisiyle antimalaryal alan ve almayan hastalar arasında kanser sıklığı karşılaştırılmış. Kanser insidansı açısından, antimalaryal alan 156 hastadan 2’sinde ve antimalaryal almayan 79 hastadan 11’inde kanser gelişmiş. Cox modeli risk oranı 0.15 olup bu da kanser görülme oranında %85’lik bir düşüş demektir. Bu büyük ve inanılmaz bir düşüş, kanser neredeyse yok edilmiş.

Şimdi de çalışmanın tasarımına bakalım. 1975’ten bugüne SLE hastaları kohorta dâhil edilmiş. Kohorta girişi zamanları zaman t_0 olarak kabul edilmiş ve hastalar kimisi kısa kimisi uzun olsa da yaklaşık 30 yıl boyunca takip edilmişler. Takip sırasında 235 hastanın 13’ünde malignite ortaya çıkmış. Ancak burada malignitenin ne zaman ortaya çıkmasıyla ilgili önemli bir yanlılık var. Otuz yıllık çalışma süresinde herhangi bir zamanda hidroksiklorokin reçetesi alanlar ilacı kullananlar, böyle bir reçete almamış olanlar ise kullanmayanlar olarak kabul edilmiş. Ancak bu yanlı. Çünkü 30 yıl takip süresinin 20. yılında hidroksiklorokin reçetesi almış hastada kanser çıkıp çıkmamasıyla takibin hemen başında hidroksiklorokin reçetesi almış hastada kanser çıkıp çıkmaması kıyaslanamaz. Burada bu kişiye antimalaryal ilaç ‘kullanıcısı’ denmiş ama ilk reçeteleri bu tarihte verildiğinden, gerçekte kullanıcı değiller. İlk defa kullanana kadar kullanıcı kabul edilemezler. Söz konusu kohort içinde bir hasta 20. yıla kadar ilaç kullanmıyorsa, nasıl olur da kullananlar grubunda kabul edilebilir? Bu ilk sorun. İkinci sorun ise kohortun izlenmesi ilk malignite gelişene kadar olduğuna göre, ilaç reçetesi alana kadar malignite gelişmesi mümkün değil, burada bir malignite olsaydı hasta, kullanmayan grubunda olurdu. Yani aslında bu

dönem boyunca kanserden muafsiniz, kanser olamazsiniz, ölemezsiniz ve reçetenizi almak için hayatta olmanız gerek. İşte bu çifte sorun ‘sonsuz zaman yanlılığı’ dediğimiz şeye sebep oluyor. Çözüm basit, süreye bağlı analiz kullanmanız gerekiyor. Bunlar temel epidemiyoloji ilkeleri, maruziyeti kişiye değil, kişi-zamana göre yapmak. Bu çeşit durumlar için süreye bağlı analiz kullanılıyor. Farmakoepidemioloji alanında bu konuda ilaç etkilerini incelediğim birçok yazım var ve bu da birçok örnekten bir tanesi, diğer birçoğu statinler hakkında. Bu çalışmalarda statinler intihar, depresyon, kanser gibi birçok şeyden koruyormuş gibi gözüküyor. Ayrıca bir öğrencim sonsuz zaman yanlılığını açıklamak üzere BMJ’de çok güzel bir yazı yayımladı, statinler ve diyabet bağlamında olayı detaylandırdı (10).

RA hastalarında antimalaryal ilaç ve malignite konusuna değinmek istedik (11), çünkü SLE hastalarımıza ait veri yok elimizde. Önemli değil, çünkü amaç sadece sorunu ortaya koymak. Bu immün periyodun ilaç kullanılmayan olarak sınıflandırıldığı ve ilk reçete sonrası hastanın kullanan olarak nitelendirildiği süreye bağlı analiz yaparak konuyu inceledik. Ağırlık oranını 1.1 olarak hesapladık, yani ilaç kullanım süresi hesaba katıldığında antimalaryal kullanıcılarıyla kullanmayanlar arasında hiçbir fark olmadığını tespit ettik.

Antimalaryal tedavi lupus sağ kalımında süreye bağlı bir etki gösterebilir. Diğer bir çalışmada kanserden ziyade hayatta kalmaya bakılmış. Yine burada problem şu ki, antimalaryal kullanımı takip sırasında herhangi bir zamanda ilaca başlanması olarak tanımlanmış, yani bu hastaları ilk ilaç öncesi kullanıcı olarak tanımladığımızda, bunlar bu süre boyunca ölümsüz olacaklar. Neticede Kaplan-Meier eğrilerinde ilacın olağanüstü etkileri görülecek. Aslında 24 aydan daha uzun süre kullanımın etkisi mortalitede %50’lik bir düşüş gibi görülmekte ama bu doğrudan ölümsüz periyodun yanlış sınıflandırılmasıyla ilgili. Bu durum bu çalışmada sonsuz zaman yanlılığına neden olmuş.

Diğer bir örnek, bazı olağanüstü etkilerin bildirildiği Lumina çalışması (12). Karıştırıcılar için düzeltme yaptıktan sonra, hidroksiklorokin 0.12

risk oranıyla böbrek hasarına karşı koruyucu olduğu görülmüş. Bu da hidroksiklorokinle böbrek hasarı riskinde %88’lik bir düşüş demek ki bu gerçekten olağanüstü. Burada hidroksiklorokin kullanımı takibin herhangi bir anında ölçülmüş, yani elbette ilk reçeteye kadar korunmanız gerekiyor. İşte yine sonsuz zaman yanlılığı.

‘Sonsuz zaman yanlılığı’ nasıl tanımlarsınız? Daha önce söylediğimiz gibi, ilk olarak sadece özete bakarak bu yanlılığı saptayamazsınız, 4 yıl kullanımla diyabet insidansındaki %77’lik düşüş demek diyabet vakalarının dörtte üçünün bu sayede önlenmesi demektir, ki bu da inanılması zor bir sonuç. Bir şey inanılır gibi değilse ben ona inanmam. Bu makalede (1) sunulan Kaplan-Meier eğrilerinde, 4 yıldan daha uzun hidroksiklorokin kullanan gruba baktığımızda, ilk 5 yılda hiçbir diyabet vakası görmüyoruz. Biz özellikle RA’da Tip 2 diyabet insidansının ne olduğunu biliyoruz. Yani burada 5 yıl boyunca yeni diyabet vakası yok, bu mümkün mü? Maruziyeti 4 yıl kullanım olarak tanımladığımızda, hiçbir vaka bildirilmeyecektir ve biz buna ‘ölümsüz zaman’ deriz; işte sonsuz zaman yanlılığı. Dört yıllık kullanımın etkisini çalışmak istiyorsanız, tüm gruplarda vakaları 4 yıllık kullanım sonrası saymaya başlamalısınız, ki bu tuzağa düşmüş olursunuz.

İşte sonsuz zaman yanlılığı bu. Gözlem çalışmaları halk sağlığında ve klinik araştırmada çok önemli bir yer tutar. Randomize kontrollü çalışmalarda birçok sorun ele alınamaz, ilaç güvenliği sorunları gibi. Randomize çalışmalarda ilaç güvenliği için değil sadece etkinlik için güç hesapları yapılır. Dolayısıyla güvenlik değerlendirmesi yapmak ve bu bilgileri tamamlamak için büyük gözlem çalışmalarına ihtiyaç vardır. Örneklerimizdeki randomize çalışmalar anti-TNF tedavisi ile kanser riskinde hafif bir artış gösteriyor ancak gözlem çalışmalarında bunu görmüyoruz. Bunun nedeni sonsuz zaman yanlılığı olmasın? Bazı çalışmalarda, sonsuz zaman yanlılığına yol açan maruziyet tanımları kullanılır, bu da randomize kontrollü çalışmalarla uyumsuzluğu açıklar. İlaç maruziyeti yanlış sınıflamadan kaynaklanan sonsuz zaman yanlılığı ilaç etkisinin küçümsenmesine yol açabilir ve bu büyük bir sorundur. İlacın sonuç üzerinde

bir etkisi yoksa sonsuz zaman yanlılığı nedeni ile önleyiciymiş gibi görünecektir çünkü etkiyi daima olduğundan az gösterecektir, SLE ve antimalaryal örneklerinde olduğu gibi.

Maruziyeti sınıflandırma konusunda hiçbir bilgi sahibi olmadan sonsuz zaman yanlılığını nasıl ortaya çıkarırsınız? Bu hususta The Journal of Allergy and Clinical Immunology dergisinde yayımlanan bir makaleyi (13) yorumladık (14). Bu makalede, rinit tedavisinde nazal kortikosteroidlerin yeri incelenmişti ve “nazal kortikosteroid kullanan hastalarda daha az astım atağı ve astım nedeniyle daha az hastane yatışı gerçekleştiği” bildirilmişti. O halde, burnu tedavi etmek akciğerleri koruyor; steroidler bir şekilde akciğerlere ulaşmış koruduğu için, akciğerleri doğrudan tedavi etmek gerekmiyor gibi bir yorum yapılabilirdi. Bu çalışmada sonsuz zaman yanlılığı olduğunu gösterdik ve editörler de yazımızı yayımlayıp “Astım ataklarının nazal kortikosteroidlerle azaltılmasının inanılması zor bir durum” olduğunu ve bunun incelenmesi gerektiğini ve inanılmayacak kadar iyi olduğuna göre muhtemelen doğru olmadığını yazdılar. İnanılmaz ise inanmayın.

Evet, şimdi reklamlar. McGill Üniversitesinde bir yaz programımız var, şu anda Montreal’de devam ediyor. Farmakoepidemioloji alanında 4 dersimiz var ve bu kursları bitiren hiç kimse bu hatayı yapmıyor! Kesinlikle! İlginç bir şey de şu: Bu makalenin yazarlarından bazıları öğrenci olarak derslerimize geldiler ve çok şükür bu hatayı bir daha asla yapmadılar, çok teşekkür ederiz.

Bir Katılımcı: Sadece bir doktor olarak sonsuz zaman yanlılığını hastalığın şiddetine bağlı gelişen diğer yanlılıklardan ayırt etme konusunda kafam biraz karıştı. İki grubunuz var, biri hafif ve diğeri daha ciddi. Bence, sonsuz zaman sadece hastalığın önemli bir parçası.

Samy Suissa: Hastalığın şiddeti konusu bu çalışmalarda önemli bir husus, buna karıştırıcı denir. Epidemiyolojide temelde 3 tür yanlılık vardır, biri karıştırıcı faktör yanlılığı, ikincisi seçim yanlılığı ve üçüncüsü bilgi yanlılığı. Sonsuz zaman yanlılığı, bilgi yanlılığı ile birlikte oluşur, konuştuğunuz

şey ise karıştırıcı faktör yanlılığıdır. Benzerle benzeri mi karşılaştırıyorum, aynı ciddiyetteki mi farklı ciddiyetteki mi hastaları karşılaştırıyorum? Bu önemli bir sorun ama bundan önce maruziyet karşılaştırma gruplarınızı nasıl tanımlar, bunu açıklamalısınız. Eğer maruziyeti sonsuz zaman yanlılığına neden olacak şekilde tanımlarsanız, farklı hasta tiplerini de karşılaştırabilirsiniz, bu da hem sonsuz zaman yanlılığına hem de karıştırıcı faktör yanlılığına neden olabilir. İyi bir çalışmada bu sorunlara değinilmeli ve bu yanlılıklar ortadan kaldırılmalıdır.

Bir Katılımcı: NNT’den bahsederken hasta sürelerini bir gruptan eksiltip başka bir gruba ekleyememek gibi bir uygulamadan şikâyet ediyordunuz.

Samy Suissa: Belki tam açıklayamamış olabilirim. Hasta süresini eksiltmek tamam ve insidans oranını hesaplamak için doğru da. Ancak, bu oranlar sonradan NNT hesaplaması için kullanılamaz çünkü NNT hasta yıla değil hastaya dayalıdır.

Bir Katılımcı: Sonsuz zaman yanlılığına yönelik önerdiğiniz çare tam olarak da bunu yapmak, maruz kalmayan hasta yılının bir kısmını çıkarıp gerçek maruz kalmayana eklemek, peki bu tercih edilir mi?

Samy Suissa: İki sorun var. Maruziyet zaman içinde değişkenlik gösterirse bu yapılabilir, oranları elde etmek için uygundur da. Ama bundan NNT hesaplanamaz. Sonsuz zaman yanlılığının olmadığı maruziyetin zaman içinde değişmediği, randomizasyonda başladığı randomize kontrollü çalışmalarda NNT kullanılabilir.

Kaynaklar

1. Wasko MC, Hubert HB, Lingala VB, Elliott JR, Luggen ME, Fries JF, Ward MM. Hydroxychloroquine and risk of diabetes in patients with rheumatoid arthritis. JAMA. 2007;298:187–93.
2. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic

- review and meta-analysis of rare harmful effects in randomized controlled trials. *JAMA* 2006 May 17;295:2275-85.
3. Bongartz T, Warren FC, Mines D, Matteson EL, Abrams KR, Sutton AJ. Etanercept therapy in rheumatoid arthritis and the risk of malignancies: a systematic review and individual patient data meta-analysis of randomised controlled trials. *Ann Rheum Dis*. 2009;68:1177-83.
 4. Rennard SI, Fogarty C, Kelsen S, Long W, Ramsdell J, Allison J, Mahler D, Saadeh C, Siler T, Snell P, Korenblat P, Smith W, Kaye M, Mandel M, Andrews C, Prabhu R, Donohue JF, Watt R, Lo KH, Schlenker-Herceg R, Barnathan ES, Murray J; COPD Investigators. The safety and efficacy of infliximab in moderate to severe chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2007;175:926-34.
 5. Wegener's Granulomatosis Etanercept Trial (WGTE) Research Group. Etanercept plus standard therapy for Wegener's granulomatosis. *N Engl J Med* 2005;352:351-61.
 6. Fleischmann R, W Baumgartner SW, Weisman MH, Liu T, White B, and Peloso P. Long term safety of etanercept in elderly subjects with rheumatic diseases. *Ann Rheum Dis*. 2006; 65: 379-84.
 7. Wolfe F, Michaud K. Biologic treatment of rheumatoid arthritis and the risk of malignancy: analyses from a large US observational study. *Arthritis Rheum* 2007 ;56:2886-95.
 8. Wolfe F, Michaud K. The effect of methotrexate and anti-tumor necrosis factor therapy on the risk of lymphoma in rheumatoid arthritis in 19,562 patients during 89,710 person-years of observation. *Arthritis Rheum*. 2007;56:1433-9.
 9. Ruiz-Irastorza G, Ugarte A, Egurbide M, Garmendia M, Pijoan J, Martinez-Berriotxo A, Aguirre C. Antimalarials may influence the risk of malignancy in systemic lupus erythematosus. *Ann Rheum Dis* 2007;66:815-7.
 10. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ* 2010;340:b5087.
 11. Bernatsky S, Clarke A, Suissa S. Antimalarial drugs and malignancy: no evidence of a protective effect in rheumatoid arthritis. *Ann Rheum Dis*. 2008;67:277-8.
 12. Guillermo J. Pons-Estel, Graciela S. Alarcón, Gerald McGwin, Jr, Maria I. Danila, Jie Zhang, Holly M. Bastian, John D. Reveille, Luis M. Vilá. Protective Effect of Hydroxychloroquine on Renal Damage in Patients with Lupus Nephritis: Data from LUMINA, a Multiethnic U.S. Cohort. *Arthritis Rheum*. 2009; 61: 830-839.
 13. Corren J, Manning BE, Thompson SF, Hennessy S, Strom BL. Rhinitis therapy and the prevention of hospital care for asthma: a case-control study. *J Allergy Clin Immunol*. 2004 ;113:415-9.
 14. Suissa S, Ernst P. Bias in observational study of the effectiveness of nasal corticosteroids in asthma. *J Allergy Clin Immunol* 2005;115:714-9.

ETKİ TAHMİNLERİ VE ŞANSIN ROLÜ

Önder Ergönül

Etki tahmini (effect estimate) olarak bilinen odds ratio, relatif risk ile birlikte güven aralığı, P değeri kavramlarına değinmek istiyorum. Biraz da dikkat çekmek için karşılaştığımız bazı sorunlardan yola çıkarak farklı detaylardan söz etmek istiyorum.

Epidemiyolojide klinikte çalıştığımız pek çok faktör var ve bu faktörlerden acaba hangisi önemli, diğerlerinden farklı, bütün meselemiz bu, bunu anlamaya çalışıyoruz. Bu açıdan bakılırsa 1600’lerden beri başlayan, modern dönemde Avrupa’da herkes fataliteden başlayarak, kim ne kadar yaşıyor, ne kadar sağ kalıyordan çıkıp, ilaçların etkisi bugün hâlâ fataliteyle devam eden bir serüvenimiz var. Bu açıdan epidemiyolojinin atılım yaptığı dönemler var. 1950-80 bana hep ilginç gelir, kohort çalışmalarının ortaya çıkması. Kohorttan kısaca bahsedeceğiz. Ve tabii 2000’lerden sonra da epidemiyoloji popüler bir alan. Neden-sonuç ilişkisinin kurulması, tavuk-yumurta meselesini, bağlamak en zor şey. Epidemiyoloji belki enfeksiyondan çıktı ama artık daha geniş olarak ele alınıyor. Bir neden-sonuç ilişkisi diye de söyleyebiliriz. Epidemiyolojinin çok farklı tanımları vardır ama en geniş tanımı belki neden ve sonuç ilişkisini araştıran bilim dalı olması. Nedenin ya da etkenin, sonucu ne kadar belirlediği de çok önemli. Bunun için bazı ‘effect estimate’ denen etki tahminleri kullanıyoruz. Bütün bu tartışmalar, yanlılık, seçici yanlılık (selection bias), confounding, bütün hepsi daha iyi bir sonuca gidebilmek için. Peki, ne yapıyoruz esas olarak? Şansın rolüne bakıyoruz. Yanlılığa, yanılma ihtimaline bakıyoruz. Burada tabii aradaki ilişki ne kadar güçlü, çoğunlukla bununla ilgileniyoruz. Sonuçta doz etki ilişkisi var mı acaba, diye. Yani ilaç dozu gibi düşünmeyelim, etkinin gücü sonucu nasıl belirliyor gibi düşünüyoruz, bakıyoruz. Bu kavramlara ve günlük hayatta kullanımlarına değineceğim.

Prevalans ve insidans kavramları zaten kullandığımız kavramlardır. Kısaca özetleyeyim. Prevalans ve insidansı anlamak şu açıdan önemli, kohort çalışmasında veya “incidence rate” denen kavramların anlaşılmasında önemli. Prevalans, bir resim çekmedir. Örneğin şu an lepra prevalansı desek Türkiye’de belki 2000 lepra vardır. Bu prevalans bize bir rakam verir, ama insidans

zamana bağlı değişen bir kavramdır. Mesela lepra insidansı 0 (sıfır)dır diyebiliriz. Gerçi son dönemde lepranın hortladığını söylüyorlar. Ama hemen hemen 0 sayabiliriz leprayı. Demek ki insidans zamana bağlı bir kavram. Ama birbirlerine de aynı zamanda dönüştürülebilir kavramlar. Yani zamanı insidansla çarptığımızda bir şekilde prevalansa da, yani toplumda var olan bütün hasta sayısına da ulaşabiliriz.

‘Incidence rate’ dediğimiz şey açık olsun diye üzerinden geçiyorum tek tek. Sağkalım veya insidans çalışmalarında çoğu kez bir kişinin çalışmaya toplam katıldığı süreyi toplayarak elde ettiğimiz person time’ı kullanıyoruz. Yine ilginç olarak şunu söyleyebiliriz, daha eski terminolojide kohort çalışmalar insidans çalışması olarak bilinirdi. Yani bir noktadan başlayarak, etkenden sonuca doğru giden ve zaman içeren çalışmalar olarak.

Ve risk kavramı önemli. Epidemiyolojide daha çok risk sözcüğü kullanılmakta. İstatistikçiler buna şans da diyebilirler. Yani istatistik ve epidemiyolojide belli temel kavram farklılıkları, isimlendirilmelerde farklılıklar olabilir. Belli bir sürede izlenen katılımcılarda saptanan etkiler. Çok basit bir formül, ama zaman içinde kullanımı karışabiliyor. Tabii ki riskin aslında bir probabilité olduğu- nu yani bir olasılık olduğunu unutmayalım.

Relatif Risk Nedir?

Relatif risk aslında bir etkene maruz kalanlardan başına ‘outcome’ gelenlerin, etkene maruz kalmayanlardan yine başına ‘outcome’ gelenlere oranıdır. Bu kadar basit. Outcome dediğimiz çoğu zaman ölüm, hastalık, tedavi başarısı gibi kullanılabilir. Dolayısıyla relatif risk dediğimizde satır bazında düşünürüz.

1. slayttaki 2X2 tablosunu satırlar bazında düşünerek relatif risk tanımını yapabiliriz. Vaka kontrol çalışması dediğimiz zaman ise, bu kez satır bazında düşündüğümüzde sonuca gidemeyiz. O zaman kolon üzerinden sonuca gitmek gerekir. Elimizde vakalar var, vakalardan yola çıkarak etkeni bulmamız gerekiyor. Bu kez toplam kişiler, diyelim ki hasta olanlardan bir etkene maruz kalanlara oranı,

2 X 2 Relatif risk tablosu

Su firması	Kolera olan	Kolera olmayan	Toplam
Southwark ve Vauxhall	844 (a)	166,810 (b)	167,654
Lambeth	18 (c)	19,115 (d)	19,133

Tahmini Relatif Risk

- S&V Riski: $a/(a+b) = 844/167.654 = \%0.53$
- L Riski: $c/(c+d) = 18/19.133 = \%0.094$

Tahmini Relatif Risk

- Relatif Risk: $\%0.53/\%0.094 = 5.35$
- Southwark ve Vauxhall suyu ile kolera ölüm riski 5.35 kat daha fazla.

1. Slayt

**Prevalans azaldığında odds oranı
relatif riske yaklaşır**

$$RR = \frac{a / (a+b)}{c / (c+d)} = \frac{a / b}{c / d} = \frac{ad}{bc} = OR$$

	Hastalık	Hastalık yok
Etken var	a	b
Etken yok	c	d

2. Slayt

hasta olmayanlardan da bu kez yine etkene maruz kalanların oranını, yani ters yönde düşünürüz. Zamansal açıdan bir terslik denemez ama düşünce sistematığı açısından ters yönde bir gidiş olabilir. Esas mesele yine dört göznlü tablodur. Bunu şunun için vurguluyorum, buradan relatif riski ürettik ve kohort çalışmalarında kullanıyoruz. Odds oranına geleceğim. Odds oranını da vaka kontrol çalışmalarında kullanıyoruz. Snow, Londra’da kolera bakterisinin henüz bulunmadığı dönemde, ortada hiçbir bakteri yokken, tamamen matematiksel, epidemiyolojik işlemlerle koleranın kaynağını bulan kişidir. Bu açıdan epidemiyoloji tarihinde önemlidir. Henüz Pasteur’ün tarihe çıkmadığı dönemde düşünün. 1850’lerde. İşte pompanın sapı diye bilinen hâlâ bir şey var. Su pompası ve o belli bir pompadan su pompalayan özel şirketler, Londra’da o zamanda özel su şirketleri varmış. Hastalanıyorlar bir miktar. Ama oraya uzak bir şirketin suyunu içenler daha az hastalanıyorlar. İşte toplam insanların sayısını hesapladığımızda ölen, kolera olan olmayan, koleralıların %80’i ölüyor zaten, bir relatif risk hesaplaması yapabiliyoruz. Satır bazında, şu şirketten su alıp da kolera olanların, bu şirketten suyu alıp da kolera olanlara oranı çok basitçe relatif riski hesaplayabiliyoruz. Anlaşılabilir, mantığı gayet düz, herkesin çok rahat kavrayabileceği bir şey. Ama benzer mantığı bazen odds oranında yürütmek güç olabilir. Çünkü odds oranı buradan, türetilmiş bir başka şeydir. Gelelim odds oranına, daha çok tartışacağımız kısım. Odds oranı, relatif risk gibi, kullandığımız günlük hayatta çoğu kez, çok rahatlıkla sanki relatif riskmişcesine kullandığımız bir rakamdır. Ama aslında tam da öyle değildir. Ve burada tabii daha çok vaka kontrollerde kullanıyoruz, sözünü ettik, bir de çok değişkenli analizlerde rahatlıkla kullanıyoruz ama odds oranını sadece bir relatif riski yakınsayan birim olarak düşünebiliriz. Nasıl hesaplanıyor? Bir assumption’a dayanıyor odds ratio. Eğer ki bu hastalık ya da gördüğümüz durum çok nadirse ancak geçerli olabilir. Yani ‘rare disease assumption’ denilen durumun nadir olması durumunda relatif riskmişcesine kullanılabilecek bir veridir.

2. slaytta gördüğünüz relatif risk hesabı formülü, a ve c ne kadar küçükse, ikinci aşamada bunları attık diyelim formülden, daha doğrusu ihmal

Olma olasılığı

- p = olasılık (probabilite)
0 ve 1 arasında
- Başarının olasılığı:
 $\Pr(y = 1) = p$
- Başarısızlığın olasılığı:
 $\Pr(y = 0) = 1 - p$

3. Slayt

Relatif risk

	DVT	No DVT
Heparin	8	92
Plasebo	18	82

$$\text{Risk}_{\text{heparin}} = 8/100 = 0.08$$

$$\text{Risk}_{\text{plasebo}} = 18/100 = 0.18$$

$$\text{Relatif risk} = \frac{\text{Risk}_{\text{plasebo}}}{\text{Risk}_{\text{heparin}}} = \frac{0.18}{0.08} = 2.25$$

4. Slayt

Odds oranı

	DVT	No DVT
Heparin	8	92
Plasebo	18	82

$$\text{Odds}_{\text{heparin}} = 8/92 = 0.087$$

$$\text{Odds}_{\text{plasebo}} = 18/82 = 0.22$$

$$\text{Odds oranı} = \frac{\text{Odds}_{\text{plasebo}}}{\text{Odds}_{\text{heparin}}} = \frac{0.22}{0.087} = 2.53$$

5. Slayt

ettik. İhmal ettikten sonra elde ettiğimiz aXd / bXc = odds oranı.

Kullandığımız rakam budur. Demek ki bu formülasyon, hastalığın ya da neyse buradaki outcome, onun çok nadir olması varsayımına dayanıyor.

Odds bir olasılık tahminidir. Olma ihtimaline p dersek, olmama ihtimali $1-p$ olacak.

3. slaytta görüldüğü üzere başarı oranı p ise onun tam tersi başarısızlık oranı, $1-p$ şeklinde olacak. Nihayetinde ben aslında odds ratio'yu anlatırken şöyle anlatıyorum. Örneğin bu bardak 4 birimden oluşuyor, bunun $1/4$ 'ü doluyorsa boş olanın dolu olana oranı 3'tür. $3/4/1/4$, 3 olarak ifade edilebilir. Ama bakın dikkatinizi çekmek isterim, proportion'dan yani yüzde gibi bir ifadeden farklıdır. Relatif riski daha çok bir yüzde ifade gibi kullanabiliriz. Ama odds oranında bunu kullanamayız. Yani bir şeyin bir şeye oranı gibi. Kaçakaç oranı gibi de söylemiştik bir ara ama pek kabul gören bir tabir olmadı.

Benzer bir şeyden, diyelim ki derin ven trombozunda heparin kullanımına bakıyoruz, heparin ve plasebo. Relatif risk 2.25 çıktı diyelim (4. slayt). Devam edip odds oranını hesaplırsak 2.53 (5. slayt). Ama buradaki rakamlar biraz düşük olduğu için, diğer tarafa göre epeyce farklı olduğu için, 2.25 ile 2.53 belki kabul edilebilir bir yakınsama. Ama bu oran yükselirse bu rakamların arası açılacaktır. Yani relatif riskten odds oranı giderek açılacak. İşte bunun BMJ'de yayımlanmış tablosu (6. slayt).

Odds oranı belli bir sayının üzerinde olursa, örneğin 6'nın üzerinde bulduğumuz zaman, relatif riskten epeyce açılmış olacaktır. Mesela bazen, bazı çok iyi dergilerde bile rastlıyoruz. Özellikle örneklem sayısı düşükse, mesela odds ratio 40 ise, ya örneklem sayısı çok düşüktür ya da başka bir hata vardır. Genelde örneklem sayısı ya da çalışmanın gücünün düşüklüğünden kaynaklanır. Belki bir asistan arkadaşımız odds oranını 15-20 bulduğunda heyecanlanır, halbuki oradaki odds oranı gerçeği temsil etmekte epeyce bir uzaklaşmış demek anlamına gelir. Biz acaba odds oranını

Risk ve Odds karşılaştırması

Risk	Odds
0.05 veya %5	0.053
0.1 veya %10	0.11
0.2 veya %20	0.25
0.3 veya %30	0.43
0.4 veya %40	0.67
0.5 veya %50	1
0.6 veya %60	1.5
0.7 veya %70	2.3
0.8 veya %80	4
0.9 veya %90	9
0.95 veya %95	19

6. Slayt

relatif riske çevirebilir miyiz? Bunu savunan bir görüş var epidemiyolojide. Odds oranı madem bu kadar saçma, bu kadar büyük çıkıyor, 40-50 gereksiz, kabaca. Belki kitaplarda yazmaz, ben kitaplarda odds oranının 6'nın üzerindeyse kullanmayın gibi bir ibare görmedim, ama böyle bir anonim bir şey vardır, 6'yı, 8'i geçmesin gibi bir yaklaşım vardır.

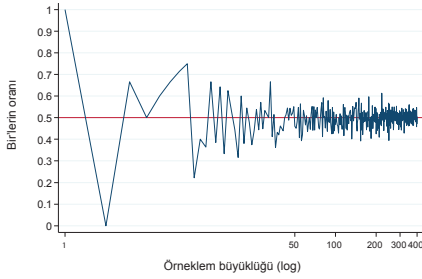
Çok basit bir formülü vardır. Bu 2 yazarın yapmış olduğu bir şeydir. Odds ratio'yu biliyorsanız odds ratio'dan türettiğiniz bir relatif risk ortaya atabilirsiniz. Ama bu sunidir. Tamamen makyaj meselesidir. Çünkü odds ratio'nun yüksek çıkması başka bir takım sorunlardan kaynaklanıyor. Ya confounder vardır, confounding yoksa zaten o zaman belki böyle çok değişkenli bir analiz yapmaya bile gerek yoktur. Doğrudan çok değişkenli analiz olmaksızın zaten confounding de yoksa sonucunuzu sunabilirsiniz. Demek ki, bazı toplantılarda da rastlıyorum, Türkiye'de yapılan toplantılarda, relatif riski odds ratio'ya çevirme eğilimi var ki bu doğru değil.

Başka bir şey risk difference/attributable risk. Bu da önemli bir kavram. Yani bizim bulduğumuz bir örnek üzerinden gideyim, zamanı da iyi kullanabilmek adına, bulduğumuz bir relatif riskin veya buna odds ratio'da diyebilirsiniz, acaba atfedilebilir önemi nedir? Toplumda özellikle halk sağlığı kavramında bu önemli, böyle kullanılabilir.

Relatif risk-1/relatif risk X 100 şeklinde bir formülle kullandığınızda aslında çok basitçe çıkarabilirsiniz. Başka sunumlarda da basitçe kullanıldı. Diyelim ki bulduğunuz bir relatif riski yüzde olarak ifade etmek istiyorsunuz, çünkü toplumsal düzeyde ya da başka insanlara anlatırken relatif riski anlatmak bazen zor olabiliyor. Bunu yüzdeye çevirmek için bu formülü kullanıyorsunuz.

Gelelim P değerine. Buraya kadar sadece etki tahmininden bahsettik. Yani odds ratio veya relatif risk veya hazard ratio da odds ratio ve relatif risk gibi. Bunlar bize sadece etkinin boyutunu anlatıyordu. Yani etki az ya da çok. Fakat acaba buna rağmen bu etki şans eseri mi yoksa şanstıan bağımsız mı? Yani biz bu olayı defalarca tekrarla-

Bir binom değişkeni için istatistik tekdüzellik ($p=0.5$)



7. Slayt

diğimiz zaman acaba aynı sonuca mı ulaşacağız yoksa bu sonucu çok tesadüfi mi bulduk? Bu açıdan bakıldığında sayıyı ne kadar arttırırsak, biliyorsunuz normal dağılım meselesini. O deneyi ne kadar çok tekrarlırsak o zaman şans faktörünü elemine etme durumumuz azalacaktır. Bu bilinen, tahmin edilen bir şey. Bizim uçlarda hesapladığımız yani, bu normal durumun dışına düşen noktalar aslında P değeri olarak ifade ettiğimiz rakamlar. Orayı bulmaya çalışıyoruz. Bazı tanınmış istatistikçilerin yaklaşımı P değerinin önemini anlam açısından bölmeye çalışırlar. Yani işte P değeri %1'in altındaysa çok güçlü bir bulgu olabilir. 1-5 arasındaysa moderate denebilir. %10'un üzerindeyse hemen hemen hiçbir şey yoktur gibi yorumlar yapılabilir. Tartışmada kullanılabilir belki ama çok güvenilir, tamamen P değeri üzerinden yorum yapmak ve bunu iletirmek pek mümkün değil. Zaten P değerini aslında kullanıma sokan, 1900'lerin başında Fisher, bir süre sonra bu P değerinden bıkmıştır. P değeri hekimlerde olan özellikle çok sık, P değeri fetişizmi denen bir kavram var. Baktığımızda hemen P değeri anlamlıysa, ki bizim Kırım Kongo çalışmalarında da sorun oydu. P değeri anlamsızsa hemen onu atmak, vazgeçmek, anlamlıysa çok şey yapmak. Ama Fisher'ın her zaman her şeyi p değeri ile açıklamayın diye bir yaklaşımı var. Bir de Altman'ın çok yerinde bir şeyle, mesela diyor ki p değeri 0.05 değil de varsayalım 0.055 oldu, 0.045 oldu yani ne olacak sonuçta, hani dünyanın sonu mu? Hep de sorulan bir şeydir. Getiriyoruz, getiriyoruz, eyvah diyor, 0.052 oldu, çok kötü, sunamayacağız bunu, etkisi çıkmadı. Hâlbuki zaten son yayınlarda anlamlıdır, anlamsızdır değil, doğrudan 3 basamaklı p değerinin sunulmasıyla yazılarda yer alıyor. Tabii örneklem sayısı büyüdükçe burada görüyorsunuz, çan eğrisinin farklı bir şekli aslında. Örneklem sayısı arttıkça istatistik düzenlilik daha da iyileşiyor ve 7. slaytta da görüleceği üzere ve elde edeceğimiz faktör şanstın daha bağımsız hale geçiyor.

Son dönemlerde bir başka şey, p değeri yerine acaba güven aralığı kullanabilir miyiz? Hakikaten bazı epidemiyoloji dergilerinde p değeri artık hiç verilmiyor. Doğrudan güven aralığı veriliyor. Güven aralığını kullanmak p değerinde ne kadar bize ne kadar yol gösterici olabilir? Olabilir, hipotez

Güven aralığının hesaplanması

$$L = \log\left(\frac{a \times d}{b \times c}\right)$$

$$SE = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

OR için %95 GA = $[\exp(L - 1.96(SE)), \exp(L + 1.96(SE))]$

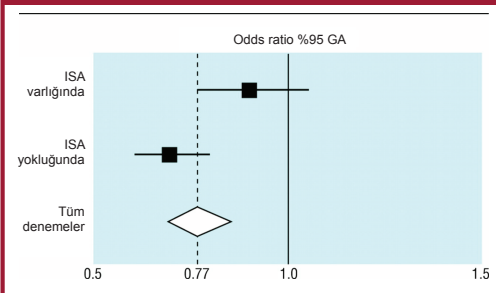
8. Slayt

testinde yani biz p değeri yerine güven aralığını kullanabiliriz. Yalnız şunu unutmayalım, güven aralığı hesaplaması 2X2 tablodaki kare üzerinden z tablosu temelinde z testi ile hesaplanmaktadır genelde. Bu önemli, birazdan karşımıza çıkacak bunun getirdiği problem. Sonuçta rahatlıkla artık güven aralığını p değeri yerine kullanabilmekteyiz. Güven aralığını hesaplamak için, kendi elimizle de hesaplayabileceğimiz formüller var. Az evvel gördüğümüz abcd, 4 gözlü tablodaki rakamları bilerek aslında kendimiz güven aralığını hesaplarız (8. slayt).

Buradaki standart error'ün 1.96 ile çarpıldığına da dikkat edelim. O da bazen başımıza işler açacaktır. Örneklem sayısı büyüdükçe güven aralığı daralacaktır. Bunu biliyoruz zaten. Örneklem sayısı ne kadar büyürse güven aralığı daralır, dolayısıyla bize daha güvenilebilir bir sonuç, yani tanım olarak düşünürsek bir deneyi 100 kez tekrarladığımızda 95'inde tek estimate dediğimiz odds ratio ve relatif risk alt ve üst sınırları içinde yer alacaktır. Tanım budur aslında güven aralığının tanımı budur.

Güven aralığı, %95 olasılıkla etki tahmininin arasında bulunacağı alt ve üst sınırları ifade eder. Gelelim gerçekten her zaman p değeri ve confidence interval'ı birbirinin yerine çok rahatlıkla kullanabilir miyiz? Daha doğrusu confidence interval'ı verdiğimizde p değerini gerçekten unutabilir miyiz? Meta analiz tablolarında görebiliriz. Şöyle bir uygulamadan da kurslarda bahsedildi. Eğer ki güven aralığı 1'i kesiyorsa, yani odds ile çalışırken bir şeyin bir katı kendisidir, dolayısıyla anlamsızdır. Etkisinin sıfır olduğunu gösteren bir şeydir. Güven aralığı 1'i kapsıyorsa bir şekilde, alt ve üst limitler 1'i kapsıyorsa o zaman istatistik açısından anlamsızdır, diyoruz. Bunu çok sık kullanıyoruz, çok rahat bir şekilde kullanıyoruz son zamanlarda. O yüzden zaten bazı dergiler güven aralığını veriyorsanız p değerini vermenize hiç gerek yok, zaten eğer ki güven aralığı içerisinde 1 yer alıyorsa o zaman bu istatistik olarak anlamsızdır diyorlar. Net bir şekilde bunu söylemek mümkün. Örneğin şu örnek kesinlikle anlamlı. 1'de oldukça uzaklaşmış, ama burada 1'i kapsadığı için p değeri kesinlikle anlamsız (9. slayt).

"1" i katıp katmamak



9. Slayt

Bir başka örnek ve bu örnekle nadir görülen bir durumu sunuyorum (10. slayt):

Gerçekten burada vakalar var, kontroller var, expose olan olmayan. Sonuçta p değeri anlamlı. Fakat güven aralığı içerisinde 1 kapsanıyor aynı zamanda. Yani normalde 1 buradaysa güven aralığına baktığımızda p değerinin anlamsız olması gerekiyor. Kesin olarak anlamsız yani, 0.05'in üzerinde olması gerekir. Ama bu nasıl oldu, gerçek bir veri ve vaka kontrol çalışması yapılmış. Böylesi bir durumla karşılaştık, bunu nasıl rapor edeceğiz? Biz bunu sadece güven aralığı üzerinden rapor etmiş olsaydık bu sisteme, o zaman p değeri anlamsız olmuş olacaktı. Ama biliyoruz ki p değeri de gerçekten anlamlı çıktı. İşte bu bize şunu anlatıyor, confidence interval'i hesaplamanın çok farklı yöntemleri var. Demek ki farklı farklı yöntemler var, mesela lojistik regresyon yaptığımızda, Woolf hemen hemen tüm programlarda kullanıyor lojistik regresyon. O zaman biraz daha geniş bakın. Woolf'da, aynı çalışmada güven aralığı, p değeri anlamlı çıkmış olsa bile güven aralığı biraz daha geniş ve 1'i kapsıyor. Demek ki lojistik regresyonda veya başka softwarelerde kullandığımız şeylere dikkat etmek durumundayız. Her zaman en azından şunu söylememeliyiz; "Güven aralıklarına bakarsak bu zaten anlamsızdır, p değeri ile eşit olarak kullanılabilir." Belli bir rezerv payı buna koymalıyız. Genelde belki %99 bunu haklı bulacağız ama bunun gibi kritik durumlarda da bir farklılık söz konusu olabilir. Sonuçta hakikaten p değerini kullanırken dikkatli olmalıyız p değeri fetişizmine düşmemek için. Düşük sayılı vakalarda özellikle, örneğin bu yürüyen Kırım Kongo tartışması tamamen bunun üzerine gidiyor. Çok sayıda arkadaşımız hâlâ az sayıdaki vakayla etkisiz olduğunu iddia ediyorlar, hâlâ ilginç bir şekilde örneklem büyüklüğü falan hesaplanmamış. Confidence interval'de dikkatli olmalıyız. Odds ratio'nun da büyüklerinden kaçınmalıyız.

Hasan Yazıcı: Çok değil, az ilgili bir soru. Benim 35 sene evvel sorduğum bir soru, hâlâ cevabı almış değilim. Yüzdelerin standart deviasyonları ve confidence interval'ları olabilir mi?

Önder Ergönül: Aslına bakarsanız, bunu software'lere verdiğinizde her şeyin confidence interval'ini

Doğru mu?
Anlamlılık varsa ($p < 0.05$), güvenilirlik aralığı hiçbir zaman etkisizliği (OR=1) kapsamaz

	Expose olan	Expose olmayan	Toplam	Proportion Exposed
Vaka	104	106	210	0.4952
Kontrolleri	12	4	16	0.7500
Total	116	110	226	0.5133
	Point estimate		%95 güven aralığı	
Odds ratio	.327044		.0748021	1.129171 (exact)
Prev. frac. ex.	.672956		-.1291707	.9251979 (exact)
Prev. frac. pop	.504717			
	chi2(1) =		3.86	Pr>chi2 = 0.0494

10. Slayt

üretiyorlar. Bana göre olamaz. Bilmiyorum, bir şeyin yüzdesinin confidence intervali deneme yaparsanız ancak olabilir. Mesela bir şeyi birden çok denediğinizde ancak olur. Ama tek bir kesitsel bir şey değil. Bir yüzdeye göre confidence interval vermek bana çok akıl kârı gibi gelmiyor. Ama bilmediğimiz başka şeyler olabilir.

Hasan Yazıcı: Çok sene evvel modaydı immuno-lojide T rozetleri yapılırdı. New England Journal of Medicine'da da yazılmış standart deviasyonlar veriyorlar. Ben yapamazsınız, dedim. Çünkü %0.98'e ya da %1.0'e nasıl confidence interval koyacaksınız? Absürd bir olay. Yüzden daha büyük şey. Tartıştık. Sonunda koyamazsın, dediler. Şimdi burada bu kadar istatistikçi varken soruyorum ben.

Standart deviasyonlar gibi dağılım parametrelerini yüzdelere koyabilir misin? Ya %98 ise, ya da %2, nasıl %100'ün üzerine çıkabiliyorsun? Yüzde-ler devamlı data değil, değil mi?

Samy Suissa: Tüm bunlar istatistik. Yüzde, güven aralığını hesaplayabilmen için yol gösterici. Eğer tahmini yöntemler kullanırsan, güven aralığının %98 ile 1.02 olan üst sınırına ulaşırsız ki bu hiç mantıklı değil. Ama burada da tartışılan bazı kesin yöntemler var, bu kesin yöntemlerle 1'i geçmeyecektir. Değerler 0-1 arasında kalır. İstatistikte tahmini yöntemler değil kesin yöntemler kullanılmalı. Ama kesinlikle oranlara göre güven aralığı verebilirsin.

Yusuf Yazıcı: Ya sadece tek bir sayı varsa, örneğin bir yorum ya da bir kesit, tekrar eden tecrübe yoksa? Güven aralığı ne ile alakalı olacak?

Samy Suissa: Oranlamada mı diyorsun? Oran da yapabilirsin. Oranda pek çok gözlem vardır, sıfırlar ve 1'ler, ve yüzdeleri söyleyip niteliği bulursun. Yani gözlem çok. Güven aralığını tek bir gözleme göre

Banu Çakır: Bununla ilgili bir şey söylemek istiyorum, proportion değil ama birçok çalışmada biz aslında prevalans hesaplıyoruz. Mesela Türkiye'de sigara içmenin ya da diyabetin prevalansı.

Söyle sıkıntılı şeyler oluyor, mesela diyorlar ki Turdeb'in çalışmasında 7,5 çıktı ya da 5,5 çıktı ama Türkiye beslenme araştırmasında 7,5 çıktı, arada 1 yıl fark var, diyabet arttı mı, azaldı mı? Özellikle böyle durumlarda bunların yanına, bunlar örneklem üzerinde olduğu için yani bir şeyi temsil etmek için yapıldığı zaman, mutlaka bir güven aralığı koymamız lazım. Bunun sebebi başka türlü karşılaştırma yapamıyoruz. Her zaman, yedi beşten büyüktür, ama bunları, güven aralıklarını mutlaka bilmemiz lazım. Özellikle saha çalışmalarında bir örnek seçiyoruz ve bu örnek aslında ülkeyi temsil ediyor. Mesela bu konuda ciddi bir sıkıntımız var. Türkiye İstatistik Kurumu, Türkiye Nüfus ve Sağlık araştırmaları verileri hep yüzde verirler bize. Türkiye'deki doğurganlık hızı % şu kadar düştü diyorlar ya da canlı doğum sayısı 25/1000 oldu diye iddia ediliyor mesela. 25/1000 gerçekten azalma mıdır, dediğimiz zaman mutlaka bizim güven aralığını bilmemiz lazım. Özellikle örneklem üzerinden seçip yapılan genellemelerde bizim için yüzdelere çok önemli. Bunu vurgulamaya çalışıyoruz ki karşılaştırma yapılabilirsin.

Odds ratio ve relatif risk konusunda biraz huzursuzum. Hazır Suissa da buradayken, bir uzman varken onu konuşmak isterim. Ben odds ratiordan relatif risk hesaplanmasına kesinlikle karşı bir insanım. Çünkü şöyle, mantık olarak relatif risk olan yerde odds ratio hesaplanmaması lazım. Yani relatif risk interventional çalışmalarda ya da kohort çalışmalarda vardır, hesaplanır. Bunların olmadığı yerlerde yani vaka kontrollerde biz odds ratio hesaplıyoruz. Dolayısıyla oradaki şey mantık olarak yanlış. Olmayan bir şeyden, probilitesini bilmediğimiz bir şeyden hesaplamamız yanlış ama eğer bir kohort çalışmayla ilgili, acaba sizin örneğiniz öyle miydi bilmiyorum. Bir kohorttan odds hesaplanırsa belki bir geriye doğru düzeltme mümkün olabilir. Ona da çok sıcak bakmıyorum çünkü niye o zaman o kohortu yapan odds ratiosu veriyor. Eğer lojistik regresyondan odds ratio veriyorsa ki o risk odds ratiodur, riskler üzerinden gelen bir şeydir. Çok gerek yok. İlla risk estimate gerekiyorsa survival analizi yapıp hazard ratio yapılabilir. Yani bu odds ratio büyüktür kabul etmeyeyim mantığına çok sıcak bakmıyorum. Mantık olarak zaten odds ratio kohortta niye kullanılsın? Bilmi-

yorum o konuda Suissa'nın da görüşü var mıdır ama o bana biraz tehlikeli geliyor. Çünkü o zaman insanlar odds ratio hesaplayıp her zaman onu risk ratio yerine kullanabilirler. Bir de biz Türkçe'de odds ratio'nun karşılığını bulamadığımız için tahmini risk ratio dedik. Şöyle sunumlar görüyorum; risk ratio'yu hesaplamadık ama tahmin ediyoruz. Bu da geçerli değil çünkü hastalık çok nadir olmadığı zaman bu da doğru bir şey değil.

Samy Suissa: Uzun cevabı ancak önümüzdeki yıl verebilirim. Kısa cevap ise odds ratioların kohort çalışmalarında kullanılma nedeni, haklısınız ihti-

yaç yok. İnsanların bilgisayarlarında sadece lojistik regresyon programları var, başka programları yok ve lojistik regresyon programları odds ratio veriyor. Kohort çalışmasında bile lojistik regresyon kullandığı için raporda odds ratio var. Size katılıyorum, kohort çalışmasında hiç ihtiyaç yok.

Ahmet Gül: Bunun tersini daha hatalı olarak görüyoruz. Retrospektif çalışma olmasına rağmen odds ratio değil de relatif risk hesaplayıp verenler oluyor. Bence diğerinden daha sık hatalı olarak yapılan bu.

Sabahtan beri dinlediklerimiz ile ilgili 2 tane makale seçtik. İlk makale, immortal time bias'ı tartışmak için seçtiğimiz bir makale. Bu çalışma, inflamatuvar barsak hastalığında infliksimabın uzun dönem güvenliliğini değerlendirmek üzere yapılmış tek merkezli bir kohort çalışması. Kısaca makaleyi özetleyeceğim önce. Güvenliliği değerlendirme açısından, randomize kontrollü çalışmaların bazı sorunları var. Örneklem büyüklüğü küçük olabiliyor, izlem süresi kısa olabiliyor, dâhil edilmeme dışlama kriterlerinin çok katı olmasından dolayı hastalıklı popülasyonu iyi yansıtmayabiliyor. Post marketing çalışmalarında güvenliliği değerlendirme açısından sorunlar var, bildirmeyle ilgili, dokümantasyon ile ilgili eksiklikler olabiliyor. Veritabanlarında ise sınırlı bir hasta popülasyonu dâhil edilmiş olabiliyor. O yüzden tam olarak bir ilacın güvenliliğini gerçek olarak yansıtmayabileceği düşünülüyor. Bu çalışmanın amacı da infliksimabın bir gerçek hayat ortamında, tek merkezde uzun dönem güvenliliğini değerlendirmek. Bu bir retrospektif kohort çalışması.

OLGULARLA TARTIŞMA

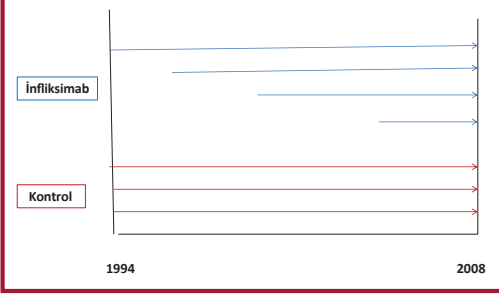
Gülen Hatemi

Bir merkezde 1994 ile 2008 yılları arasında değerlendirilmiş bütün inflamatuvar bağırsak hastaları çalışmaya alınıyor. Başlangıcın 1994 olmasının nedeni, infliksimabın o zaman kullanılmaya başlanmış olması. Hastalar bir tane bile infüzyon almış olsalar infliksimab grubuna alınıyorlar, hiç biyolojik almamış olanlar da kontrol grubunu oluşturuyor.

Örnek olarak çizdiğim bu şekilde (1. slayt) sunu göstermek istiyorum: çalışmada 1994 yılında, infliksimab ilk çıktığında başlamış olanlar da var ama ondan sonra 2008 yılına kadar ne zaman başlamış olursa olsun, infliksimab alanlar infliksimabı ilk aldıkları anda kohorta girmiş oluyorlar. Kontrol grubundaki hastaların tümü ise 1994 yılından itibaren girmiş oluyorlar. Yani bizim yan etkiler açısından değerlendirdiğimiz süre kontroller için 1994 ile 2008 arasındaki sürenin tamamı iken, infliksimab grubu için sadece infliksimab almaya başladıkları noktadan itibaren olan zaman. Yani infliksimab başlanmadan önce ve 1994'den sonraki zaman aralığındaki yan etkileri infliksimab grubu için değerlendirmiyoruz. Bu süre içinde bir yan etki olmamış olması lazım. Çünkü bu insanlar

*Romatoloji Bilim Dalı
İstanbul Üniversitesi
Cerrahpaşa Tıp Fakültesi*

Çalışma dizayını



1. Slayt

ölmüş olsaydı veya kanser geliştirmiş olsaydı infliksimab almamış olacaktı, dolayısıyla bu-
raya girmemiş olacaktı. Metotta bir de şöyle
bir cümle var; ‘gender has been anonymised for
ethical reasons.’ Hangi etik nedenle bir ilacın yan
etkisinin cinsiyetlere göre verilmemesi gerektiği
bir tartışma konusu.

Sonuçlarda infliksimab grubunda 743, kont-
rol grubunda 666 hasta olduğunu görüyoruz. P
değerleri anlamlı görünüyor ama klinik olarak
baktığınızda pek anlamlı bir şey yok. Ancak izlem
süreleri farklı, infliksimab grubunda ortalama 5
yıl kontrol grubunda ise 12 yıl. Bunun nedeni de
biraz önce bahsettiğim gibi infliksimab alanların,
sadece infliksimab başladıktan sonraki dönem-
lerinin değerlendirilmesi, kontrol grubu için ise
1994’ten 2008’e kadarki tüm sürenin deęerlendi-
rilmiř olması.

Vedat Hamuryudan: Hastaların bazıları herhangi
bir zamanda girdikten sonra infliksimab almadan
belli bir periyot takip ediliyorlar muhtemelen.
Daha sonra infliksimab alıyorlar, o zaman inflik-
simab grubuna geçiyorlar ve onun için daha kısa
oluyor.

Gülen Hatemi: Hasta yılı olarak baktığımızda da
infliksimab grubundaki hastaları 3775 hasta yılı
izlendiğini görüyoruz. Kontrol grubundakiler ise
6704 hasta yılı izlenmişler. N’lerini de hatırlatırsak
infliksimab 743, kontrol grubu 666 idi.

Çalışmada bildirilen ciddi yan etkiler şöyle: İn-
fiksımab grubunda 15 hastada malignite gelişmiş,
12 ölüm, 59 enfeksiyon, 12 infüzyon reaksiyonu ve
18 diğer yan etki görülmüş. Buna karşılık kontrol
grubundaki 666 hastanın 16’sı yani %2.4’ü ölmüş.
Bu tabloyu da ben hazırladım makalenin kendi-
sinde yoktu. Odds oranını hesapladığımızda 1.33
(%95 GA 0.56 – 3.0), p=0.45 olduğunu görüyoruz.

Malignitelere baktığımız zaman infliksimab
grubunda 743 hastanın 15’inde, yani %2’sinde
malignite geliştiğini görüyoruz. Kontrol grubun-
daysa 666 hastanın 37’sinde yani %5.6’sında
gelişmiş. Odds oranı 0.97 (%95 GA 0.56 – 1.65).

İki grubun karşılaştırması

	İnflksimab	Kontrol	OR	p
Mortalite	12/743 (%1.6)	16/666 (%2.4)	1.33 %95 GA 0.56-3.0	0.45
Malignite	15/743 (%2)	37/666 (%5.6)	0.97 %95 GA 0.56-1.65	0.91
Ciddi enfeksiyon	59 in 48/743 (%6.5)	77 in 62/666 (%9.3)	0.67 %95 GA 0.46-0.99	0.048

2. Slayt

İnfliksımab grubunda 48 hastada 59 ciddi enfeksiyon gelişmiş. 666 kontrolde ise 62 hastada 77 enfeksiyon gelişmiş. Bunun odds oranı da 0.67 (%95 GA 0.46 – 0.99) çıkıyor.

İnfliksımab grubunda, 12 ölümün 4 tanesinin infliksımab ile ilgili olduğu düşünülmüş. Burada ölüm zamanı dikkati çekiyor. Üç hasta da infliksımab başlandıktan sonraki ilk 4 ay içinde, erken dönemde ölmüşler.

İnfüzyon reaksiyonları %17 hastada görülüyor, çok ciddi şeyler değil kolay manage ediliyor. Otoimmün olaylardan, 2 hastada lupus var, 1 otoimmün hemolitik anemi, 2 demyelinating bozukluk, bir optik neuritis geliştiği bildiriliyor. Bir miktar da dermatolojik önemsiz reaksiyonlardan bahsediyorlar.

Sonuç olarak deniyor ki infliksımabın uzun dönem tedavisi güvenli ve iyi tolere edilebilir gibi görünüyor, ciddi yan etkiler infliksımab kullanmayan inflamatuvar barsak hastalıklarına benzer ve de infüzyon reaksiyonları ciddi bir sorun oluşturmuyor. Bu konuyu tartışmaya açmak istiyorum. Sizce bu iki grubun karşılaştırması böyle mi yapılmalıydı, bu çalışma tasarımı uygun mudur, ne düşünüyorsunuz, diye sormak istiyorum.

Mektupta şöyle bir düzeltme yapılmış: infliksımab grubunda ortalama izlem süresi 5 yıl, diğer grupta 12 yıl. Arada 7 yıl var. Bu 7 yılı 743 hastayla çarpıp, o hasta yılını, kontrollerin hasta yılına eklemişler. Bu şekilde yeniden odds ratio'yu hesaplamışlar. Yani bu 743 hastayı 7 yıla çarpıp 5201 hasta yılını, kontrol grubundaki 6704 hastaya ekleyip 11.905 hasta yılı üzerinden hesap yaptıkları zaman ölüm için relatif oran 2.37'den 1.33'e iniyor, kanser için de 1.73'ten 0.97'ye iniyor, diye yorum yapılmış. Sizce doğru mu?

Vedat Hamuryudan: Daha önce sen önermiştin böyle bir şeyi, gene retrospektif çalışma. Demiştin ki infliksımab alındığı anda hastalık süresi, yaşı uygun başka bir kontrolü de alıp onunla paralel götürmek çözüm olabilir mi, diye soruyordun Samy'ye, değil mi?

Gülen Hatemi: Evet yani alternatif olarak nasıl dizayn edilebilir, diye düşünürken böyle bir şey geldi aklıma. Her hasta girdiği an 94'te, 95'te, 2000'de bir tane de infliksımab kullanmamış başka bir hasta alıp böyle bir kohort oluşturmak, olabilir mi acaba?

Vedat Hamuryudan: Bu tabii prospektif dizaynı gerektirir önceden. Varsayacaksın.

Gülen Hatemi: Geriye dönük yapılabilir mi? Elimizde infliksımab kullananlar var, kullanmayanlar var.

Hasan Yazıcı: Wolfe'nin verilerine de benziyor. Onu yapamıyorsun. Yaparsın gibi görünüyor, geriye dönük olarak, o olamıyor. Çünkü bir sürü faktör var infliksımab almayanda, infliksımab almamayı sağlayan çok faktör var. İngilizlerin verilerinden bir türlü ölümcüllük ve kanser çıkmadı. İngiliz National Registry'den bir türlü çıkmadı, çıkmadığını çok iyi biliyorum. Enfeksiyon çıktı, onlar da farkındalar ama çıkamayacağını da söylemiyorlar. Böyle bir açmazdalar çünkü onu toparlayamazsın ortaya. Öbür grupta, zaten romatoid artritliler var ortada, bunların bir kısmına bir şey veriyorsun. Onu vermediğini oraya koymak diye bir şey. Sevgili arkadaşlar bu ciddi bir olay. Bütün bu olan bitende kontrollü klinik çalışmanın son 20 senede canına okunması yatıyor aslında. Kontrollü klinik çalışmalar işe yaramaz dendi, şu oldu bu oldu ve kendimizi kanıtlamak yolunda veri toplamaya çalışıyoruz. Bu observational verilerin biraz gerisinde kendini kanıtlamak çabası var. İlaç şirketlerinin de fevkalade işine geldi bu. Behçet'le uğraşıyoruz Behçet'te bir tane kontrollü klinik çalışma çıktı anti TNF'lerle. Amerika'da kontrollü klinik çalışma dediğin anda drug approval için diye düşünülen alet haline geldi bu iş. Öyle değil. Araştırmacının oluşturduğu fevkalade güzel bir alet bu aslında. Ama bu aletin yerine ne toplarsa toplansın, veriyi gelemeszin. Onların çok büyük bir kısmı hastalar kendileri gidip bildiriyorlar. Çok kötü bir epidemiyoloji o esasında.

Melike Melikoğlu: Hocam o zaman bu immortal time bias'tan kurtulmak mümkün değil. Ne kontrollü çalışma yaparsanız, ne retrospektif, ne

prospektif. Çünkü tanım gereği, bir zaman koyduğunuz zaman ileriye ya da geriye yönelik, zamana göre değiştiği için immortal bias retrospektifte ayrı, prospektifte de yine immortal biasle beraber başlıyorsunuz. Ne kontrollü çalışmalar, çünkü metoteksat kullanan ve başka bir şeye ihtiyacı olmayan hastayla benim sıfıncı dakikada infliksimab başladığım ya da ondan 5 yıl önce retrospektif serimde infliksimab başladığım hasta aynı hasta olmuyorlar. Ne background olarak ne hastalık yükü olarak aynı hasta olmuyorlar. O zaman bir immortal bias'tan matematiksel olarak bu çalışmalar nasıl sıyrılır?

Gülen Hatemi: O daha çok confounding bias'a giriyor aslında. İkinci soru 'Siz bu çalışmayı nasıl dizayn ederdiniz?' idi.

Yusuf Yazıcı: Kolay bir çözüm yok, o belli zaten ama Wolfe'nin veritabanındaki sorun, bizim de burada yapmaya çalıştığımız veritabanı gibi 'consecutive' her hastayı en başından itibaren, hasta bir şekilde semptomundan dolayı doktora gelecek. Hasta bir yerde her zaman yeni hasta. Bu sistemi her yere yaparsanız o hastalar medikal sisteme girdikleri andan itibaren dataları toplandığı zaman hem özellikleri belli hem hangi ilacı ne zaman başladıkları belli, niçin durdurdukları belli, öyle olduğu zaman retrospektif olarak analiz edilse bütün data prospektif toplandığı için bütün bayesleri en aza indirecek tek model o zaten. Ki o yüzden yapmaya çalışıyoruz. Yoksa ne zaman Corrona database olsun hastaları consent ediyorlar. Consent ettiğiniz zaman zaten ayrı bir hasta grubu çıkıyor. O yüzden observational çalışmada consent olmaması lazım bence. İnsanlar nasıl vergi veriyorsa, askere gidiyorsa consent yok onda, bu tip bazı outcomelerin da consentsiz, hastaysan, hastaneye geliyorsan bu bilgiyi vermek durumundasın. İnsanlık yararına gibisinden bir anlayış için. Ama Amerika'da böyle bir anlayış yok tabii. Avrupa'da biraz var. Ancak öyle yapılsa herkes girdiği için nerede ne olduğunu bilerek, yoksa geri dönüp şuradan başlayalım, şunlar başladılar ile bence olacak bir şey değil.

Gülen Hatemi: Ama o yine çözüyor mu bilmiyorum, çünkü bizim yaptığımız veri tabanındaki gibi

herkes bir sıfır noktasında başladı ama o arada yine belli bir noktada almaya başlayacaklar almaya söz konusu ilacı, yine değerlendirirken o grubu ondan sonra değerlendirmiş olacaksınız. Diğer kontrol grubunu başladığınız andan itibaren.

Yusuf Yazıcı: Ama bütün data olduğu için onu adjust edebilirsin. Şimdiki bilmiyoruz daha öncesini. Wolfe'nin hastası database'e girdikten önceki dönemde ne yaptığını bilmiyoruz.

Gülen Hatemi: Ama ölmemiş ve kanser gelişmemiş.

Yusuf Yazıcı: Evet ama romatoid artritleri varmış, başka ilaçlar almış, bilmiyoruz. Biziminde ölmemiş, doğru. Romatoid artrit geliştirmiş ölmediği için, tamam ona yapacak bir şey yok. Bazı şeyleri bilmemize imkân yok. Bunu kabul edelim. Bütün olay şu, eğer insanlarda bu parayı harcama şeyi yoksa, üniversitede, devlette, ilaç firması yapacak, bu ilaç firmasının suçu değil. O suç bizim suçumuz. NIH'in bütçesinin %70'i kanser araştırmasına gidiyorsa ve de 40 senede kanser mortalitesi 25'ten 18'e indiyse sadece o para ziyan olmuş para. Çünkü sigarayı bırakırsınız, kilo verirsiniz, biraz düzgün yerseniz bütün kanser tedavilerinin yaptığından overall benefit daha iyi oluyor. O parayı alıp başka şeyler yapsalar çok daha yararlı. Öyle bir takıntı olduğu için o bizim suçumuz. Vergi ödeyenler olarak biz sebep oluyoruz. Para olmadığı için yapılamıyor ilaç firması dışında randomize kontrollü trial.

Hasan Yazıcı: Bundan daha evvel yapılmaya çalışıldığı zamanlarda bundan daha mı çok para vardı? Diyelim ki NYU denek başına 5000 USD alıyor. Hayır, öyle değil kardeşim. FDA diyor ki ilaç şirketleri bize yapmak için plasebo da vermek zorundasın, ben ilacı alırım. İstiyorsan bunu ver sen bunu kanıtlamak zorundasın. Bunu kanıtlamanın yolu da üniversiteye gidip bu kadar paranın dönmemesine değil, bu kadar paranın dönmesi tamamen iyi bir kısmı gereksiz ilaç çalışmaları, bir tane Fransa'da yapıyorsun, bir tane Almanya'da, bir tane İspanya'da yapıyorsun. O bütçeyi muazzam kısmak mümkün.

Yusuf Yazıcı: Bunların hepsi özel şirket, istediğini yapar. Gidiyor Çin’de yapıyor, mesela Pfizer’in %30 hasta Çin’den geliyor çünkü Amerika’da hasta bulamıyor. Ama ona zaten engel olamazsın. Devlet de doğru neyse onu yapması lazım. En çok para devlette var, devlet tutup da NIH’in bütçesi %80’ini kansere verirse geri kalan %20’nin 18’i de kardiyolojiye verirse hiçbir şey olmaz. Bir tane daha çalışma çıktı, kanser riskini herhangi bir genetik prediktörden daha iyi diyor. Artık vaz geçelim bu işten ‘family history’ alalım, parası ucuza gelen bir şey. O parayı alıp başka yere yapmak lazım, bir de 50 sene evvel bilinmeyenle şu anda bilinmeyen aynı değil. Şu anda neyle uğraşıyoruz, relavant ne olacak, hasta metotreksat aldı. Ondan sonra biyolojik mi alsın, çok da ucuz ilacı mı alsın? Onun bir efekti var, bilmekte fayda var. Daha iyi mi, daha kötü mü? Ama onu yapmak aynı değil, daha pahalı bir ilaç. Kimseye de zorunlu sen bize ilacı da plaseboyu da vereceksin, diyemezsin. O zaman Rusya gibi olur.

Ahmet Gül: Hem Hasan Hocanın hem Yusuf’un söylediğine ek olarak yöntemle ilgili bir şey söylemek istiyorum. Sonuçta klinik araştırmalara kontürler açtırmak zorunda oldukları kesin ve bu sadece ilaç çalışmalarına sınırlı olmamalı. Hatta bu sadece ruhsat almaya yönelik çalışmalarla sınırlı olmamalı. Gerçek hayatta oldukça fazla sayıda gözlemsel veya retrospektif verilerden de bilgi elde etmemiz mümkün. Bizim bu çalışmaları yaparken yazarken ve okurken dikkatli olmamamızdan kaynaklanıyor. Eğer herhangi bir çalışmanın, bu bir veritabanı da olabilir, bir klinik çalışma da olabilir, nerelerde hata olabileceğini bu verinin ne derece güvenilir olabileceğini biliyorsak buradan çıkarımlarımız olabilir. Sorun bizim bunları mutlak gerçekler gibi okuma niyetimizden kaynaklanıyor. Dolayısıyla ben kesinlikle sadece kontrollü klinik çalışmalar değerlidir, diyemem. Tabii onların bir değeri vardır ve belki çok daha yüksek yerededir. Ama bugün biliyoruz ki o çalışmaların da çoğu suni. Gerçek hayatı çok da yansıtmayan settinglerdedir. Pratik hayata döndüğümüzde onların da sorunları çıkar. Yeter ki gözlemsel olsun, retrospektif olsun. O çalışmaların her birindeki hata nerededir, nerelerde hata olabilir? Ne derece

güvenilir olarak bunları okuyabiliriz? Bu konuda dikkatli olalım.

Samy Suissa: Bence iyi çalışmalara ihtiyacımız var. Randomize ya da gözlemsel çalışmalar, iyi çalışmalar olmalı. Bu yetersiz çalışmaya örnek. Ve yine randomize kontrollü çalışmalar için de aynı şeyi söyleyebilirim. Bu gözlemsel bir çalışma, bunlara ihtiyacımız var ama iyi çalışmalara ulaşmak için bazı standartlar var. Mesela burada açık immortal time bias var ve bununla baş etmenin birçok yolu var. Burada gösterdiğin örnek sadece bir yolu, başka yollar da var. Sanırım bunları da söyledin. Bence bu randomize kontrollü çalışmayı örnek olarak alabiliriz. İnfliksımab kullanan hastaların demografik özelliklerine bakarak ve hastalık süresini de göz önünde bulundurarak bu noktada bununla en çok uyumlu olanı buluyoruz. Yaş, cinsiyet, vs... En uyumluyu bulup, bu benim ilk kontrolüm diyerek kenara koy. Bunda infliksımab burada başlanıyor, bu belli anlara bakıp, eşleşeni bu belli anlarda bulmak zorundasın. Bu bölümü kullanamazsın, çünkü bu noktada kullanmıyorsun. Böylece burdaki immortal time’ı kaldıracaksın. Tüm bunlar infliksımaba başlamak için, burada ölemezsin, infliksımaba başlamak için yaşaman gerekiyor. Ölenlerin hepsini kaldırırıyorsun. Buraya baktığında en iyi eşleşmeyi görüyorsun. Bu ve bu da öyle. Yani bu randomize klinik çalışmayı andırıyor. Karakter özelliklerine göre eşleştiriryorsun, böylece immortal time bias’ı kaldırırıyorsun. Sonra da şüpheye düşüyorsun, “Bir dakika neden bu infliksımab başlamış da bu başlamamış, database de bazı özellikler bulabilir miyim, sihir yapmak için?” Yani eğilim skorları kullanabilirsiniz, confounding’den kurtulmak için bu tür araçlar kullanmaya başlayabilirsiniz. Ama önce immortal time bias’tan kurtulmanız lazım, ilk yapmanız gereken bu. İmmortal time bias’tan kurtulmak için gerçek bir dizayn yapmanız gerekir, sonra selection bias sorunlarından endişelenirsin, eğer kayıplar varsa confoundinge sıra gelir. İnfliksımab başlamayıp ölen hasta, kontrol grubunda ölü diye belirtilmiş, burada ise yer almıyor, çünkü yaşamaları gerekiyordu. Bu ham bias’leri yok etmen gerekir, sonra confounding bayes için endişelen. Anahtar şu; iyi bilimin standartları var. Randomize çalışmalar da bias verebilir, gözlemsel çalışmalar da. Ama çok

iyi yapılmalı gerekir, her ikisi de kullanılabilecek güncel bilgiyi verebilmek için profesyonelliğimizle uygun şekilde yapılmalıdır.

Tartışmak istediğimiz ikinci makale de randomize kontrollü çalışmaya bir örnek. Bu çalışma erken romatoid artritte başlangıçta kullanılan kısa süreli yüksek doz prednisolonun metabolik parametreler üzerine etkinliğini değerlendiren bir çalışma. Glukoz metabolizmasının belirleyicileri insülin sensitivitesi ve beta hücre fonksiyonu. Romatoid artritte de insülin duyarlılığının bozulmuş olduğunu biliyoruz. Bunun bir miktar hastalık aktivitesiyle bozulmuş olduğuna dair de çalışmalar var. Ve glukokortikoid ile ilgili de veri var. Bir takım kesitsel çalışmalar var, bunun açık insülin duyarlılığını bozduğuna ve de tip 2 diyabet gidişini predict edebildiğine dair. Ama diğer taraftan da glukokortikoidlerin glukoz toleransını düzelterek antienflamatuvar etkileri ve hastalık modifiye edici etkileri sayesinde glukoz metabolizmasını düzelterek faydalı olabileceğine dair de bir takım kısa süreli çalışmalar var. Ama bunların hepsi de açlık glukozu üzerinden yapılan ölçümlere bağlı. O yüzden bu çalışmada 2 dozda, yani 60 mg/gün ve 30 mg/gün dozunda 1 hafta boyunca aktif erken romatoid artrit hastalarında bir hafta içinde glukoz toleransı, beta hücresi fonksiyonu, insülin duyarlılığına olan etkisi incelenmiş. Burada 2 yıldan daha kısa hastalık süresi olan romatoid artrit hastaları var, DMARD ve glukokortikoid kullanmamışlar, aralarında hiç tip 2 diyabet olmadığını söylüyor. Hastaları alırken bilinmeyen kalp hastalıkları, ciddi böbrek yetmezliği olmayan hastalar bunlar. Ve de hastaları iki gruba randomize ediyorlar, bu randomize aktif kontrollü, tek kör ve çok merkezli bir çalışmanın bir bölümü bu. Büyük kısmının sonuçları yayımlanmamış, sadece bu glukoz metabolizması kısmını görebiliyoruz. Bir sorun bu çalışmanın single blind olduğunu söylüyor ama kimin kör olduğunu bilmiyoruz. Doktorun olduğunu tahmin edebiliriz, 30 mg prednisolon grubuyla ilgili 60 mg olsun diye şu kadar plasebo aldılar diye bir laf etmiyor. Larger trial diyor ama double blindin parçası olup olmadığını da bilmiyoruz. Hastalar 21 hasta 60 mg grubunda, 20 hasta 30 mg grubunda olacak şekilde randomize ediliyor ve de 75 mg oral glukoz testiyle 0. (sıfır) ve

ve 7. günlerde 0, 10, 20, 30, 60, 90 ve 120 dakika yani kısa aralıklarla plazma glukozları, insülin ve C peptid düzeyleri ölçülüyor.

Gruplar arası farklar independent samples t test, Mann-Whitney U testiyle, grup içindekiler paired sample t testi veya Wilcoxon signed rank testle değerlendiriliyor. Bir de Multivariate regression analizleriyle metabolik parametrelerle hastalık özellikleri arasındaki ilişkiler araştırılıyor ve de prednizonun AUCg üzerine yaptığı etkinin ve beta hücre fonksiyon ve insülin duyarlılığına olan etkinin belirleyicileri regresyon analiziyle araştırılıyor.

Çalışmada prednison 60 mg ve 30 mg alan iki grup arasında, baseline özellikleri açısından karşılaştırıldığında independent t testi, Mann-Whitney U testi, ki-kare testiyle hiçbir anlamlı fark olmadığı söyleniyor. Burada dikkatimizi çekecek olan şey, başlangıçta hastaların yarısından fazlasının aslında iki grupta da glukoz toleransı bozulmuş görünmesi. Ve de daha önce bilinmemesine rağmen bu çalışmayla ortaya çıkıyor ki, 60 mg prednizon grubunda 3 tane hastanın da over tip diyabeti var.

Sonuçlarda glukoz parametrelerine bakarsak eğer, 7 günlük tedaviyle ne 60 mg grubunda ne de 30 mg grubunda açlık grubunda anlamlı bir değişiklik olmadığını görüyoruz. Yine 4 saatlik area under the curve'e baktığımızda da glukozun iki grupta da anlamlı bir fark yok. Glukoz için area under the curve'ün ESR ve CRP ile korele olduğunu görüyoruz. Başlangıçta glukoz metabolizması bozuk olan 7 tane hastanın 7 gün sonra, tip 2 diyabet olarak klasifiye edilecek hale geldiğini görüyoruz. Böylelikle başlangıçta %7 olan tip 2 diyabet oranı bir hafta sonra %24'e çıkmış oluyor. Diğer taraftan başlangıçta glukoz metabolizması bozulmuş olan 9 hastanın da 7 gün sonra normal glukoz metabolizması değerlerine kavuştuğunu görüyoruz.

Glukoz metabolizmasındaki kötüleşmenin belirleyicileri araştırıldığında demografik özellikler, hastalık özellikleri ile ilgili fark olmadığı, sadece hastalık süreleri ilerleyen grupta ilerle-

meven gruba göre anlamlı olarak farklı olduğu görülüyor. Odds ratiosu 1.068, hafif bir artış var gibi görünüyor.

İnsülin ve c-peptid düzeylerine baktığımızda prednison 30 mg alan grupta açlık insülin sensitivitesinin daha fazla bozulduğunu görüyoruz. Ama daha yüksek doz, yani 60 mg alan grupta anlamlı bir fark yok. C-peptid düzeyinde 2 grupta da anlamlı bir değişiklik var.

Sonuçta deniyor ki aktif romatoid artritli hastalarda glukoz metabolizması sıklıkla bozulmuştur. Bu inflamasyonla ilişkili olabilir ve yüksek doz prednison ile kısa sürede tedavi glisemik kontrolü daha fazla bozmaz ve glukokortikoidin erken romatoid hastalarında diyabetojenik ve antiinflamatuvar etkileri birbirini dengeliyor olabilir deniyor.

Sınırlılıkları olarak da uzun dönem etkilerin bilinmemesinden, bunun kısa dönemli çalışma olduğundan bahsediyorlar. Bir de beta hücre foksionu ve insülin duyarlılığı için surrogate markerler olan oral glukoz toleransı ile ölçülebilen değerler kullanıldığını, asıl altın standartların kullanılmadığını çünkü bunun çok zor olduğunu söylüyorlar.

Sormak istediğim sorular şunlar: İlk olarak, 30 mg prednison grubunda sanki 60 mg'a göre daha fazla bir glukoz metabolizmasında bozulma var gibi görünüyor. Bununla ilgili yorum yapmak ister misiniz? Tartışmada kendileri şöyle söylemişler, bu oral glukoz tolerans testinde ve insülin düzeylerinde çok fazla variability var, kişilerin hem kendi arasında hem birbirleriyle arasında. O yüzden böyle bulunmuş olabilir, diyorlar. Ama tabii bu oldukça kötü bir açıklama çünkü o zaman yaptığınız bütün çalışmanın da aslında güvenilirliğini sorgulamanıza yol açacak. Başka nasıl yorumlarsınız?

Bir Katılımcı: Bir değişkenin variability'si yani kendi içinde bulunan değişkenliği zaten o değişkenle yapacağınız herhangi bir klinik çalışmanın gücünü belirleyen etkenlerden biridir. O yüzden o değişkeni zaten kendi içinde olan değişkenli-

ğini baştan hesaba katarsanız güç-kuvvet hesabı yaparsınız bu çalışma değişimiyle ilgili. Glukozun değişimiyle ilgili herhangi bir kuvvet hesabı yapmıyor bildiğim kadarıyla.

Bir Katılımcı: Ben çalışmanın dizaynı açısından değil de bir klinisyen olarak, biz hangi romatoid artrit 30 ya da 60 mg kortizon veriyoruz? Bu benim hangi klinik sorunu cevap verecek? Biz klinik pratikte 30-60 mg kortizon vermiyoruz ki. Bu çalışma 6 mg ile yapılsa benim için daha çok şey ifade ederdi.

Gülen Hatemi: Biz bunu metodolojik incelemek için aldık ama verenler de var aslında.

Hasan Yazıcı: Bir grup var, biz dergi kulübünde bunu okurken daha evvel ben de hemen aynı şeyi söylemişim ancak Hollanda'da bir grup ilk başta büyük doz steroid verip de ondan sonra işi önlediğine dair bir takım veriler var, onun için yapıyorlar bunu.

Yusuf Yazıcı: Şimdi tam o çalışma o zaten. 60'la COBRA'yı yaptılar bir şey gösterdi, yan etki de çok göstermedi. Sonra başka bir çalışma yaptılar çünkü 10 sene sonra Hollanda'da kimse COBRA'yı kullanmıyor ve niye kullanmıyor diye baktıklarında steroid dozu çok yüksek buluyoruz, dediler. Peki 30 olsun o. Tam işte abesle iştigal çalışması bu. Bunu çalışmanın hiçbir anlamı yok çünkü 30'u baştan kullanmaları zaten hata. Ama randomize kontrollü çalışmayı, üniversite çalışma olarak bunu yapıyor. Bu da aynı, ilaç firmasının yaptığı gereksiz çalışmanın üniversitedeki gereksiz çalışma örneği. 30 olsa ne olacak, kullanılacak mı? Hayır. Zaten kullanmıyorlar hiçbir anlamı yok, TNF'ler var, biyolojikler varken. Yan etkisi de şey ama işte bunu buraya yapmış vaziyetler.

Hasan Yazıcı: Yıllardır bu grupla kurslar veriyoruz. Ama rahatlıkla söyleyebilirim ki bu çalışma Arthritis&Rheumatism'deki geçen yılın en kötü çalışması. Bir problem şu, bu çalışma tek kör çalışmaydı değil mi? Kimdi tek kör olan?

Vedat Hamuryudan: Çünkü large double blind

kontrollü çalışmanın bir bölümü gibi düşündü, yani eğer large double blindin, tek körü varsa...

Gülen Hatemi: Ben tersini düşündüm çünkü 30 mg'lık grubun plaseboda aldığı nedir, buna dair bir açıklama yok.

Yusuf Yazıcı: İki dakika steroidlere niye takılıp kaldılar noktasına açıklama yapayım. CAMERA çalışmaları yakın zamanda yayımlandı: 2 yıl, 10 mg prednizolon + metotreksat, tek başına metotreksata karşı. Mektubumuz bunun için kabul edildi, çünkü çalışmanın sonunda eğer steroid kullanılıyorsa daha etkin sonuç alındığı söyleniyor, bu iyi ve 10 mg alanla hiç steroid almayanlar arasında hiç yan etki farkı yok. Vereceği zarar şu, diğer insanları prednizolonun işe yaradığını göstermek için 5 mg ya da daha düşük dozda prednizolon kullanmaya ikna etmek zor olacak. Biz zaten biliyoruz ki, mektubumuzun tartışma konusu da bu, yüksek dozlarda prednizolon için ikna etmeye çalışıyorlar ama bunun zararı büyük olacak. Bu randomize kontrollü çalışma için boşa harcanmış efora iyi bir örnek. İki yıl boyunca prednizolon 10 mg'ı her gün kullanıp kötü yan etki olmayacağını düşünmek gerçeğe sığmıyor.

Samy Suissa: Bir yıl süresince benimle çalışmaya Montreal'e gelen Will Dixon ile bir paper hazırladık. Prednizolon ya da oral steroidler ve enfeksiyon riskleri ile ilgili sorularımız için Quebec database'ini kullandık. Biz de enfeksiyon ve doza göre etki riskleri bulduk. Bu gözlemsel bir çalışma, kabul edilebilir, gerçek dışı değil.

Yusuf Yazıcı: Mesele şu ki, hiçbir şey görmediler.

Samy Suissa: Ya da yine güçle ilgili. Güvenilirliği görmek için ellerinde sayıları yok. Sayılar etkinlik için, güvenilirlik için değil. 'Eğer bir şeyi göremiyorsan orada değildir' diyemezsin. Sadece elinde görebilmek için yeterli rakam yok.

Gülen Hatemi: İkinci sorun, örneklem büyüklüğünün nasıl hesaplanmış olduğu. Bu çalışma glukoz metabolizmasındaki değişikliklere göre power edilmiş bir çalışma değil. Son olarak da Tablo 1'de baseline özelliklerini istatistiksel olarak karşılaştırmış olmalarına dikkat çekmek istiyoruz.

Bu daha önceki Nesnel toplantılarında da tartıştığımız bir şeydi aslında. Burada yorum yapmak istediğiniz bir şey var mıdır?

Melike Melikoğlu: Bu çalışma üzerinde değil ama genel olarak bu immortal bias çalışmaların birbiriyle kıyaslanırken pek çok bayesi, özellikle de Samy Suissa buradayken sormak istiyorum. Kontrollü çalışmalar pek çok şeye cevap vermekten acizler. Şöyle bir dizayn bizim işimize yarar mı? Open çalışma, retrospektif randomize. Yani bir grup hasta alacaksınız, şimdi biz neyi tartışıyoruz, guideline tartışıyor Yusuf, diyor ki ben katılmıyorum bu metotreksat alıyor. Bu guidelinede algoritmayı artık biz izleyemiyoruz. Yani ben kombine de başlasam iyi geliyor, metotreksata 60 mg steroid eklesem de iyi geliyor. Metotreksata infliksimab eklesem de iyi geliyor. Hem erken hastada iyi geliyor, hem geç hastada iyi geliyor. Ben de diyorum ki o zaman biz de bir hasta grubunda, ne olursa olsun, bana geliyor, romatoid artriti var, sıfır hasta. Ve ben artık kontrollü çalışmalardan cevap bulamıyorum. Kime ne başlayacağım ile ilgili. O yüzden retrospektif, open ama randomize.

Yusuf Yazıcı: Bunun tam aynısını 2-3 sene evvel informed consent ile ilgili olarak bir etikçi, galiba da gastroenterolog, söylemiş. Etik dersi veriyormuş orada hep nasıl olmamalıdır örneği verirken yavaş yavaş aklına yatmaya başlamış ki, aynen öyle diyor. Adam İngiliz. National Health System'da nasıl olsa ilacın parasını devlet veriyor. Ve de diyor ki eğer aktif ilaçlar varsa, diyelim ki biyolojikleri vermiş, metotreksatı vermiş. Nasılsa bir aktif ilaç başlanacaksa, o yüzden o aktif ilaç randomize olarak normal hayatta da başlanmalı. Çünkü hastanın hangi aktif ilaç, birbirinden farklı olmadığı için bilmesine karar vermesine, ya da doktorun bilmesine gerek yok. O yüzden de o National Health Service'in benefitini almak için bu hastanın yapacağı bir anlaşma. Tam örneğini vermiş. vermiş. Doktora gidiyorsun metotreksatı verdi, cevap vermeyince daha sonra alacağın ilaç randomize. Open label olabilir ama çok güzel compared efficacy hem ucuz tarafından hem birçok hasta tarafından aynı onu söylemişlerdi. Bence olur.

Gülen Hatemi: Bunu enfeksiyoncular uzun süre yaptılar aslında. Özellikle, hematoloji ve onkolojide yatan hastalara, mesela belli bir bakteri ürememişse, antibiyotik başlayacakları zaman iki türlü şemaları vardı. Birine onu birine onu başlıyorlar ve outcomelerini değerlendiriyorlardı. Bu 5-10 yıl önceydi. Artık etik kurullar izin vermiyor. Mesela biz Behçet'te azatioprin ve siklosporine dirençli

üveit hastalarını interferon veya anti-TNF'ye randomize etmeyi istedik ama SGK'nın parasıyla artık onu yapmak mümkün değil.

Yusuf Yazıcı: Gerçekten bilinmiyorsa bu mu, bu mu daha iyi diye vermemekte sorun yok, kimseye harm ya da undertreatmen olmuyor. Herkese aktif olduğu ilacı veriyorsunuz ve bilmediğimiz için etik bir sorun yok bence.

ROMATOLOJİDE KLİNİK ARAŞTIRMALAR VE NESNEL TİP VIII

EDİTÖR
HASAN YAZICI

S. AYDIN
B. FELDMAN
G. HATEMİ
M. İNANÇ
E. LESAFFRE
İ. ŞİMŞEK
B. ÜSTÜN
H. YAZICI
Y. YAZICI
S. YURDAKUL

Borderline Mean Pulmonary Artery Pressure in Patients With Systemic Sclerosis

Transpulmonary Gradient Predicts Risk of Developing Pulmonary Hypertension

Christopher J. Valerio, Benjamin E. Schreiber, Clive E. Handler, Christopher P. Denton, and John G. Coghlan

ARTHRITIS & RHEUMATISM
Vol. 65, No. 4, April 2013, pp 1074–1084

SİSTEMİK SKLEROZLU HASTALARDA “SINIRDA” ORTALAMA PULMONER ARTER BASINCI

Murat İnanç

*Romatoloji Bilim Dalı
İstanbul Üniversitesi Tıp Fakültesi*

Bu toplantıda tartışacağımız makale Royal Free Hospital'dan kıdemli yazarları Christopher P. Denton ve John G. Coghlan olan “Borderline Mean Pulmonary Artery Pressure in Patients with Systemic Sclerosis” (CJ Valerio CJ et al. Arthritis Rheum 2013;65:1074-8).

Pulmoner arteriyel hipertansiyon (PAH) sistemik sklerozun (SSk) ölümcül bir komplikasyondur ve sıklığı yaklaşık %10'dur. Tanıda zorluk yaratan problem, kesin tanının invaziv bir yöntem olan sağ kalp kateterizasyonu ile konulmasıdır. Sağ kalp kateterizasyonu yapılacak hastaların seçimi konusu önemlidir. Bu konuda Doppler ekokardiyografi ile tarama gerekli bir yöntemdir ve Türkiye'de de romatologların birçoğu bunu çok iyi öğrenmiştir. Bununla birlikte ekokardiyografik taramanın PAH'ın saptanmasında yetersiz kaldığı hastalar vardır. Son 7-8 yıldır üzerinde tartışılan ‘sınırdaki pulmoner hipertansiyon’ kavramı bu konuda riskli hastaların erken tanınmasında rolü olabileceği düşüncesiyle ortaya çıkmıştır. Normal toplumda pulmoner arter basıncı (PAB) ortalama 14 mmHg'dir. Standart sapması yaklaşık 3 olan bu değeri dikkate alırsak normallerin PAB değerlerinin 20 mmHg'den az olması bekleniyor. PAH sınırı ise 25 mmHg'den başlıyor. Yani sol ventrikül normal, pulmoner arter tıkanma basıncı (wedge pressure) normal, sol ventrikül hastalığı olmayan biri için sağ kalp kateterizasyonu ile PAB 25 mmHg ve üzerinde olduğunda başka bir hastalığı yoksa PAH tanısı koyuyoruz. Demek ki PAB 21-24 mmHg

olan gri bir alan var. Bu araştırma grubu da, bu gri alanda bulunan hastalarda uzun dönemde ne oluyor sorusunu araştırmak istemiş.

Çalışmada ele alınan temel sorular; a) Sınırdaki PAH (ortalama PAB 21-24 mmHg) olan hastalar normal PAB değerleri olanlara göre PAH gelişme riski artmış mıdır? b) Sınırdaki PAH olanlarda sağkalım normal PAB olanlardan farklı mıdır?

Ben bu başlığa baktığımda iki şey görüyorum, birincisi ben de bu alanda çalıştığım için yazarları tanıyorum, önemli bir merkez. Bu durum okuyucu olarak bizleri etkiler. İkincisi de yakın zamanda bana nefes darlığı ile gelen ama ortalama PAB değeri sağ kalp kataterizasyonu ile 24 mmHg çıkan ve ne yapacağını bilemediğim bir hasta aklıma geldi. Şu anda da bir tedavi almıyor. O şekilde de devam ediyor, bir yıl sonra bir daha sağ kalp kataterizasyonu yaptık, yine 24 mmHg, dolayısıyla böyle bir gri alan var. Bu araştırmacılar da nasıl bir soru sormuşlar bu durumda, bu 'sınırdaki' PAH olma riskleri yüksek mi? Birinci soru. İkinci soru; normallerle karşılaştırıldıklarında sağkalımları daha kötü mü? Şimdi bu nasıl bir seri, nasıl bir kohort? Sonuçta bu bir kohort çalışması gibi duruyor ama dikkat edelim, hasta alımı üzerine konuşacağız. Burada çeşitli merkezlerden Royal Free'ye hastalar gönderiliyor. Ne zaman gönderiliyor, SSK hastaları gönderiliyor, ama bu SSK hastalarının sağ kalp kataterizasyonu endikasyonu varsa gönderilme olasılığı daha yüksek. Ve aslında biz bu oranı bilmiyoruz. Yani İngiltere'deki SSK'ların hepsini toplamak istiyorlar ama bunda %100 başarılı olmak mümkün değil. Başka merkezlere de gidiyor hastalar, sağ kalp kataterizasyonu olması gereken hastaları topluyorlar çünkü kataterizasyon santral olarak yapılıyor. Bu çalışmanın iyi tarafı o, ama hasta seçimi birçok merkez tarafından belirli kriterlere göre periferde yapılmış. Bu konuda önceden belirlenmiş kriterleri olması çalışmanın olumlu taraflarındandır. Bu kriterler daha önce yayımladıkları kriterler. Ekokardiyografik kriterleri var, akciğer solunum fonksiyon testiyle ilgili kriterleri var, bir de bunlar olmasa dahi, hasta benim nefes darlığım var diyorsa, klinisyen ikna olduysa kataterizasyon için gönderiliyor. Sonuç olarak gelen hasta havuzunda belirgin bir heterojenlik var.

Bir Katılımcı: Bu çalışmayı okurken bana hep retrospektif bir çalışma gibi, hani kendi olgularını sonradan değerlendirmişler gibi hissettiriyor, çünkü hasta hakkında çok fazla kriter var.

Murat İnanç: Gözlemsel nitelikli kohort çalışmaları ve uzun süreli prognoz çalışmalarının çoğunun karakteri budur. Çünkü bu sistem kurulduğunda, SSK hastaları bu merkeze gönderilmeye başlandığında, sağ kalp kataterizasyonu yapıldığında bu kavram ('sınırdaki PAH') henüz yoktu. Bu merkez 20 senedir bu konuyla uğraşıyor, bu kavram 2008'de önem kazandı. Ama sonuçta ellerinde bu hastaların verileri var. Bu çalışma, (a) özellikle bu konuyu araştırmak için düzenlenmiş bir çalışma değil, (b) bu konuda gerekli veriler ellerinde var, birçok prognoz çalışmasında da bu bilgilerin bir şekilde değerlendirilmesi lazım. Bu çalışmalarda kısmen bir taraflılık olması kaçınılmaz. Yani burada hem sistemik hata olabilir, hasta alımı açısından, çünkü diyelim ki bir merkezde daha şüpheli bir hekim var, "Nefesim biraz daralıyor" diyen bütün hastaları ileri tetkik için gönderdi. Pahalı bir şey, hastayı Londra'ya gönderiyorsunuz, öbür tarafta bir tek testlere dayalı olarak, test patolojik çıkınca hastayı gönderen bir hekim var. Bunların oranlarını maalesef bilmiyoruz. Dolayısıyla tam homojen hasta alımından bahsetmiyoruz. Burada bir seçim taraflılığı problemi olabilir. Ayrıca bütün prognoz çalışmalarında olduğu gibi birçok bilmediğimiz 'confounding' yani karıştırıcı faktör de olabilir. Ama yine de bu merkezin hastaları böyle bir hasta grubunu toparlamak için en uygun serilerden birisi. Sayı yetersizliği sorunu olmaması önemli, çünkü bu hastaları bulmak da zor, birazdan 1-2 örnek daha göstereceğim. Bu hastaların gerektiği şekilde değerlendirildiği tecrübeli bir merkez bulmak da zor.

Kısaca akış şemasına bir bakalım. Sekiz yüz hastaya sağ kalp kataterizasyonu yapılmış. Veri tabanını çok iyi tutan bir yer ama gönderildikten sonraki durumu biliyorlar, önceki durumu bilmiyorlar. Bunların 505'inde pulmoner hipertansiyon zaten saptanmış, tanı konmuş hastalar, bunu bir tarafa ayırıyoruz. Yani 25 mmHg'nin üzerinde çıkmış kateter, şu ya da bu patoloji. Üç yüz'den fazla hastada ise PH saptanmamış, esas hedef

hasta grubumuz bu. Burada nasıl bir kohort oluşacak? Öncelikle önemli akciğer hastalığı olanlar, interstisyel fibrozu olanlar buradan çıkarılacak. Çünkü orada başka bir patoloji söz konusu. Demek ki bu 90 hasta da buradan çıkarılmış. Geriye kalıyor 228. Bizim kohortumuzun iki ayağını oluşturacak olan hasta havuzu bu 228 hasta. Bu havuzun bir grubu basıncı 21-24 mmHg olan yani 'sınırdaki' PAH olanlar, diğer grubu ise normal çıkanlar. Çünkü sonuçta her iki hasta grubunda da aynı işlem yapılıyor. Solunum testleri bozulursa, eko bozulursa veya semptomatik bir bozulma olursa her iki gruba da yeniden sağ kalp kataterizasyonu yapılıyor. Ama başlangıçta bu ayrımı nasıl yapıyorlar, hangi parametreye göre? İlk önce sağ kalp kataterizasyonundaki pulmoner basınca göre ayırıyorlar. Bu şekilde gittiğimiz zaman, önce normallere bakalım. Yani pulmoner basıncı katederde normal çıkmış, ikinci kez kriterlere uyduğu için seçilip kataterizasyon yapılmış. O zaman ne olmuş, bunlardan 11'inde pulmoner hipertansiyon bulunmuş, bir bölümünde yine sol kalp hastalığı, akciğer, bunlar çıkarılmış. Bütün bunların altındaki düşünceyi de şöyle kısaca açıklayalım. PAH konusunda ilaç tedavi seçenekleri olmasına rağmen bu konuda yayımlanan yeni bir derlemeyi de dikkate alırsak hâlâ 5 yıllık sağkalım %50'lerde, yani bütün bu ilaçlara rağmen sağkalım düzelmiş değil. Ancak tarama ile bulunan bir grup hastada sağkalım düzelmiş. Bu araştırma grubu da yeni bir erken tanı çabası gösteriyor, yani sınırdaki basıncı olanlara tanı koyarak veya onların PAH gelişme riskini belirleyerek erken tedavi etmek istiyor. Aslında temel amaç bu. Sonuçta bu grupta ilk sağ kalp kataterizasyonu normal olmasına rağmen ikinci kez yapıldığında 7 hastada PAH geliştiği belirlenmiş.

Bir Katılımcı: Tesadüfen mi iki grupta da normal olanlar ve sınırdaki basıncı olanlar 38 kişi?

Murat İnanç: Evet tesadüf. Zaten oranları farklı. Hasta seçimi açısından şunu düşünelim. Bu bir SSK kohortu değil. Bu, SSK hastaları içinden, PAH riski yüksek olduğu bilinerek, ama ispat edilmemiş olarak seçilmiş bir grup. Amaç, izleme ile bunu kanıtlamak ve erken tedavi vermek. Burada pahalı tedavilerle ilgili endikasyonun yerine ge-

tirilmesi de söz konusu. Mesela Amerika Birleşik Devletleri'nde başka bir grubun çalışması var. Onlar kataterizasyonda sınırdaki basınç değerlerini görünce, "Klinisyen olarak benim görüşüm bu hasta riskli, ben buna da ilaç vereyim," diyebiliyorlar. Ama burada basıncı 25 bulmadıysanız ödeme sistemi açısından o ilacı veremiyorsunuz.

Hasan Yazıcı: Murat birkaç şey var, bir tanesi, ekokardiyografiyle sağ kalp kataterizasyonu arasındaki ilişkinin kötü olduğunu söylüyor...

Murat İnanç: Doğru.

Hasan Yazıcı: Ancak az önce söylediğin gibi, "Bir daha sağ kalp kataterizasyonunu tekrarlamamın kriterlerinden bir tanesi ekokardiyografinin bozulması..."

Murat İnanç: Evet.

Hasan Yazıcı: Peki çok güzel de öyle olduğu zaman birinci cümle, ikinci cümleyle çıkıyor gibi.

Murat İnanç: Doğru.

Hasan Yazıcı: Onu bırakıp da acaba doğrudan doğruya bunların ekokardiyografilerinin ne olduğunu söyleselerdi. Bu birinci sorum.

İkinci sorum, triküspitteki jet hızı meselesi var. O makalenin sonunda, tartışmaya gelirken diyor ki; bütün bunlarla düzeltildikten sonra triküspitteki jet hızı, triküspit gradyanı tekrar sabit kaldı, ama sonra tartışma bölümünü okumaya başlıyorsunuz ki triküspit gradyanı en önemli şey. Peki, o ikisi nasıl bağdaşıyor?

Murat İnanç: Hocam, iki sorunuza ayrı ayrı cevap vereceğim. İlki, bu ekokardiyografi ile kataterizasyon arasındaki uyumsuzluğu bu grup ve dünyada başka gruplar göstermiş ama şöyle bir uyumsuzluk var, eğer ekokardiyografide 60'ın üzerinde sistolik basınç ölçüyorsanız uyum çok yükseğe çıkıyor, yani uyumsuzluk %100 değil, 40-50 gibi... Değerler arttıkça iş kolaylaşıyor, dolayısıyla giriş kriterlerinden birisi ekokardiyografi, birisi solunum fonksiyon testi, birisi de semptom. Bütün dünyada ekokar-

diyografide karar verilemeyen ama semptomatik olduğu için kateter yapıp PAH tanısı konan hastalar da var. Çok önemli bir mesele de burada kardiyologların tutumu. Bu çalışmada da bir kardiyolog var. Romatologlarla çok uyumlu çalışan bir kardiyolog, Gerry Coghlan. Son Nice PAH kongresinde kardiyologlar bu çalışma konusunda kendi aralarında büyük bir tartışmaya girdiler. Kimisi çeşitli ekokardiyografi yöntemleriyle bu hastalara PAH tanısı konulabileceğini söyledi. Yani burada dikkat etmemiz gereken bir şey de şu; ekokardiyografi metotları standart değil. Birçok çevre merkezde yapılmış ekokardiyografi değerleri var. Burada gerçek hayat söz konusu. Kontrollü bir çalışma değil. Tek merkezde bütün hastalara ekokardiyografi yapılmış, ondan sonra kataterizasyon yapılmış bir durum yok. Çok merkezde herkes kendi kardiyoloğuna ekokardiyografi yaptırmış, aslında Türkiye’de de hepimizin karşılaştığı durum.

Şimdi hızla ilerleyelim. Bu borderline grupta da 38 hastaya katater yapılmış. Rakamlar önemli değil, oranlar farklı zaten. Burada da 16 hastada PAH saptanmış. Yani özet olarak, her iki grupta da bir grup hasta PAH’a ilerlemiş. Ama basınç sınırda bulunmuşsa oran biraz daha yüksek. Bu aslında çok da beklenmeyen bir şey değil. Bu hastalık ya da bu patoloji muhtemelen bir süreç yani ‘sınırda’ basınç olduğu zaman da ilerliyor, normalden de bir grup hasta bu konuda riskli, siz bir kesitsel alanda ölçüm yapıyorsunuz. Belki ölçümü 3 ay sonra yapsanız olay tamamen değişecek. Yani tek bir ölçüme göre ayırıyorsunuz.

Ahmet Gül: 25’in sınır olarak belirlendiği çalışmalarda aslında daha geniş spektrumlu bir tarama var. Az evvel Hasan Hoca’nın sorduğu sorunun veya kohortu nasıl tanımladığının cevapları biraz orada var. Oralara dönüp bakıldığında onların prospektif takibini yapmak belki böyle bir seriyi takip etmekten daha akılcı olabilirdi gibi geliyor bana. Yani o toplum verisinde sınırın konulduğu yerin aslında bir 10 yıl sonraki ya da 20 yıl sonraki seyrini, sınırı doğru yere mi koyduk, nerede kümelendi sorusunun cevabı, belki tanımlandığı yerde.

Murat İnanç: Buradaki güçlük şu, biz özel bir hasta grubuyla ilgileniyoruz. Burada hasta SSK

olacak, belli özellikler taşıyacak, yani çok seçilmiş bir hasta grubu. Normal toplumdan çok farklı bir seyir de olabilir. Prognoz çalışmaları gösteriyor ki 25’in üzerinde basınç olanlarda mortalite yüksek, böyle bir veri var.

Şimdi istatistik analiz olarak benim anlayabildiğim ‘State of art’, yani ellerinden gelen her şeyi yapmışlar. Survivor eğrileri, regresyon analizi vb... Ama şunu da düşünmek lazım, bu çalışma prognoz belirleyicilerini ortaya koymak için yapılmış bir çalışma değil. Bu konuda bir hazırlık yok. Örneğin her hastada serumda NT-proBNP ölçülmemiş, ürik asit ölçümü yok... Bu ölçümlerin önemini gösteren bir çalışma da yayımlandı (DETECT çalışması), gerekirse konuşuruz. Dolayısıyla buradan ne çıkar diye bir araştırma bu. Daha çok kardiyoloji yöntemlerinin kullanıldığı bir araştırma, üstelik kullandıkları bütün parametreler de birbiriyle ilişkili. Hiçbiri bağımsız değil. Triküspitteki gradyan pulmoner basınçla bağlantılı, o sağ atrium basıncıyla bağlantılı, ama bu ölçüm daha iyidir, diye bir şey çıkarmışlar sonuçta. Bizim için belki önemli olan 23 hastada PAH geliştiği tespit edilmiş. ‘Sınırda PAH’ olan grupta aşağı yukarı %19, normal grupta %5. Bu durum beklenen bir durum. Sağ kalp kataterizasyonu ile PAB değerleri daha yüksek olanlarda prognoz daha kötü olmuş. Çok önemli olan 3 yıllık sağkalıma bakarsanız, %82. Böyle bir merkezde üstelik böyle hastaları bulana kadar çok yakın izlemiyorlar. Ama bulduktan sonra oldukça iyi izliyorlar. En iyi protokolleri, en yeni ilaçları kullanıyorlar. Buna rağmen örneğin lupustaki sağkalımların son derece altında. Dokuz tanesi ölmüş, 1 tanesi komorbidite kaynaklı. Bir de gene şöyle bir şey var, bir kriter belirlemişler, yürüme hızına, solunum fonksiyon testine göre. Daha iyi giden hastalar da var. Yani aslında kataterizasyon, evet basıncı yüksek olanlarda riskli hastaları ayırıyor, ama daha iyi prognozu olan hastaları çok ayırmıyor. Bunlara da PAH diyoruz. Ama aynı tedavileri gördükleri halde bu hastalar daha iyi de seyredebiliyor. Yani elimizde bu konuda mükemmel bir belirteç yok.

Bu çalışmada prospektif olarak toplanmış bir hasta verisi var elimizde. Evet, hasta seçimi olarak

Çalışmanın güçlü ve zayıf yönleri

Güçlü Yönler

- Tecrübeli merkez ve araştırmacılar
- Yüksek riskli hastalardan oluşan kohort
- Sağ kalp kateterizasyonunun tek merkezde yapılması
- Takipte hasta kaybı olmaması
- Güçlü istatistik analiz
- Yeni bir yüksek riskli grubun tanımlanması
- Sağ kalp kateterizasyonu için hasta seçiminde aşama
- Erken etkin tedavi için hasta seçiminde aşama

Zayıf Yönler

- Bir 'inception' kohort değildir
- Çok sayıda merkezden hasta gönderilmişdir
- Seçilmiş yüksek riskli hastalar incelenmiştir
- Bir prognoz belirleyicisi bulmak üzere tasarlanmamıştır
- Retrospektif bir analizdir
- Takip versinde eksikler vardır
- Ölüm nedenleri sistematik olarak incelenmemiştir
- Tedavi etkisi bilinmemektedir
- Belirlenen eko verisinin (pulmoner akım gradyanı) bağımsız bir etken olduğu ispatlanamamıştır
- Çalışma sonuçlarının dikkate alınması sağ kalp kateterizasyonu ve tedavi uygulamalarını artırabilir

1. Slayt

sorunları var. Sonuçta 'sınırdan' ortalama PAB değerleri olan hastalar seçilmiş, belirli özellikleri olan SSk hastalarında, buradaki karşılaştırmayı söyledim. Bir de regresyon analiziyle pulmoner akımın diğer ekokardiyografik parametrelere göre, riski daha iyi gösterdiğini ortaya koymuşlar. Bu teknik bir şey. Bunun çoğu bizim elimize gelen ekokardiyografilerde bulunmamaktadır. Bunu kardiyologlarla tartışmak da mümkün olmayabilir. Sonuçta diyorlar ki, bu çalışma bizim SSk PAH hastalarında erken tanı koymamızı sağlayabilir. Tarama yeteceğimizi biraz daha arttırdı. Bu hastalarda daha sonra gelişen PAH tamamen iyi prognozlu değil, gördüğünüz gibi ölen hastalar da var. Bu hasta grubunun yakın takibini öneriyorlar, yerinde bir öneri. Çalışmalarında kendilerinin yazdığı negatif noktalar; retrospektif bir analiz tabii ki. Hasta seçimi sorunu var, yani çalışmaya girerken çok seçilmiş hastalar var, sağ kalp kateterizasyonu tek zamanda yapılıyor, aslında bir süreçten bahsediyoruz, dolaşısıyla bunun gerçekte alt sınır olduğunu söylüyorlar. Yani daha sık kateterizasyon yapabilirsek tanı olasılığının artacağı anlamı çıkıyor. Ve çok önemli bir şey daha var yazının içinde. Bunun için yazıya dönmenizi öneririm. Burada hastalar öldü, diyoruz. Sonuçta çoğunun neden öldükleri de belli değil, yani sistemik bir inceleme yok. PAH'ın burada en önemli etken olduğu kabul edilebilir ama birçok hasta komorbidite nedeniyle ölüyor. Buradaki PAH tanısıyla prognoz arasında ilişki var ama mutlaka neden sonuç ilişkisi değil.

Bir Katılımcı: Murat Hoca söylemeden geçemeyeceğim, aslında Hasan Hoca sorar gibi oldu. Biz kendi aramızda okuduğumuzda, tartışma bölümünün ilk paragrafında triküspit gradyanının prediktif olmadığını söylüyor. Bütün makale bunun üzerinden giderken onu hâlâ anlayamıyorum.

Murat İnanç: Tabii burada şu çelişki var, ekokardiyografik parametreyle hasta seçemiyoruz ama bizim elimizdeki değişkenlerden prognozla en çok ilişki gösteren bu parametre. Teknik olarak çalışmanın bu parametreler arasındaki farkları araştırmak için tasarlanmadığını düşünebiliriz.

Slayt 1'de bir çalışmanın güçlü ve zayıf yanlarını ortaya koymaya çalıştım. Sonuçta bu yüksek

riskli hastaların oluşturduğu bir kohort. Seçilmiş, santral olarak sağ kalp kataterizasyonu standart olarak yapılmış hastalardan oluşuyor, bu önemli. Takiplerin kalitesinin yüksek olduğu anlaşıyor ama sağ kalp katateri grubunda takiplere gelmeyenlerden hiç bahsedilmemiş. Demek ki yok gibi kabul etmek zorunda kalıyorsunuz, hâlbuki bu konudaki literatürü okuduğunuz zaman izleme bilgisi çok önemli. Çok iyi bir istatistik analiz yapılmış. Nedir literatüre katkısı? Muhtemelen yeni bir yüksek riskli grup. Tahmin edilemez yüksek riskli bir grup değil de, bunu ispat eden veya ispat etmeye çalışan bir çalışma. Yine sağ kalp kataterizasyonu için hasta seçiminde bir iyileşme göze çarpıyor. Hastalar belki bu dönemde tedavi olsa daha iyi olacak gibi bir izlenim ortaya çıkıyor. Ama tabii bu bir 'inception' kohort değil, hasta girişi tamamen heterojen ve farklı merkezlerden. Bunu dikkate almak lazım. Farklı merkezlerden seçimde bir standardizasyon zor, ekokardiyografiler farklı yerlerde yapılmış, bir prognoz belirleyicisi bulmak için planlanmamış, dolayısıyla hiçbir biyobelirteç incelenmemiş. Onlarla bir karşılaştırma yok. Ekokardiyografi şu parametresi çok önemlidir diye karşımıza geliyor aslında bu konuda yapılmış istatistiğe baktığınız zaman, bütün parametreler istatistik olarak anlamlı, çoklu analiz yapılırken belirli bir parametre çıkmış. Retrospektif bir analiz. İzleme verileri tam değil. Elimizde tam bir tedavi verisi yok, son klinik takipler bazı hastalarda yok, bunun için Ulusal veritabanlarına bakmışlar ama bunların ne kadar sorunlu olabildiğini biliyoruz. Ölüm nedenleri bilinmiyor. Tedavinin etkisi bilinmiyor. Burada, buldukları parametrelerin diğerlerinden farklı olduğuna dair elimizde değerli bir veri yok. Riskler nedir? Sağ kalp kataterizasyonuna çok daha fazla sayıda hasta gönderilmesi, tabii bunun sağlık sistemine bir yükü var. Bu çalışmanın sonuçları elimizdeki kısmen etkili ilaçların çok daha fazla veya gereksiz kullanılması sonucunu da doğurabilir.

Yasemin Kabasakal: Çok teşekkür ediyorum. Ben sadece şunu sormak istiyorum. Hastalar farklı merkezlerden geldi, hasta alımı çok değişken diyorsunuz. Peki, hastalar bu merkeze gönderildiğinde tekrar eko yapılsa, kendi standartlarına giren hastalar arasında, ben bu hasta grubunu

alacağım ve bunu inceleyeceğim, diyerek kendileri eko yapsalar, o standartlara girenlerle devam etseler, daha uygun bir çalışma olabilir mi bu? Ona göre dizayn edilmemiş mi?

Murat İnanç: Benim anladığım kadarıyla bunun iki cevabı var. Birincisi gönderen merkez açısından olay bir probleme dönüşebilir. Gönderen merkeze "Hayır, ben uygun bulmadım" diyemiyorlar. Çünkü bazı semptomlara dayanarak göndermiş. Her gönderen de katateri yapıyor, bir kere o aşamayı hasta geçmiş oluyor. Onun için ayrıca bir risk değerlendirmesi yapmıyor, tabii kendi izledikleri hastalarda zaten heterojenlik yok. Royal Free'de başlamış, değerlendirilmiş hasta daha tecrübeli ellerde, muhtemelen daha değişik yöntemlerle değerlendiriliyor. Ama dışarıdan gelenler muhtemelen daha serbest kriterler uygulanarak geliyor. Bunu standardize etmek anladığım kadarıyla problemli. İngiltere'nin öbür ucundan bir hasta gönderilince "Hayır, biz uygun bulmadık" deyip geri gönderemiyorsunuz.

Emmanuel Lesaffre: Eğer sorularım fazla teknik gelirse affedin. Ben bir istatistikçiyim, mazur görün. İstatistik tekniğin son noktası. Pulmoner hipertansiyonda sağkalım oranlarına bakarsanız, iki sorum olacak. İlki, değişmez kural mıdır, bir kez saptanmışsa tersine çevirmek mümkün değil midir? İkinci sorum; pulmoner hipertansiyon taraması sabit midir? Taramanın ne sıklıkta yapıldığını göremiyorum. Buradaki sorun, zamanını net olarak bildiğin ölümden farklı olarak, sadece tarama çalışmalarınızı yaptığınızda sorunun pulmoner hipertansiyon olduğunu anlıyorsunuz, iki tarama arası rastlanır, bu yüzden 'sansürlü değişken' denir ve genellikle makalelerde göz ardı edilir. Bu, onkolojide progresyonsuz sağkalıma baktığınızda çok sık yapılır.

Murat İnanç: İstatistiğe yakın birisinden bu eleştiriyi bekliyordum. Onkoloji ile kıyaslama yapmanıza memnun oldum çünkü romatoloji alanındaki hastalarla çok benzerlik gösterirler. İlk soru tarama. Tarama yılda bir kez yapılıyor. Dünyaca kabul gören bu. Çünkü az sayıda riskli hasta için yüksek mortalite riski söz konusu. Kimse daha sık yapalım diyemiyor. Diğer konu, PAH gerileri mi?

Kardiyolojide uzman değilim ama uzmanlar eğer PAH ise sol kalp hastalığı ya da interstisyel akciğer hastalığına bağlı olmadan her zaman ilerleyici olduğunu düşünüyorlar. Varsayılan bu, klinik tecrübeler de bunu doğruluyor. Üçüncüsü tarama aralıkları ile ilgiliydi, kimse bilmiyor. Ölçümler yılda bir kere yapılır. Bu haliyle teknolojiye son nokta. En iyisi değil ama eğer prospektif çalışma yapıyorsanız, birkaç hafta önce DETECT isimli bir çalışma yayımlandı, sağ kalp kateterizasyonu veya ekokardiyografiyi yılda birkaç kez yapılabilirsiniz. Maalesef pratikte durum bu değil ve istatistik olarak bu çalışmada uygulananlar belki de ideal bir metot değil.

Ahmet Gül: Her konuşmacı, konuşmasını acaba ideal dünyada ben bu çalışmayı nasıl yapmayı dü-

şünürdüm diye tamamlarsa belki çok daha yapıcı da olabilir. Mesela bu çalışmayı Murat sen planlasaydın ve her türlü imkânın da olsaydı, nesini düzeltsen sonuçlar daha güzel olurdu?

Murat İnanç: Bu konuda riskli hastalara prospektif bir çalışma yapılabilir, ki yapıldı. Baktığımız zaman önümüzdeki hafta Ankara’da bir toplantıda da tartışacağız. DETECT çalışması diye bir çalışma. Risk faktörü olanlar, prospektif olarak sağ kalp kateterizasyonuna girdiler ve izlendiler. Oradan risk faktörleri, tabii onun içinde “biomarkerlar” da var, ürik asit, NT proBNP ve oradan hangi hastanın daha riskli olduğuna dair bir skala çıkarıldı. Ama o skala da henüz başka hasta gruplarında değerlendirilmedi ve doğrulanmadı.

ACR/EULAR ROMATOİD ARTRİT SINIFLAMA KRİTERLERİNİN ÇOK ERKEN İNFLAMATUAR ARTRİTİ OLAN HASTALARA UYGULANMASI ÇALIŞMASINA ELEŞTİREL BİR BAKIŞ

Gülen Hatemi

Romatoloji Bilim Dalı
İstanbul Üniversitesi
Cerrahpaşa Tıp Fakültesi

A critical appraisal of

Application of the 2010 ACR/EULAR classification criteria in patients with very early inflammatory arthritis: analysis of sensitivity, specificity and predictive values in the SAVE study cohort

Iuliia Biliavska,^{1,2} Tanja A Stamm,² Jose Martinez-Avila,² Thomas W J Huizinga,³ Robert B M Landewé,⁴ Günter Steiner,² Daniel Aletaha,² Josef S Smolen,² Klaus P Machold²

ARD 2013

Bu çalışmanın amacı ACR/EULAR 2010 kriterlerinin, başlangıçta bu kriterler geliştirilirken kullanılmamış olan randomize kontrollü bir çalışmanın kohortunda validasyonunu yapmak. ACR/EULAR 2010 kriterleri anımsayacağınız gibi 3 aşamada geliştirilmişti. İlk başta erken artritle gerçek hastalar vardı. İzlemede bir yılın sonunda metotreksat kullanmayı gerektirmiş olmayı RA kabul etmişlerdi. İkinci aşamada paper hastalar üzerinde kriterler sınanmıştı. Üçüncü aşamada ise başka kohortlarda validasyonu yapılmıştı. Burada kritik olan RA kabul edilen durumun, hastada bir yıl içinde metotreksat başlanmış olması olarak tanımlanması. Bir de skorum yapıyor, 4 domaini var, eklem tutulumu, seroloji, akut fazın yüksek olması ve semptom süresi olarak. Bu skorlamaya göre 6 puan alan hastaya RA deniyor.

Servet Akar: Gülen çok özür dilerim, bu çalışmaya ihtiyacımız var mı?

Gülen Hatemi: Böyle çalışmalara ihtiyaç olabilir tabii, eğer düzgün yapılırsa, iyi kohortlarda yapılırsa bence var.

Servet Akar: Aslında söylemek istediğim bir şey vardı. Bu ACR/EULAR kriterlerini, yazarların girişte vurguladıkları, kendilerinin söylediği en az 3 tane ya da daha fazla valide eden çalışma var.

Gülen Hatemi: Çalışmalar var ama genellenebilirliğini söyleyebilmek için, daha farklı yerlerde, daha farklı hasta kohortlarında sınanmasında fayda var.

Servet Akar: Onlar aslında kendi farklılıklarını vurgulamışlar, tek bir merkezde değil, birden fazla merkezde hasta aldık, hatta tek bir yerde değil, birden fazla ülkede aldık ve üstelik daha geç değil, erkenleri aldık diye bir şeyler söylemişler. Buna katılıp katılmadığını merak ettim.

Gülen Hatemi: Bunların hepsinden bahsedeceğim. Bu randomize kontrollü çalışma, Servet'in de söylediği gibi çok merkezli ve uluslararası yapılmış. Ve böyle olmasından çalışmanın güçlü tarafı olarak bahsediyorlar, bundan birazdan bahsedeceğim. Metotta bu kohortun nasıl bir kohort olduğu açıklanıyor, öncelikle çok erken artrit kohortu, 16 haftanın altında, en az bir tane inflame eklemi olan, daha önceden DMARD kullanmamış hastalardan oluşuyor. Bu hastalara, eğer travmaları, septik artrit, gut, sadece distal interfalangeal eklem tutulumu gibi semptomlarını açıklayacak başka bir olası nedenleri yoksa, bu çalışmaya katılmaları teklif ediliyor. Çalışma plasebo kontrollü, 120 mg tek doz metilprednizolon çalışması. Katılmayı kabul edenlere başlangıçta bir eklem sayımı yapıyorlar, akut fazlarını ölçüyorlar, RF'lerini, anti CCP'lerini ölçüyorlar, daha sonra da 2., 12., 52. haftalarda hastaları yeniden değerlendiriyorlar.

Endpointler ise 2010 ACR/EULAR kriterleri ile değerlendirildiğinde 1987 RA kriterlerine göre, 52. haftada RA gelişiminin sensitivite, spesifitesine, pozitif prediktif, negatif prediktif değerleri. Bunun için de altın standart olarak kabul ettikleri 3 tane durum var. Bir tanesi o araştırmacının romatoid artrit tanısı koymuş olması, ikincisi bir DMARD başlamış olması, üçüncüsü de bunlardan herhangi birini yapmış olması. Bu çalışmanın bence en önemli sorunlarından bir tanesi burada. Bu kriterler geliştirilirken zaten metotreksat almış olmayı romatoid artrit için standart kabul etmişlerdi. Bu kriterleri valide ederken geri dönüp yeniden endpoint olarak bir DMARD başlamayı aldığınız zaman burada bir circular logic söz konusu oluyor. Diğer taraftan herhangi bir DMARD başlanmış olmasını endpoint olarak aldığınızda spondiloartropatisi veya herhangi bir bağ dokusu hastalığı olanları da RA olarak sınıflandırmış oluyorsunuz. DMARD kriterini koymak hem circular logic olması açısından hem de aslında romatoid artrit dışındaki

bir sürü inflamatuvar hastalığı romatoid artrit kabul etmesi açısından sorun.

Sonuçta yazıdaki akış şemasında 12 aylık izlemi olan 303 hastanın dahil edildiğini görüyoruz. Bunlardan 144'ü RA, 98'i andiferansiye artrit, 61'inin de diğer tanıları var. RA'ların %7'si kadarı DMARD almıyor, nedeni de belirtilmemiş. Andiferansiye artritlerin %9'u, diğer tanıları olanların da %29.5'i DMARD alıyor. Bize yazıda verilen flow chart'ta bu şekilde gösterilmiş. Ancak yazıyı okuyunca anlıyoruz ki aslında bu çalışmaya randomize edilen hasta sayısı 386. Bu hastaların 80 tanesi kaybolduğu için, 6 tanesinde de 'protocol violation' olduğu için bu hastalar hiç değerlendirmeye alınmamışlar. Ve gerçekten hastaların çeyreği bu, yani hastaların dörtte biri değerlendirmeye alınmamış. Ancak böyle bir çalışma yaparken, mutlaka bütün kohortun dahil edilmesi gerekir. Kaybolan hastalar hafif oldukları için gelmeye gerek duymamış olabilirler. Veya çok ağır oldukları için başka bir yerde dertlerine çare arıyor olabilirler. O yüzden böyle bir çalışmayı yapmak için mutlaka ilk baştaki kohortun hepsini çalışmaya almak lazım. Üstelik orijinal SAVE çalışmasına baktığımız zaman görüyoruz ki, bu hastalar, çalışmadaki inflamatuvar artrit kliniklerine gelen hastalar arasında çalışmaya girmeyi kabul edenler. Daha hafif olan hastalar bu çalışmaya girmeyi ve 120 mg prednizolonu almayı kabul etmemiş olabilirler. O yüzden, gerçekten de bu dizayn, çalışmanın amacına uygun değil. Dahası orijinal çalışmadaki sayılar ile bu sayılar uymuyor, orijinal olarak 12 ay sonunda romatoid artrit, seronegatif spondiloartrit ve andiferansiye artrit olarak bildirilen hasta sayıları bu çalışmada verilenden farklı. Hasta sayılarındaki fark büyük değil, tüm gruplarda toplam 10 hastanın altında. Bu durum sonuçları çok etkilememiş olabilir. Ancak bu uyumsuzluk, çalışmada belirtildiği gibi, hastaların orijinal çalışmada araştırmacı ne tanı verdiyse o gruba alınmamış olabileceğini, ikinci çalışma yapılırken başka bir değerlendirmeye alınmış olabileceklerini akla getiriyor.

Çalışmada sensitivite, spesifite, pozitif ve negatif prediktif değerler verilirken yeni kriterler eskisi ile karşılaştırıldığında 'substantial increase in sensitivity and somewhat less specificity' deni-

yor. Ancak görülüyor ki sensitivitede substantial dedikleri fark 0.80 ile 0.65 arasındaki fark. Diğer taraftan somewhat farklı dedikleri spesifikite farkı ise, 0.63 ile 0.76 arasındaki. Neden birine substantial, diğerine somewhat dediklerini anlamak zor.

Bir başka sorunlu nokta pozitif prediktif değerler verilirken hasta sayılarının verilmemesi. Normalde sensitivitede ki farkın pozitif prediktif değeri daha fazla etkilemesini, spesifikitenin negatif prediktif değeri daha fazla etkilemesini bekleriz. Ama bu çalışmada, iki kriter seti arasında sensitivite açısından 'substantial' bir fark varken pozitif prediktif değerleri birbirine benzerdi deniyor. Buradaki pozitif prediktif değer hesabının bir problemi daha var: pozitif prediktif değer hesaplanırken, matematiğin içine hasta sayıları da giriyor. Yani sizin aldığınız popülasyondaki pozitif dediğiniz insanların sayısı da, bütün pozitif olan insanların sayısı da giriyor. Biraz önce tartıştığımız gibi, başlangıca göre ortadan kaybolmuş hasta var. Bu hastalar hesaba alınmadığında hesaplanan pozitif prediktif değer geçerli bir pozitif prediktif değer olmuyor. Çünkü baştaki gerçek sayıları, gerçek oranları biliyorsunuz.

Bu bahsettiklerimiz dışında area under the curve'ler de hesaplanıyor ve 2010 kriterleri için area under the curve'ün hem RA tanısı için hem de DMARD başlanmak için 1987 kriterlerinden daha iyi olduğu söyleniyor.

Çalışmanın iyi yönlerinden biri metotreksat kullananlara tüm DMARD kullananlardan ayrı olarak bakmış olmaları. Bu açıdan bir fark olmadığı söyleniyor. Diğer bir olumlu yön de 1987 kriterlerinde RA tanısı için semptom süresinin 6 haftanın üzerinde olması gerekliliği nedeniyle, semptom süresi 6 haftadan az olanları ayrıca analiz etmiş olmaları. Bu analiz ile de benzer sonuçlar aldıklarını bildirmişler.

Tartışmada, bu çalışmanın bu konuda yapılan ilk çok uluslu, çok merkezli kohort çalışması olduğu söyleniyor. Burada önemli bir sorun var. Orijinal çalışmada: "Patients with arthritis of short duration are frequently seen by rheumatologists after considerable delay" ve de o yüzden fast track referral policy'leri ve open door klinikleri veya ikisi birden alındı, diyor. Orijinal çalışmayı da çok ayrıntılı okudum. Bu hastaları değerlendirenler romatolog mu, böyle open door established romatoloji klinikleri gerçekten var mı, anlayamadım. Ama romatologlara ulaşmak zor diye, erken hasta alabilelim diye bir takım open door kliniklerden hasta almışlar ki bu bence, bu sonuçların genelleştirilebilmesi açısından sorun olabilecek bir durum. Ayrıca çok sayıda merkez dahil edilmiş, merkezler arasında sonuçlar açısından herhangi bir fark olup olmadığı da bildirilmemiş.

Son olarak yazarlar, "Moreover the 2010 criteria have been developed for classification, while diagnosis of RA can still be made in an individual patient even when failing the classification criteria and visa versa" demiş. Bu da buradaki kafa karışıklığını çok iyi özetleyen bir cümle oluyor. Sensitivitesi kötü diye yeni kriterler yapıyorlar. Ama daha sonra bu kriterleri spesifikitesi yüksek olsun diye, çalışmalarda kullanmak için yaptıklarını söylüyorlar. Ondan sonra da dönüp yine de tanı kriteri olarak da kullanılabileceğini söylüyorlar. Sonuçta burada büyük bir kafa karışıklığı var.

Teşekkür ederim.

Yusuf Yazıcı: Senin de bahsettiğin gibi, en büyük sorun ilaç başlamaları. Tabii bütün bu çalışmanın çoğunun Hollandalılar tarafından yapılması ve Hollandalıların anti-CCP pozitif romatoid artrit, anti-CCP negatif romatoid artrit farklı olduğuna inanmaları. Birisini anti-CCP pozitif görürlerse metotreksatı başlayacakları, görmezlerse başka ilaç verecekleri için o laboratuvar testleri de gereksiz yere daha fazla skorlanıyor, 3 point olsun, sedimentasyon, CRP olsun. Tabii o asıl sorun. Ötekisi, hiç kan testi pozitifliği yokken romatoid faktör var bir tek, ama gerek yok romatoid faktörün pozitif olmasına eskisinde, o yüzden spesifikitesi de eskisinde yeniden daha iyi oluyor, çünkü başka romatoid hastalığı olup da CCP'si ve romatoid faktörü pozitif olan bir sürü insan var. O insanları da katmaya başlıyor bu yeni kriter, eskisinde onlar klinik examde katılmıyorlardı, romatoid faktör önemli kalmıyordu. Dört tanesini tutturduktan sonra şey

oluyordu. O yüzden de ben ikisinin farklı olduğunu inanmıyorum zaten, arasında bir fark var, birinde anti-CCP pozitif, birinde negatif ama tedavi yanıtı olsun, prognoz olsun, bir farkları yok. Çok suni bir ayırım bence ve Hollandalılar buna inandıkları için ve datanın büyük kısmı Hollanda'dan geldiği için sonuç böyle olmuş, diyorum.

Gülen Hatemi: Çok güzel bir nokta gerçekten bu. Yanlış olarak bir ağırlık vermiş oluyoruz biz.

Nurullah Akkoç: Ben başka bir şey vurgulayacağım. Aramızda Tevfik Bedirhan Üstün var Dünya Sağlık Örgütü'nden. Çok güzel bir sunumu vardı. Maalesef çok fazla kişi katılamadı, ama onun sunumundan öğrendiğimiz bir şey Dünya Sağlık Örgütü'nün bu ICD-10 kodlarının yeni revizyonunda bir şekilde sınıflama kriterlerini de işin içine sokmaya çalıştıklarını söylediği için, tam da burada olması çok iyi oldu, çünkü romatoid artrit için olan şey aksiyal spondiloartrit için de oldu. Her ne kadar bunlar aslında diagnostik kriter gibi geliştirilmeye çalışılsa da sonuçta dönüp dolaşıp tekrar klasifikasyon kriterleri haline geliyor. Bu romatolojide çok zor yani bir diagnostik kriter geliştirmek bütün problem bence bu. Ana sorun şu herhalde, bu anti-TNF ajanları gibi pahalı ajanlar ortaya çıktıktan sonra, daha pahalı ajanlar çıktıkça geri ödeme kurumları için bir takım kriterler gerekiyor. Bence sıkıntı galiba bu.

Gülen Hatemi: Ama SUT'ta da kriter doldurma gerekliliği yok, DAS28'in belli düzeyin üzerinde olması gibi gereklilikler var.

Yaşar Karaaslan: Ben teknik kısmıyla çok değil ama politik açıdan, bir hastanın romatoid artrit olup olmadığını anlamak için burada klinisyenin metotreksat başlatmış olmasını kabul ediyorlar. Bence bu da sıkıntılı bir durum, gerçekten romatoid artrit olup olmadığını anlamak için klinisyen olarak en pratik şey, 2 sene sonra bir daha bakarsınız, 2 sene sonra hâlâ romatoid artritse o romatoid artrittir ama başlangıçta RA dediğiniz birçok hasta aslında undiferansiye artrit. "Ben verdim düzeldi, çok iyi oldu" aslında onlar ilaç verdiğiniz için düzelmiyor onlar zaten andiferansiye artrit. Yani bizim esas problemimiz andiferansiye artritle

romatoid artrit ayırmaksa kriterlerde, bu yeni kriterlerin ben çok işe yaramadığını, daha da işi karıştırdığını düşünüyorum. Gerçekten ne kadar iyi çalıştığını anlamak için 2 yıl sonra bu hastalar kaybolmuyorsa bir daha baksınlar. İki yıl sonra hâlâ romatoid artritse o zaman bu kriterler gerçekten iyi çalışıyor demektir. Yani bu yol iyi bir yol olabilir. Çünkü başlangıçta metotreksat başladığınız hasta bile gerçekten romatoid artrit olmaya bilir, yani pratikte metotreksat başladığımız sonra lupus çıkan hastalarımız da oluyor.

Murat İnanç: Ben o zaman başka bir görüş seslendireyim. Ben arkadaşlarım gibi düşünmüyorum, bence yeni kriterlerin katkısı var. Ben de anti-CCP ve romatoid faktör pozitif hastalığın farklı bir hastalık olduğunu düşünüyorum. Evet, bizim kendi yaptığımız çalışmada da negatif hastalarda da eroziv hastalığı olan var ama yani hepimiz kabul etmeliyiz ki öbür hasta grubu daha ağır seyrediyor. Klinik tecrübemiz de böyle, dünyadaki birçok çalışma da böyle. Bu çalışmada klasifikasyon, diagnostik kriter konusunda hocamın yazdıklarını tabii ki okudum ama üzgünüm ki mediko legal olaylar sürdüğü sürece dünyada uzun bir süre diagnostik kriter olmayacak. Çünkü Yaşar Bey'in dediği gibi 2 sene beklenip ondan sonra kendisinin kötü tedavi edildiğine inanıp, ondan sonra da bu diagnostik kriteri yapana dava açıp kazanabilirler. Bizde olmasa da başka yerlerde. Dolayısıyla tanı kriteri diye bir şeyin yani bu kriter konusunda oldukça uzun çalışmış biri olarak söyleyeyim, hem lupus kriteri hem skleroderma kriteri konusunda, tanı kriteri diye bir şey uzun bir süre olmayacak. Bu belki özel kliniklerde olabilir yani göz kliniği "Ben Behçet'in gözünü şöyle tanıyım" recommendation olarak yapabilir ama bütün kriterlerin istisnası olacağı için başlarına legal bir şey gelmesini istemiyorlar. Böyle bir mesele de var. Onun için birçok kriter sınıflandırmaya kriteri olarak devam edecek.

Brian Feldman: American College of Rheumatology olarak şu anda yayımlanan diagnostik kriteri ile sınıflandırma kriteri farkını anlatan bir çalışmamız var. Ve neden ACR'nin diagnostik kriter yapmayacağını birçok sebeple anlatıyor bu çalışma. Biraz önce bahsettiğiniz nedenlerin hepsi bu yazıda var. Bunların sınıflandırma kriteri olduğu gerçeği göz

önünde bulundurulduğunda ve amaç çalışma için homojen hastaları bir araya toplamak olduğu için spesifitenin daha yüksek olması gerekmiyor mu?

Gülen Hatemi: Evet iddiaları o. Daha yüksek olmalı.

Yusuf Yazıcı: Ben bu konuda sunum yapacağım. Bu bahsettiğiniz problem sonucu nasıl etkiliyor, romatoid artrit olmayan hastalara romatoid artrit deniyor. Tersini yapmayı denerken lupus hastaları randomize kontrollü çalışmalara RA olarak giriyor.

Nurullah Akkoç: Buna cevaben, bence aslında bu kriterleri yaparken diagnostik kriterler yapmayı amaçladılar. ASAS klasifikasyon kriterlerinde de AS için bu oldu ama daha iyi rakamlara ulaşamadılar, bu yüzden sınıflandırma kriteri olarak adlandırdılar.

Gülen Hatemi: Ama varılan noktada daha iyi sepisifiteye ulaşılamadı, yani sınıflandırma kriteri olarak da daha iyi olmadı.

Brian Feldman: Buna cevabım, son 6 yıldır ACR'de yeni kriterlerin gelişimini denetleyen komiteye başkanlık ediyorum. Bu daha en başından sınıflandırma kriteri amacıyla geliştirildi. Şu olabilir, belki çalışmadaki uzmanların çoğu sınıflandırma kriterinden çok tanı kriterleri düşünmüş olabilirler ve bu yüzden yüksek sensitivite düşük spesifite ortaya çıkmış olabilir.

Murat İnanç: Ama yetersiz tartışıyoruz, ben itiraz ediyorum çünkü olanları anlamamız lâzım. Şimdi bu çok önemli bir paper, duyarlılıkta da bir artış var bugün kriterlerde. Yeni lupus kriterine bakarsanız, orada da spesifite daha düşük, sensitivite daha yüksek. Tabii burada tedavi ajanlarının çok olmasının ve tedavi ajanlarına daha geniş bir alan tanınması isteğinin, yani yumuşak bir ifadeyle böyle söylüyorum, bunun da etkisi var. Skleroderma kriteri de daha duyarlı ve daha az spesifik olacak, yani bundan sonra böyle olacak, trend böyle.

The ability of synovitis to predict structural damage in rheumatoid arthritis: a comparative study between clinical examination and ultrasound

Maxime Dougados,¹ Valérie Devauchelle-Pensec,² Jean François Ferlet,³ sandrine Jousse-Joulin,⁴ Maria-Antonietta D'Agostino,⁴ Marina Backhaus,⁵ Jacques Bontin,⁶ Gérard Chales,⁷ Isabelle Chary-Valckenaere,⁸ Philip Conaghan,⁹ Richard J Wakefield,¹⁰ Frédéric Etcheperre,¹¹ Philippe Gaudin,¹² Walter Grassi,¹³ Désirée van der Heijde,¹⁴ Xavier Mariette,¹⁵ Esperanza Naredo,¹⁶ Marcin Szkudlarek¹⁷

SİNOVİTİ SAPTAMADA KULLANDIĞIMIZ FARKLI YÖNTEMLER İLERİDEKİ RADYOGRAFİK HASARI SAPTAMADA BİRBİRİNE BENZER SONUÇLAR VERİR Mİ?

Sibel Zehra Aydın

Tartışmak istediğimiz araştırma romatoid artitte yapılmış bir ultrason çalışması. Çok merkezli, çok uluslu bir çalışma olarak tasarlanmış. Çalışmanın ana amacı sinoviti saptamada kullandığımız farklı yöntemlerin (fizik muayene, griskala ve Doppler ultrason) ilerideki radyografik hasarı saptamada birbirine benzer sonuçlar verip vermediğini göstermek. Araştırmaya 16 merkezden, alanda iyi bilinen ve çok sayıda araştırması bulunan çok sayıda araştırmacı katılmış gözüküyor. Araştırmacılar dizayn başlığı altında asıl amacın yanısıra tekrarlanabilirlik, eksternal validasyon, değişimi saptama ve ayırma kapasitesi (reliability, external validity, sensitivity to change and discriminate capacity) gibi hedefleri olduğunu da belirtiyorlar.

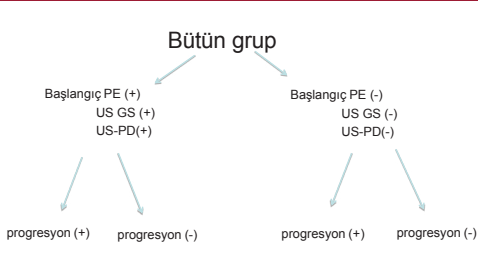
Hasta alım kriteri son derece geniş belirlenmiş: Romatoid artrit hastası olmak ve klinisyene göre antiTNF ihtiyacı olması. Çalışmaya alırken tek objektif kriterleri, hastaların 6 ya da daha fazla şiş eklemi olması; başka hiçbir objektif kriterimiz yok.

Çalışma 4 aylık bir takip çalışması gibi görünüyor. Bazalde bir klinik değerlendirmeleri var, ultrasonları var ve direkt grafileri çekiliyor. Dördüncü ayda ultrasonları var, klinik değerlendirmeleri var, ikinci yılın sonunda hiçbir klinik değerlendirmeleri yok, sadece direkt grafileri çekiliyor.

*Romatoloji Bilim Dalı
Koç Üniversitesi Tıp Fakültesi*

Emire Seyahi: Dört aylık antiTNF tedavisi veriyorlar, ondan 2 sene sonra böyle bir çalışmayla

İstatistik (1)



1. Slayt

o datayı değerlendirmek mi istiyorlar? Ellerinde herhalde mevcut olan bir şey vardı, bunu uzatmışlar. Bu arada ne tedavi aldıkları hakkında bilgimiz var mı?

Sibel Aydın: Hiçbir fikrimiz yok.

Üç grup analiz yapıyorlar, önce bazalde power Doppleri pozitif, fizik muayenesi pozitif ve gri skalası pozitif olan eklemleri belirliyorlar ve sadece o eklemlerin kaçında progresyon olup olmadığına bakıyorlar. İlk karşılaştırmaları bu.

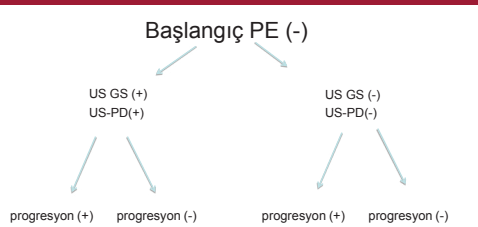
İkinci karşılaştırmaları, bir modalite negatif, diğer modalite pozitif olduğunda, sub gruplarda progresyon ne kadar. Yani fizik muayenesi normal, ultrason /gri skalası pozitif hasta ne kadar progresse oluyor? Ultrasonu negatif, fizik muayenesi pozitif hasta ne kadar progresse oluyor, bunları karşılaştırıyorlar.

Üçüncü olarak yaptıkları değerlendirmede 4. ayda hastaya tekrar baktıklarında hasta remisyonunda ise o eklem ya fizik muayeneye göre ya ultrasona göre bu hastalarda sürekli remisyon olması halinde radyografik progresyon anlamında bir farklılık var mı?

İlk başta 70 küsur hastayı çalışmaya dahil ediyorlar ama izlemin tamamlanabildiği hasta sayısı 59. İlk giriş özelliklerine baktıklarında bir farklılık olmadığını data vermeden söylemişler. Hasta başına 32 eklem tarıyorlar, dolayısıyla 1888 eklemeye ulaşıyorlar ve ultrasonun sensitivitesini daha yüksek olarak buluyorlar. Ama bölgelere ayırıp baktıklarında aslında birçok bölgede ultrason daha sensitif gibi değil, fizik muayene daha çok sinovit olduğunu söylüyor. Gri skalaya göre sadece ayak için içine katıldığı zaman ultrasonun sensitivitesi daha yüksek gibi gözüküyor.

Servet Akar: Burada aslında yöntemde klinik sinoviti, ultrasonun hem gri skalası hem de power Doppler skorlamasını ayrıntılı veriyorlar, peki araştırmacılar sonuçlarda bu skorlamaların sonuçlarına göre mi bu sonuçları sunmuşlar yoksa, var-yok şeklinde mi?

İstatistik (2)



* Aynı analiz US (-) ama PE (+)

2. Slayt

Sibel Aydın: İkisini de vermişler aslında, bir “var” ya da “yok” diye ayırıyorlar bir de derecelendiriyorlar, onun 0 (sıfır) ile 3 arasında bir skorlaması var, 0 ile 1’i de, yani 1 kabul edilebilir dediklerinden bunu da 2 ayrı analiz olarak veriyorlar.

Dördüncü aydaki değerlendirmeyi aynı klinisyen, aynı ultrasoncu yapıyor. Direkt grafler tek bir merkezde skorlanıyor. Metotta pek çok sıkıntı var, onlara tek tek geleceğim.

İlk sonlanım noktaları, ikinci yıldaki progresyonda bir fark var mı? Rakamlara ayrıca girmeye gerek yok ama sinoviti değerlendirmede ne kullanılırsa kullanılsın bir eklemden herhangi bir yöntemle sinovit var diyorsanız progresyon riski daha fazla gibi görünüyor; odds ratio olarak baktığımızda klinik değerlendirme sanki ultrasondan daha iyi gibi gözüküyor.

Emire Seyahi: Bu araştırmacıların daha önceki yayınlarındaki bulgularıyla çalışmıyor mu?

Sibel Aydın: Çelişiyor. Metottaki sıkıntıları yüzünden çeliştiğini düşünüyorum, ona geleceğim.

Diğer modalite normalse progresyon riski ne kadar diye baktıkları zaman, yine modaliteden bağımsız olarak tek bir sinovitin gösterilmesi progresyon riskini arttırıyor. Fizik muayenesi normal, ultrasonu pozitifse de bu hastaların progresyon riski var, tam tersine ultrasonu negatif ama fizik muayenede sinovit saptarsak da progresyon riski var gibi görünüyor.

Persistent sinovit açısından baktığımızda da eğer ultrasonda hâlâ persiste eden inflamasyonu görüyorsak bu hastaların progresyon riski var gibi gözüküyor. Ama sadece fizik muayenede “Ben sinovitinin devam ettiğini düşünüyorum” diyorlarsa bunun progresyon riskinde anlamlı bir artış var gibi gözüküyor.

Hangi noktalarda kısıtlılıkları var? Önemli bir nokta çalışmaya seçilen hasta grubu, rutinde izlediğimiz hasta grubunun tamamını yansıtıyor mu? Aslında yansıtmıyor, çünkü çok seçici olmuşlar, sadece antiTNF gerektiren bir grup var, bunu

bütün RA grubuna uyarlamak mümkün değil, ama bunun içinde de her antiTNF gereken grup da birbirinin aynı olmayabilir. Benim hastam buraya uyar mı, bilemiyorum.

On altı merkez var ve ultrasoncuların bireysel olarak özelliklerine baktığımızda gerçekten uluslararası kurslarda eğitmenler ve bu işi iyi yaptıklarını biliyoruz. Ama kendi kullandıkları kriterlere göre, 70 tane ultrason yapmış olmak bu çalışmaya girebilmek için yeterli gibi gözüküyor. Bu kesinlikle hem yeterli değil hem de uluslararası kabul görmüş rakam değil. Yani biz bu çalışmayı Türkiye’den yapmış olsaydık ve 70 ultrason yapmış merkezi katmış olsaydık, çok büyük olasılıkla ciddi eleştiri alırdı bu rakam.

Farklı merkezlerde değişik cihazlar kullanıyorlar. Kullanılan cihazların bir kısmı aslında eski cihazlar. Cihazlara ait çok fazla detay yok ama sadece kullandıkları pronun frekansını 7-12 mHz olarak veriyorlar. Halbuki hepimiz rutin hayatta yüzeysel dokuyu görmek için daha yüksek frekansa çıkıyoruz ve frekans ne kadar düşerse kullandığımız datanın güvenilirliği o kadar bozulmaya başlıyor. Sensivitesini çok birebir etkileyecek bir parametre. Bir de çok indirekt bir sonuç ama bu merkezlerin hemen hepsi son 5-6 yıldır bundan çok daha iyi problemler kullanıyorlar ve bu yazıları sürekli yayımlıyorlar. O yüzden acaba bu data çok daha erken mi toplanmıştı, ultrasonları gerçekten çok mu eskiden yapılmıştı gibi bir endişem var.

Tanımlamaları netleştirmişler mi? OMERACT’ın sinovit tanımlaması çalışmalarda en çok kullanılan tanımlama ve “Bunu kullandık” diye vurguluyorlar, burada bir soru işareti var gibi gözüküyor ama bir ultrason çalışmasında makine ayarlarının ne olduğu, ekibin bunda nasıl konsensusa vardığı çok önemli. Söyledikleri bir cümle var; “Biz bu çalışmayı yapmadan önce bir araya geldik, 1,5 günlük bir konsensus toplantısı yaptık ve orada tüm tanımlamaların üzerinde hemfikir olduk,” diyorlar ve buna referans veriyorlar. Bu dedikleri referansa gidip baktığımızda ise o çalışmada 7 ultrasoncu var, bu 7 ultrasoncu Belçika ve Fransa’dan gelmiş ultrasoncular gibi gözüküyor. Yedi aktif hasta ve bir cihaz var. Yani aslında bu çalışmada kullanılan

cihazlar, araştırmacılar pek örtüşmüyor gibi gözüküyor. Yirmi sekiz eklem değerlendirmesini yapıyorlar ve ayak bunların içerisinde değil. Sonuca baktığımızda da aslında bu grubun tekrar edilebilirliği son derece düşük gibi gözüküyor. Hatta o yazının tartışmasında bunu, “Diğer mevcut literatüre göre çok daha düşük bir tekrarlanabilirliğimiz vardı” diye vurguluyorlar. Dolayısıyla bahsedilen konsensus mevcut çalışmanın konsensusu mudur, çok net değil. Ama onun dışında her hastayı aynı şekilde takip etmişler, uygulamaları standart, körlüğü sağlamışlar.

Çalışmanın amaçlarından bir tanesi tekrarlanabilirliğinin test edilmesi iken buna ait hiçbir veri olmadığını görüyoruz. Site edilen önceki yazıya gittiğimizde tekrarlanabilirliğinin kötü bile olduğu gözüküyor. Ve tekrar edilebilirlik, güvenilirlik ultrason çalışmasında çok önemli bir nokta.

Kaç merkezden hasta kattıklarını söylemiyorlar, ben 16 merkezi sadece yazar listesinden bakarak söylüyorum. Ne kadar pratik olarak hasta katılım kısmına katılmıştı, burası belli değil. Ama nadir bulunan bir hastalıktan ya da nadir bulunacak bir hastalık aktivitesinden bahsetmiyoruz. Neden tekrarlanabilirlik bu kadar önemli bir sorunken 52 hasta için 16 merkeze gitmeleri gerekli? Ben olsam ne yapardım: Bunu az sayıda merkezde, sadece araştırmacılar arasındaki uyumun iyi olduğunu biliyorsam ve kendi tekrarlanabilirliğim iyiye o noktada çalışmaya girerdim.

Önemli dedikleri sonuçlardan bir tanesi de ultrason ve fizik muayene arasındaki uyumsuzluk. Bu rakamları yaklaşık hesapladım ama fizik muayenede normal buldukları eklemlerin %40 ultrason pozitifliği var, ultrasonu normal eklemlerin %30’unda fizik muayene pozitifliği var.

Servet Akar: Normal bulunan dediğimiz eklemler aslında normal eklemler ama pozitif eklemler dediklerimiz arasında şüpheli eklemler de var. Şüpheli dediklerimizde sinovit var gibi kabul etmiş gibi görünüyor sonuçlarda.

Sibel Aydın: İki tür analiz var, yani şüpheli dedikleri grade 1, 0-3 arasında hem fizik muayene hem

ultrason için. 0 ile 1’i aynı kategoriye koyup 0-1’e karşı 2 ile 3’ü de karşılaştırıyorlar, 0’a karşılık 1-2-3’ü de karşılaştırıyorlar.

Servet Akar: Demek istediğim bütün odds ratio değerleri 0 vs 1-2-3?

Sibel Aydın: Burada verdiklerimiz öyle.

Servet Akar: Yani şüpheli şeyler de var.

Sibel Aydın: Tabii. Supplemente gittiğimiz zaman görüyoruz. Normal yazıda o rakamlar yok.

Emire Seyahi: Ben daha da temel bir soru soracağım. Fazla ultrason çalışmalarına aşına olmadığımından belki, cahilce olabilir, bağışlayın. İlk başta okuduğumda bu eklemlerin toplanmasını, bir tarafa, iki dik oturmuş şekilde toplanmasına ne diyorsun? Hani insan bazında hasta bazında değil de bu araştırmalar, bir insanın inflame bir, benim 2. Mcp’m inflame ise 3. de büyük olasılıkla öyledir, değil mi?

Sibel Aydın: Kesinlikle öyle. Eleştirilerimden biri de o zaten. Kabaca toplam 140 eklem tutulmuş, progrese olmuş gibi gözüküyor. Ama bu 140 eklem, 52 hastada 3’er eklem tutulmuş da olabilir, 5 hastada 30 eklem tutulmuş da olabilir, bu konuda hiçbir fikrim yok. Yani ben bu datayı direkt alıp kendi hastama uyarlayabilir miyim?

Emire Seyahi: Peki genelde nasıl yapılır ultrason? Hasta bazında mı tercih edilir?

Sibel Aydın: Hasta bazında yapıyoruz, subgroup analizi yaparken bölge bazında yapıyoruz. Çünkü bölgeden bölgeye de değişir. Yani el bileğinin ultrason bulgusuyla, ayak bileğinin ultrason bulgusu arasında fark var. O yüzden bölgelere ayırmak da mantıklı.

Emire Seyahi: Bu eskiden beri Hasan Hoca’nın hep eleştirdiği bir şeydir. Behçet’in gözlerinin analizini yaparken iki göz mü alalım, tek göz mü alalım, yoksa gözleri toplayalım mı tartışmasına getiriyor.

Sibel Aydın: Burada muhtemelen sorun şöyle: Radyolojik progresyonu değerlendirirken çok geniş bir skalada çok ufak rakamlardan bahsediyoruz ya, o bulgu gerçekten bizim o eklemimize denk düşmüyorsa belki bu değerlendirme daha doğru olabilir ama bunu analiz etmek lazım, kesinlikle hastaya da uyarlamak lazım. Hastaların kaçını ilgilendiriyordu konusunu da bence netleştirmek gerekiyor.

Bir vurgulamam gereken şu olabilir. Bütün ‘confounding’ faktörleri için içine katmışlar mı? ‘Generalized estimating equations analysis’ denen bir yöntem kullandıklarını söylüyorlar. Ama bu sadece metotta var, ondan sonra bununla ilgili hiçbir detay yok ve bunun için de cinsiyet, hastalık süresi gibi progresyon etkili bazı faktörleri koyuyorlar ama RF’si yok, ACPA’sı yok, sigara öyküsü yok, kullanımı yok, yani gerçekten de belki riski olan hastaları tanımlamak mümkün olabilir, ultrason sonuçları sadece bir grubu ilgilendiriyor olabilir. Bunu hiç bilmiyoruz. Bizim hastamız bu raya uyar mı, bilmiyoruz.

Bir de 4. aydaki inflamasyondan bahsediyor. Ondan sonraki 20 ay boyunca ne olduğu hiç belli değil. Belki de orada sadece erken remisyona girebilmekten bahsediyordu, ondan sonraki aktivasyonlar ya da ilaç kullanımlarının etkisi hakkında da hiç fikrimiz yok.

Hasan Yazıcı: Burada analizlerde iki olay var. Birincisi 4. aydaki klinik - ultrason ilişkileri ve sonra da 2. yıldaki ultrason. Şimdi burada ilk ve 4. ay arasında odds ratiolar kullanıyorlar. Hâlbuki elle-rindeki veriye göre herkesin başlangıçta sinoviti var, ultrasonla biliyorlar. Dördüncü ay sonunda ileriye dönük prospektif veri kullanıyorlar. O prospektif veriyi kullandıklarında, beklenen odds ratio değil, niye bunu odds ratio kullanıyorlar, birinci soru bu. Odds ratiolar daima rölatif risklerden sayısal olarak daha fazla büyüklerdir. Eğer karşında odds ratio ile rölatif riskin farkını pek bilmeyen bir popülasyon olduğunu düşünürsen, ben bunları rölatif risk kullanmak istemiyorum deyip prospektif olmasına rağmen, öbür istatistiği kullanıp, olan ilişkiyi, daha yüksek gibi gösterebilirsin. Bu bir.

İkincisi, dördüncü ay ile ikinci yıl arasında buradaki veri retrospektif olabilir ama bunu anlatırken de odds ratio kullanıp diyorlar ki, 4. aydaki klinik bulgu sinoviti, 2. yıldaki progresyonu gösterdi. Hâlbuki tam cümle öyle değil, onun odds ratiosu buydu, o zaman sinoviti varsa üç misli daha fazla erozyon olması ihtimali var, diye düşünebilirsin. Hayır, erozyon olanlar arasında 2. yılda, 4. ayda sinoviti olanların oranı, olmayanlara göre daha fazla. İki cümle birbirinin aynı değil. Yani ‘miyokard enfarktüsler arasında kolesterol çok yüksek olur’la, ‘çok yüksek kolesterolü olanlar iki sene sonra miyokard enfarktüsü olur’lar, farklı şeyler aslında. Bunu bilmemelerine imkan yok.

Prof. Lesaffre’e bu konuda ne diyor?

Emmanuel Lesaffre: Öncelikle söylemem gereken, bu tür makaleleri okumak benim için çok karmaşık. Detayların arasında kayboluyorum. Birinci soruna cevaben, rölatif risk yerine odds ratio kullanmak, kullandıkları istatistiksel tekniğin tercihi doğrultusunda. Genelleştirilmiş tahmin denklemleri kullandıklarını söylemişler, buna göre aynı bireydeki eklemelerin birbiriyle bağlantılı olduğunu hesaba katarak sofistike lojistik regresyon olarak bakabilirsin. Yani lojistik regresyon bağımsız gözlemlerde kullanılmış, eğer şimdi bu bağ için düzeltmeyi yapmadan, bunu eklemelere uygularsan o zaman sorun var demektir. Lojistik regresyonda denklemleri tahminleyen software mevcut, çok net değil, rölatif risk model için de var, yani aslında genelleştirilmiş tahmin denklemi kullanmalarının bir nedeni de bu olabilir. Neden genelleştirilmiş tahmin denklemi kullandıklarını belirtmiyorlar. Belki okuyucu için ek bilgi olabilir. Sonar öngöründen söz edildiğinde, ben rölatif riskin geleceği tahmin ettiriciliğinden daha gerçekçi öngörüler görmek isterim. Sorunun cevabı, odds rationun rölatif riskten daha büyük olduğu doğru, ama büyük ihtimalle odds ratio analizini yapan softwareleri olduğu için odds ratio kullandılar.

Hasan Yazıcı: Ama doğru olan elindeki dataya uygun software kullanmaktır, software’e göre datana bakmak değil.

Emmanuel Lesaffre: Katılıyorum ama hayat bu. İnsanlar neye sahipse onu kullanıyorlar. İlişkiyi anlamlı ya da anlamsız diye belirleyebilirsin, ama yapılabilir. Belki sonuçtan memnun kalırsın ya da istediğin şekilde yorumlarsın. Eğer odds ratioyu rölatif risk olarak yorumlarsan o zaman sorun var. Analizin kendisi problemli değil eğer doğru yorumlanırsa.

Yusuf Yazıcı: Bir de anladığım kadarıyla progresyon olup olmadığını 0-1 diye mi yapmışlar? Ama tabii bu 1 olması için en azından, Van der Heijde cihazında beş ünite olması lazım. Beşin altını insan gözü ayırt edemiyor. Tek tek baktıkları için en az 5 olmalı. Beş ünite de ilerleme çok görülen bir şey değil, %20 falan gibi.

Sibel Aydın: Eklem bazında bakıyorlar. Diyorlar ki, o eklem için, eklem aralığında daralma ya da erozyon var mı? Bu ikisini de ayrı ayrı veriyorlar, yani o eklemde erozyonda ilerleme var, eklem aralığında ilerleme, daralma var ya da herhangi birisi var. Supplementta verdiği şeylerden birisi o. Genel skoruna bakarak karar vermiyor progresyona.

Yusuf Yazıcı: Total skorları nedir bu hastaların sonuçta?

Sibel Aydın: Bilmiyoruz, hasta bazında data yok çünkü.

Nevsun İnanç: Ben cevap vereyim. Şimdi, radyolojik skorlama ile ilgili Van der Heijde grubunun bir de 'simplified erosion narrowing score' diye bir şeyi var, onda tek tek bir patoloji varsa 1 olarak veriyor. Ve bunun da geçen yıl ya da ondan önceki yıl, diğer kendi yaptıkları bu 0-448 ile korelasyonu da gösterdiler. Burada radyolojik skorlama olarak sanıyorum o simplified kullanmışlar. Bir de Emire'nin sorusuna, bu eklemleri niye tek almışlar, niye kişileri tek başına almamışlar, onunla ilgili bir şey söylemek istiyorum, bunu ben de yapıyorum çünkü. Aslında ultrason çalışmalarında yapılan bir şey bu. Şimdi burada sadece sinoviti değerlendiriyorlar. Yani o kişinin RF'si, anti-ccp'si diğer klinik özelliklerini işin içine katmıyorlar, sadece sinoviti değerlendirdikleri için sinovitin etkisi araştırılıyor. Bu nedenle, eklemler tek tek sanki bir bireymiş

gibi, tek faktörmüş gibi alınabiliyor, çünkü ultrason vakit alan bir şey. Yani bir hastayı bütün olarak aldığınız zaman, 10 eklemine, 20 eklemine aldığımız zaman, örneğin burada 20 eklemi, di-yelim 6 eklemi varsa, onlar daha sonra 6 eklemi tekrar yazıp, tekrar ultrasonla değerlendiriyorlar, bu da çok vakit alıyor. O nedenle tek tek birey olarak kabul edildiği zaman, istatistikte küçük farkları değerlendirme şansı oluyor.

Hasan Yazıcı: İşte o şans her şeyi tamamen mahvediyor. Yanlış olan o. Yani bir insanın sağ el bileğinde sinovit olduğu zaman, sol el bileğinde sinovit olma şansı daha fazla, 3. daha fazla, 4. daha fazla. O bakımdan çok klasik, yıllar öncesinden romatoid artritte erozyon çalışmasını hatırlıyor musunuz? Kirwan çalışması. Kirwan çalışmasında kortizon iyi geliyor mu, 100 hasta 200 el idi. Çok klasik yanlışlardan bir tanesi. Yüz hastanın, 200 eli üzerinde yapamazsın. Çünkü girerken o ki-karelerin hepsine, her bir gözlem öbüründen bağımsız olması lazım. Bir insanın sağ kulağı, sol kulağından bağımsız. Çift organ sorunu.

Nevsun İnanç: Hocam, klinik parametrelerini katmadıkları için belki olabilir diye düşünüyorum. Burada sadece sinoviti değerlendiriyorlar.

Sibel Aydın: Ama bazı parametreleri katıyorlar. Her şeyi katmıyor değiller.

Ahmet Gül: Tartışmaya zemin olması için aynı tartışmayı romatoid artritte de psöriatik artritte de aynı istatistik yöntemle yapabilir miyiz? Yani çok sayıda eklem tutulma şansının yüksek olduğu bir hastalıkla, tek bir eklemde tutulma riskinin yüksek olduğu bir durumdaki karşılaştırmada aynı istatistik çözümler bize doğru yanıtları verir mi? Yani bu psöriatik artrit çalışması olsaydı örneğin. Sadece yöntem tartışması için diyorum.

Hasan Yazıcı: Bu yöntem meselesi değil. İlk olarak bir düzeltme faktörü kullanılabilir. İkincisi sağ el, sol el olarak ayrı verilebilir. Biyolojik olarak anlamı yok, ama yaptığın hesabı, en sonunda bir olasılık yapıyorsun. Bizim misafirlerden bu konuda söz almak isteyen var mı?

Yusuf Yazıcı: Son bir şey söyleyeceğim. Patoloji var mı, yok mu, dediniz ama patoloji, mesela benim sol parmağım da 4 defa kırıldı. Bir patoloji var ama fonksiyonu etkileyen bir şey yok. Fonksiyonu etkilemiyorsa orada patoloji olduğunu söylemek de zor. Sırf o yüzden klinik tedaviyi bilmeden, yani ormanda ağaç düşerse, kimse duymamışsa ses çıkarmış mıdır, orada patoloji var diyorlar ama o bence bir patoloji değil. Hem total skora bakmadıkları sürece bütün uzun dönem sonuçlarında total skorlar ilişkili, o da iyi bir korelasyon değil hasta sonuçlarına göre mesela. Yani bunu neye dayandırmışlar, onu da anlayamadım.

Emmanuel Lesaffre: Bence bu sorunun cevabı zaten verildi, burada kullanılan metot genelleştirilmiş tahmin denklemi, yani şüphesiz her bireyde iki el ya da 200 eklemın bağlantılı olduğunu hesaba kattılar, p-değeri için göz önünde bulundurdular.

Metot, daha önce de söylediğim gibi bir çeşit lojistik regresyon yöntemi, tek öznede çoklu gözlem ve bağlantılı data. Ve bu tür bağlantılı datalarda, bağımsız datalar kadar fazla bilgi olmaz. Yani bağlantılı datayı aynı şeyi 100 kez tekrar ederek alabilirsin. Ya da şöyle söyleyeyim, her zaman öğrencilerime verdiğim bir örnek vardır; Hasan hakkında aynı dedikoduyu 100 kez yaparım, bir

süre sonra “Hadi ama biz bu hikâyeyi zaten biliyoruz” derler. Ama 100 farklı kişi Hasan ile ilgili aynı dedikoduyu yaparsa dedikoduya inanmaya başlarlar. Bağımsız ve bağımlı arasındaki fark bu.

Hasan Yazıcı: Aradıkları şey 4 ayda, eklemlerdeki şişliklerdeki sinovit. İkinci yılda ise erozyon sayısını arıyorlar. Hasan’ın 25 eklemi katılmışken, John’un 2 eklemi katılmış olabilir, yani 25 eklemlerle Hasan daha ağır basacaktır hesapta.

Emmanuel Lesaffre: Hayır, bu 25 eklemın bağlantısı hesaba katılıyor, neden söz ettiğini anlıyorum, bazıları bir şekilde daha ağır basıyor ama her eklemın hesaba dâhil edildiğini düşünüyorum.

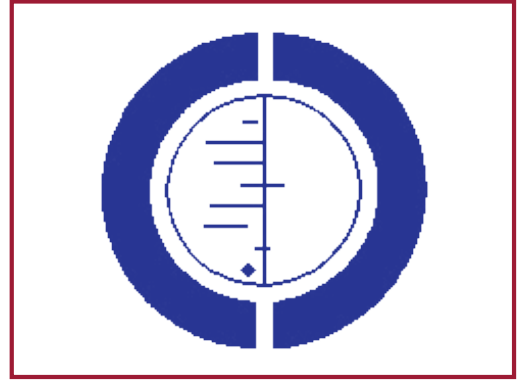
Hasan Yazıcı: Ama ilişkili olan hesaba dâhil edilir. Söylemek istediğim rahatsız edici olan, korelasyon. O zaman bize bunu öğret. Net korelasyon problemleri, söylemek istediğim bu.

Emmanuel Lesaffre: Evet doğru. O zaman 30 saatlik kursa hazır mısın?

Emire Seyahi: Artık bunu bitiriyoruz. Sibel çok teşekkürler. Sonuçta oldukça iyi bir dergide yayımlanmış, tamamen alanındaki en önemli yazarlar tarafından yapılmış fakat niye yapıldığını anlayamadığımız bir çalışma.

KANITA DAYALI TİP İÇİN NE KANIT VAR?

Hasan Yazıcı



1. Slayt



2. Slayt

Sevgili arkadaşlar bu 1. slayt, malum, kanıta dayalı tıp, Cochrane Foundation logosu. Büyük bir tevazu içinde (!) onun yanına ben de kendi logo-mu koydum (2. slayt) . Benim logo kör çalışmanın önemini simgeliyor. Behçet hastalarına iğne yapınca 48 saat sonra paterji oluyor ya, bir hastada paterji testi pozitif mi, değil mi diye bakıyoruz. Tabii yorumu yapan hekim testi koyduğu hastanın gerçekten Behçet'ten şüphelendiği hasta mı, yoksa sadece tekrarlayan aftöz stomatiti olan bir hasta mı olduğunu baştan bilmezse – yani kör gözlem yaparsa – paterji testinin sonucunu çok daha yansız yorumlar. Logo bunu simgeliyor.

Önce kısaca 'Evidence-based medicine', yani 'kanıta dayalı tıp' nedir, onun tarihinden bahsedeceğim. Ben kanıta dayalı tıbbı, eski ve yeni kanıta dayalı tıp diye ikiye ayırıyorum.

*Romatoloji Bilim Dalı
İstanbul Üniversitesi
Cerrahpaşa Tıp Fakültesi*

Esasında kanıta dayalı tıp çok eski bir kavram. Daha 17. yüzyılda ünlü İngiliz düşünür F. Bacon Hipokrat'a çatıyor. "Hipokrat'ın aforizmaları, adı üstünde kanıta dayalı değil." diyor. 18., 19. yüzyılda İngiltere'de tıbbi istatistik başlıyor. Arkasından stetoskop, mikroskop bulunuyor. Lister'in aseptisi antisepsiyi, Pasteur'ün mikropları keşfetmesi kanıta dayalı tıp değil de nedir? Aynı şey röntgen ışınları, antibiyotikler, polio aşısı için de geçerli.

Yeni kanıta dayalı tıp ise bence 1948'de başlıyor. 1948'de ne oluyor? İngiliz Tıp Araştırma Grubu'nun (MRC) 1948 yılında yayımladığı, akciğer tüberkülozunda streptomisin işe yararlığını sınavan çalışma, kanıta dayalı tıbbin kalesi olarak kabul edilen kontrollü ilaç çalışmasının ilk örneği olarak kabul edilir. Bu arada etikçiler için ufak bir parantez açayım, esasında bu çalışma için hastaları randomize ederken hastaları sadece değişik iki merkezde izlemişler ki, hastalar açısından, körlüğü sağlasınlar. Çalışmanın yapıldığı yıla dikkatinizi çekerim. Ünlü Nuremberg deklarasyonu 1947 yılında, MRC çalışması 1 yıl sonra. Ancak ortada aydınlatılmış onam filan yok!

Kanıta dayalı tıbbi anlatırken kullandığım, çok sevdiğim, yıllardan beri öğrencilerime anlattığım bir çalışma var. Çalışma tonsillektominin gerekliliği hakkında (1). 1960'ların başında New York'ta bir grup çocuk alıyorlar ve 3 tane kulak burun boğaz uzmanına "Bu çocuğun tonsillektomi'ye ihtiyacı var mı?" diye soruyorlar. Uzmanlar, kafa kafaya verip, tartışıp sonunda çocukların 1/3'ünün tonsillektomiye gereksinimi olduğunu söylüyorlar. Çalışmayı düzenleyenler bundan sonra ilk 3 uzman tarafından muayene edilip de tonsillektomi olmalarına gerek olmadığı söylenen çocukları ikinci 3 kişilik uzman grubuna gösteriyorlar. Bu ikinci grup da ilginç olarak, çocukların kabaca 1/3'ünde bademcikler alınsın, diyor. Çalışmanın son aşamasında ise ilk 2 grup KBB uzmanı tarafından da tonsillektomi gereği açısından beraat etmiş çocukları, son kez, 3. defa, yeni bir 3 kişilik KBB uzmanına gösteriyorlar. Sonuç fevkalade ilginç. Bu sonuncu uzman grubu da, şaşmaz şekilde, çocukların 1/3'ünün bademcikleri alınsın diyor. Durum gayet açık. Bir çocuğun tonsillektomi gereksinimi, her şeyden önce, onu gören KBB

uzmanı sayısının bir fonksiyonu! Bu çalışmadan sonradır ki, doğal olarak tüm uygar dünyada 'bademcik alma' modası giderek azalıyor. Bu da tabii eski kanıta dayalı tıbbın bir diğer zaferi.

Şimdi biraz da 1960'ların başında yeni kanıta dayalı tıba neden gereksinim duyulmuş, onu tartışalım. Peki, neden 1960'larda 70'lerde kanıta dayalı tıp yapmak isteyen insanlar ortaya çıkıyor? Çok ilginç, yeni bir kitap var, 'Philosophy of Evidence Based Medicine'. Bu kitapta yeni kanıta dayalı tıbbın kurucularından ünlü Chalmers çok çarpıcı bir şekilde anlatıyor (2). Yazar, hekimliğinin ilk yıllarında İsrail-Gazze şeridinde çalışırken çok sayıda kızamıklı çocuğa bakıyor. Hastalara, Londra'da School of Tropical Medicine'da öğrendiği gibi, koruyucu amaçla antibiyotik kesinlikle yazmıyor. Çalıştığı sağlık merkezinde bir de Filistinli hekim arkadaşı var. O ise her gelen kızamıklı çocuğa antibiyotik veriyor. Bu Chalmers'a Londra'da öğretilenin tam aksisi. Ancak çarpıcı olarak Filistinli hekim arkadaşının baktığı çocukların prognozu Chalmers'ın baktığı çocuklara kıyasla çok daha iyi. Yazar çok haklı olarak bundan kendisine Londra'da, en ileri tıp merkezlerinde öğretilen tıbbın, Filistin'de geçersiz olduğu sonucunu çıkartıyor ve kendi kendine şu önemli kararı veriyor: "Bu gözlemden sonra artık bana her öğretileni sorgulamaya başladım." diyor.

Biraz da kendimden bahsetmek istiyorum. Ben 1969'da tıp fakültesinden mezun oldum. 1966'da patoloji okurken ünlü ve gerçekten de çok iyi bir patoloji hocamız vardı, Prof. Talia Bali Aykan. İltihabın nasıl oluştuğunu anlatırken, "İltihabı başlatan esas olay kan hücrelerinin damar duvarından dışarı çıkmasıdır." derdi. Peki, Talia hocaya göre hücreler bazı durumlarda neden damar dışına çıkıyordu? Bunun nedeni açıktı. Damar duvarında bir 'locus minoris resistentia' yani 'direncin en az olduğu bir bölge' oluşuyor ve kan hücreleri de buradan dışarıya çıkıyorlar.

İltihap olayının başlangıcını açıklamakta 'locus minoris resistentia' açıklaması, adeta kutsal bir din kitabından alınmış gibiydi. Doğal olarak, tartışılabilir, çürütülebilir bir kavram değildi ancak bu açıklama, tıp bilimi ve uygulamasında

birçok diğer açıklama gibi, Latince söylediğinde adeta bilimsel bir nitelik alıyor, sınavlarda da sorulabiliyordu.

Daha sonra, internlük yıllarımda Philadelphia'da Schamberg kremiyle tanıştım. Dr. Schamberg çok ünlü, Philadelphia'lı bir dermatologtu. Kendi adıyla anılan, bacaklarda görülen Schamberg ülserlerini tanımlamıştı. Söz konusu ülserlerin bir numaralı ilacı da doğal olarak (!) Schamberg kremiydi ve biz internlerin önemli görevlerinden biri de Schamberg kremi hazırlamaktı. Hastane eczanesinden steril bir kap içinde alelade beyaz bir krem gelir, hastadan ülserin çapına göre 10-20 cc alınan venöz kan bu kremle bir güzel karıştırılır ve ortaya çıkan pembe macun büyük bir bilgiçlikle ülserin üzerine, kalın bir tabaka halinde, sürülürdü. Kremin kerametini sorgulayan olmadığını gayet iyi hatırlıyorum.

İlerleyen yıllarda, romatoloji yan dal uzmanlığı eğitimine başladıktan sonra, konumuzla ilgili ilginç bir anım var. Çalıştığım hastanede genellikle büyük bir grup halinde hasta vizitleri yapan nefrologlar beni ve hocamı gördükleri her yerde "Siz romatologlar, hekimliğin elitisiniz" diye laf atarlardı. Anlaşılır gibi değildi. Ortada bilimsel açıdan elit diye nitelenebilecek bir grup varsa, nefrologlar kesinlikle bizden daha elitler. Gün boyunca 'anyon açığı' hesaplamak, hastanın asit baz dengesi için devamlı Henderson-Hasselbalch denkleminde baş vurmak hep onların işiydi. Hâlbuki aynı yıllarda, biz romatologların en popüler ölçü aletimiz ise grip strength (belirli bir dereceye kadar şişirilmiş bir tansiyon aleti) aracılığıyla, hasta elinin manşonu sıkma gücünü ölçmektir.

Sonra yavaş yavaş anlamaya başladım ki aslında nefrologların elitismden anladıkları onların aritmetik formüllere bağlı, mekanistik düşünce tarzları. Hâlbuki onlar bizi, grip strength ölçmemizden dolayı elitlikle itham etmiyorlar (!) esas kastettikleri duyarlılık, özgüllük gibi bazı temel ölçü kavramlarına bizim güncel tıp uygulamasında onlardan çok daha fazla yer vermemiz. Unutmayalım, bu toplantıda sabahdan beri dilimizden düşürmediğimiz duyarlılık kavramı, bir zamanların

PubMed'i Index Medicus'a ancak 1985'te girmiştir. O zamana kadar sensitivite deyince sadece allergologların sensitivitesi anlaşılırdı. Daha da ilginç, duyarlılığın etrafına bir güven aralığı (confidence limit) koymak ancak 1995'te ünlü Alvan Feinstein ve ark. tarafından yapılmıştır (3). Diğer bir deyişle 9/10 ile 900/1000 oranlarının her ikisi de %90 duyarlılık gösterirler. Ancak 10 hastanın 9'unda bir testin pozitif bulunmasıyla 1000 hastanın 900'ünde pozitif bulunması arasında dağlar kadar fark vardır.

İşte güven aralığı bize bu farkı işaret eder. Maalesef günümüzde dahi herhangi bir gözlemin duyarlılığından veya özgüllüğünden söz eden yazarların büyük bir kısmı bu saptamalarına birer güven aralığı koymak disiplinine girmemişlerdir.

Evidence Based Medicine Group'un JAMA 2002 yazısının (4) Yeni Kanıtı Dayalı Tıbbın başlangıcı olduğu kabul ediliyor. Nesnel Tıp toplantıları-mıza daha evvel katılmış olan Peter Tugwell de bu makalede çok önemli bir isim. Önce bu ünlü makalenin başlığına dikkat etmek gerek, "Kanıtı Dayalı Tıp: Tıp uygulamasını öğretmeye yeni bir yaklaşım." Yani yazının hedef aldığı kitle eğitici-ler ve doğal olarak öğrenciler. Bunu unutmamak gerek. Yazı ne diyor:

1. Hepsinden evvel yerleşik bilgiyi her zaman sorgulayın. Size doğru diye öğretilen her şeyin kanıtını arayın.
2. İkinci önemli nokta, hastalıkların mekanizmalarını bilmek o hastalıklara çare bulmak için yeterli değil. Hatta sadece buna dayanarak tedaviyi ayarlamak tehlikeli bile olabilir.
3. Buna adeta koşut, temel bilimlerdeki bilgi birikiminin güncel uygulamasını her zaman sorgulayın.
4. Hekimlerin, aritmetik, mantık bilmeleri ve çok okumaları gereklidir. Tıp literatürü ancak böyle değerlendirilebilir.

Sana yeni bir hasta başvurdu. Önce hastayı dinle, muayene et, sonra da hemen kütüphaneye koş. Biliyorsunuz günümüzde kütüphaneye cep aletinizden de ulaşabiliyorsunuz! Görüyorsunuz, burada önemli olan, demin bademcik ameliyatları

örneğinde altını çizdiğimiz gibi, geleneksel ve hocanın dediğini devamlı sorgulamak gerek.

Yakın bir geçmişte anladık ki tıp araştırma ve uygulamasında günümüze dek çok başarılı olmuş biyomedikal model, bir dizi hastalığı anlamamızda yetersiz kalıyor. Yani pnömokok pnömoni yapar, tüberküloz basili tüberküloz yapar, kolunu buradan kırsan şu olur modeli, birçok hastalığımızı izah edemiyor. ‘Biyopsikososyal model’ esas itibarıyla multifaktöryel hastalıklarda işe yarıyor. Söz konusu modelde laboratuvar testleri, radyografi vb. ikinci planda kalıyor. Burada önemli olan hastanın öyküsü. Aynı şey prognozu saptamakta da geçerli. Romatologların uğraştıkları hastalıkların büyük bir kısmı biyopsikososyal modelle daha iyi açıklanabilir. Eşit derecede önemli olarak böyle hastalıkların tedavisinde doktorun yaptıkları yanında hasta-çevre etkileşimi de ön planda. Geçtiğimiz yıl toplantımıza katılan Halsted Holman, esasında eski veya yeni, bir kanıta dayalı tıp savunucusu değil ancak söz ettiği Riley sıçanlarının çok ilginç bir öyküsü var (5). Bu inbred sıçanlar hayatlarının belli bir evresinde ve ancak muayyen bir virüsle enfekte sıçanlarca emzirilirlerse kanser oluyorlar ve belli bir süre sonunda ölüyorlar. Hâlbuki aynı fareleri alıp iyi bir ortamda tutarsan, kafesleri iyi ışıklandırılmış ve temiz olursa ve iyi beslenirlerse bu fareler kötü koşullarda yaşayan ve kendilerinden habire kan alınan farelerden anlamlı olarak çok daha fazla yaşıyorlar. Bu modeli biyoloji ile izah etmek pek mümkün değil. Stres gibi kavramlar koymak lazım. Esasen bir sürü izah edemediğimiz olay var. Ve yine çok önemli, yeni kanıta dayalı tıpta bir merkezden kontrol altına almak teması var.

Romatolojik hastalıklara baktığımız zaman bu söylediklerim çok önemli. Neden? Bir sürü romatolojik hastalık, esasında bir yapıt. Yani biz ne dese o oluyor. Gut, kol kırığı, septik artrit gibi bazı durumlar dışında hepsi bizim yaptığımız hastalıklar. Bunu bir zaptı rapta almak lazım. Ağrı gibi, tanınması, tartışılması, ölçmesi zor olan bir olay, bizim hastalıkların en önemli semptomu ve kimi örnekte sonucu. Demin bahsettiğim biyomedikal modelin romatolojik hastalıkların izahına yetmiyor. Ted Pincus ve ark. geçen yıl önemli bir yazı

yayımladı (6). Malum, tanı ve tedavide anamnez, fizik muayene ve laboratuvar önemli. Ancak çeşitli hastalıklarda bunların önemi değişiyor. Örneğin şeker hastalığına tanı koyarken en önemli laboratuvar, hipertansiyon tanısında ise fizik muayene. Hâlbuki romatoid artrit tanısında önce anamnez, daha sonra da bir miktar fizik muayene önemli.

Romatolojide bir türlü kantitatif düşünceyi ön plana çıkaramıyoruz. Uzun zamandır hastalık tanı kriterleri hakkında düşünüyorum, bu işte ne kadar çuvalladığımızı burada biraz daha anlatmak istiyorum. Hiç iyi yapamıyoruz bu işi. Hepsinden evvel neden tanı kriterlerine gerek var? Gayet basit. Tanı koyup hastalarımıza yardım etmek, hem hastaya hem de aileye hastalığın doğasını anlatmak için tanı önemli, bu bir. Bunun yanında, gerek temel gerekse de klinik bilimlerde araştırmalarda hastalık tanı kriterleri çok önemli. Tanı kriterlerinin bir üçüncü önemli oldukları alan ise topluma, sağlık otoritelerine, politikacılara ve nihayet gerek hasta bakımı gerekse bilimsel araştırmalara kaynak desteği yapanlara hastalıkları anlatmak yükümlülüğü.

Tanı kriterleri yıllardan beri çok yanlış yapıyor diye düşünüyorum. Zaman alıyor, para alıyor, vakit kaybı oluyor. Tanı kriterleri hazırlarken sirkularitenden kaçmak diye bir kavram var, hayır bütün kriterler sirküler olarak yapılırlar. İkincisi, diagnostik kriterlerle klasifikasyon kriterleri ayrı yaparım, diyorlar. Değil şimdi, 300 yıl sonra da klasifikasyon kriterlerinden ayrı diagnostik kriterler pek yapılamayacak. Diagnostik kriteri aramak aslında devri daim makinesine benziyor, çünkü diagnoz sadece tek bir hastada klasifikasyon yapmaktır. Beyindeki mantık, frontal korteksin işlemesi, klasifikasyon üzerine. Eğer bunu tek hasta karşısında yaparsan adı diagnoz, tanı, oluyor. Başka bir sihir yok. Aritmetik tamamen aynı, duyarlılık, özgüllük, hastalık sıklığını bilmek hepsi aynı, klasifikasyonda da var, diagnozda da var. Onun için devamlı olarak ‘ben diagnostik kriter yapacağım,’ absürt bir laf. Klasifikasyon yapacaksın, onu diagnozda kullanacaksın.

Bunları ilk söyleyen ben değilim. James Fries diyor ki “Kriterler tanım üzerine sirküler olarak

Erken RA için ACR/EULAR 2010 kriterleri

- Avrupa ve Kuzey Amerika'dan 12'şer araştırmacı (%24'ü kadın)
- Kontrol grupları yok
- En az bir eklemi tutan iltihaplı artritli olan hastada, 1 yıl içinde metotreksat başlanmış olması romatoid artrit göstergesi olarak alınmış

3. Slayt

yapılması gereken şeylerdir ama şu anda da yaptıkları o günlük bilgiler ışığında etraftaki geleneksel mantık neyse bunu etrafa aktarmak için yapıldı" (7). Kriter yapmanın amacı bu. Sonra devam ediyor. "Tabii mükemmel bir dünyada klasifikasyon kriterleri diagnostik kriter de olur". Fries tanı koymak sadece ve sadece bir hekimin tek bir hasta karşısında yapabileceği iştir, diyor gayet açık olarak.

Hastalık kriterlerini ne denli yanlış hazırladığımız açısından biraz ünlü ACR/EULAR 2010 tanı kriterlerinden (8) bahsetmek istiyorum. Şimdi 3. slayta bakalım, ne yapıyorlar? Önce tanı kriterleri komitesi oluşturmak için bir grup hekim alıyorlar, 12 tane Amerika'dan, 12 tane Avrupa'dan. Hekimlerin %25'i kadın diyorlar. Anlaşılr gibi değil. Yani RA tanı ve tedavisinde kadın veya erkek hekim arasında önemli bir fark mı var? Diğer fahiş bir yanlış, bu kriter hazırlama çalışmasında kontrol grubu yok.

Diğer önemli bir nokta, araştırmacılar "biz bu kriterleri erken RA'ya tanı koymak veya diğer hekimleri artritli hastaları romatologlara referetmeleri için koymadık" diyorlar. Esas amaç hastaları klinik çalışmalara almakta kullanmak diyorlar. Bu mantıksız, hatta bir yerde etik dışı bir laf. Bakın Hollanda'da yapılan bir çalışmada (9) 2010 kriterlerine göre tanı konmuş hastaların, bir yıl sonra, %18'inde bu tanının yanlış olduğu ortaya çıktı ve bu hastalar arasında 6 kanserli de vardı. Dinleyiciler arasında ilaç firması temsilcileri de var, herhalde bir süre sonra lenfoma olacak hastalara anti-TNF vermek istemezler. Bütün bunlardan öte, yazarlar tartışmalarında, yazılarının bir kısıtlaması (limitation) olarak "bu kriterlerin özelliği günümüzde bilinen bilgiler temelinde hazırlanmış olmalarıdır" diyorlar. Oldukça abes bir laf. Zaten tüm kriterler o günlük bilgiler ışığında yapılır.

Dediklerim tabii sadece RA'de kriter yapmaya ait değil. Bir süre ACR/EULAR'ın vaskülit kriterleri komitesine ben de katıldım. Ancak anlaşılmadık, komiteden ayrıldım. Her toplantıda ileride, klasifikasyondan ayrı, tanı kriterleri de yapacağız diyorlardı.

Peki, niye oluyor bütün bunlar: Aritmetik bilmiyoruz gibime geliyor. Sadece biz bilmiyor değiliz. Çok güzel bir kitap var, 'Math on Trial' (10) diye, tavsiye ederim. Toplum ve hatta yargının aritmatikle ne denli uzak olduğunu çok güzel anlatıyor. Ünlü Dreyfus davasından son yılların sansasyonel Amanda Knox cinayetine kadar birçok örnek veriyor. Bu olayda İtalyan yargıçlarının DNA verilerine olan güvensizliklerini okumak çok ilginç. Belki daha da çarpıcı Amerikan Yüksek Mahkemesinin herhangi bir dava sürecindeki olasılıklar üzerinden savunulan delillere pek yüz vermemesi.

Şimdi size aritmetikle olan kötü ilişkimizin diğer bir örneğini anlatacağım. Tayvan'da yapılmış bir çalışma (11). Anti-TNF grubu ilaçları kullanan RA'lı hastalardan oluşan ulusal bir kohorttaki kanser riskini yaş ve cinsine uyarlanmış ulusal kanser riskiyle karşılaştırıyorlar. Buldukları sonuç inanılır gibi değil. Birinci yıl sonunda buldukları 60 kat artmış risk 8 yıl sonra tersine dönüyor ve RA'lılar arasında kanser riski azalmış gibi görünüyor. Yani romatoid artritli ol, kanserden kurtulursun gibi bir durum. Bu yazıyla ilgili dergiye bir mektup yazdık ancak yazarların yanıtı hiç tatmin edici değildi (12). Dediğimizi adeta anlamamışlardı. Hâlbuki olan oldukça açık gibi. Söz konusu kohortta belirgin bir sol sansürleme var ve bu sansürün en önemli ögesi de büyük ihtimalle kanser geçirmiş veya artmış riski olan RA'lılara anti-TNF yazılması ve dolayısıyla onların bu kohorta girememesi. Burada esas üzücü olan yazarlara bu kuşku belirtildikten sonra dahi konuya eğilmek gereğini duymamaları.

Her nedense anti-TNF'lerin yan etkilerini dile getirmeyi düzgün yapmıyoruz. Örneğin, anti-TNF ajanlarla malign melanom riskiyle ilgili JAMA'da çıkan yeni bir yazıda (13) aynen şöyle diyor: "Bu nedenle mutlak risk, binde 6 olarak, oldukça azdı." Mutlak riskin binde 6 olması demek 159 tane hastada bu ilaçları kullanırsan bir tanesinde malign melanom çıkar demek. Peki, bu olay küçümsenecek bir risk mi? Bakın ben senede 159 defa Kadıköy vapuruna biniyorum. Bu vapur 159 seferinden bir tanesinde batacak deseler, binmekte tereddüt eder miyim, etmez miyim?

Kanıtla dayalı tıptan bahsederken bir de, oldukça kötü bir huy olan, kendimizi kanıtlama çabamızdan bahsetmemiz gerek. Kısa bir süre önce anlaşıldı ki uzun yıllardır hurda, işe yaramaz, DNA diye bildiğimiz DNA'nın çok önemli işlevleri vardı (14). Bu doğal olarak çok önemli bir buluş ve benim yereceğim tabii ki bu buluş değil, onun dünyaya duyuruluşu. Bu önemli buluşun çeşitli öğeleriyle ilgili alt buluşlar aralarında Nature ve Science gibi çok ünlü dergilerinden bulunduğu 30 ayrı dergide yayımlandı. Bu bilimsellik açısından çok yanlış bir davranış. Sözümlü ettiğim üst buluşa ait alt buluşların hepsi aynı anda mı tamamlandı da bu çalışmaları gerçekleştiren araştırmacılar bu bulguları aynı anda yayımladılar. Elindeki verileri yıllarca gizli tutup da toplumda en etkili olacak şekilde yayımlamaya kalkışmak bilimsel açıdan en azından ayıptır ve reklamcılığa girer.

Kendini kanıtlama yolunda diğer bir ayıp, son zamanlarda adeta moda olan, $p=1.0$ lafı. New England Journal of Medicine örneği en ünlü dergilerimizde bir çalışmanın sayısal sonucu olarak A grubunda, diyelim 28.7, B grubunda da 28.8 gibi iki ortalama veriliyor. Ondan sonra da yazarlar bunun istatistik olarak anlamsız olduğunu belirtmek için $p=1.0$ eklemesi yapılıyor. Açık olarak altını çizeyim. Bu esasında insan aklına saygısızlık ve tümüyle göz boyamak için yapılıyor.

Söz ettiğimiz kendini kanıtlama çabasının son zamanlarda gözlediğimiz bir başka ve ilginç boyutu daha var. Bilirsiniz bilimsel makale yazarken çalışmanızın kısıtlamalarını da yazınızda açık ve seçik olarak belirtmeniz bilimselliğin bir vazgeçilmezidir. Bir çalışmanın gerek iç, gerekse dış tutarlılık (internal and external validity) sorunlarını o çalışmayı yapanlardan daha iyi kimse bilemez. Yıllardan beri gözlemim temel bilimler makalelerinde öz eleştiri, klinik bilimlerden daha az. Bu varsayımı sınamak için bir çalışma planladık. Bu amaçla üç temel romatoloji dergimize belirli zaman dilimlerinde yayımlanmış özgün makaleleri aldık. Aynı makalelerin tartışma bölümlerinde; gerek belirli anahtar sözcük (limit veya ondan türeyen sözcükler) kullanarak elektronik, gerekse geleneksel okuma yöntemiyle iki bağımsız araştırmacı öz eleştiri aradı. Özetle gözledik ki bilim

çalışmaları ve klinik bilim makaleleri kabaca %75 öz eleştiri yaparken aynı oran temel bilim makalelerinde en fazla %75 (15).

Son olarak kanıta dayalı tıp veya yeni kanıta dayalı tıp hakkındaki güncel tartışmaya değinmek istiyorum. 2011 yılında S. Hickey ve H. Roberts tarafından yazılmış bir kitap (16) kanıta dayalı tıbbi çok eleştiriyor. En büyük yanlış da, deniyor ki, grup için iyi olan bir şey her zaman birey için de iyi olmaz. Hâlbuki kanıta dayalı tıp bunun tam aksini söylüyor. Diğer önemli bir eleştiri nedeni ise yeni kanıta dayalı tıp araştırmaya gerekli ağırlık vermiyor, tüm dedikleri yerleşik bilgiyle ilgili.

Bu eleştirilere çok katılmıyorum. Önce ikinci eleştiriyi yanıt vereyim. Doğru, kanıta dayalı tıp araştırma ağırlıklı değil, ancak araştırma yapmanın da demiyor. Esasında problem bizde gibi, kanıta dayalı tıpta hiçbir sorun yok. Bunun yanında birinci eleştiri de haksız. Grupta gözlenenin bireye ne kadar uygulanabileceğine karar vermek için bize akıl bahşedilmiş. Bence kanıta dayalı tıbbi inkâr etmek Lavosier'i inkâr etmek gibi bir şey.

Girişte altını çizdim. Yeni kanıta dayalı tıbbın ana amacı 'öğretmek'. Hâlbuki biz kanıta dayalı tıbbın öğretici boyutunu hemen tümüyle dışlayıp uygulamaya koyulmuşuz. Bunun yanında aritmetiğimizizin kötü olduğunu bir türlü kabul edemiyoruz. Üçüncüsü, topluma, hekim olarak, olasılıklarla çalıştığımızı söylemıyoruz: Dilimin döndüğü kadar açıklamaya çalıştım. Yukarıda değindiğim ECR/EULAR erken RA kriterlerindeki temel sorun bu. Topluma, sağlık otoritelerine dönüp bir hastada 1-2 aydan daha az süreli artrit varsa bu hastaya RA tanısı ancak belirli olasılıklarla konur. Yine aynı hastaya, kesin tanı koymadan metotreksat başlayabiliriz. Bütün bunları açık seçik gerek hastalarımıza gerekse topluma açıklamamız, öğretmemiz gerek. Kanıt dayalı tıp ancak bu açıklıklarla işler.

İzzet Fresko: Hocam çok güzel bir konuşmadı, teşekkür ederiz. Ben şimdi soracağım sorunun cevabını bilmiyorum aslında fakat kendi aramızda zaman zaman tartışıyoruz. Bunların hepsi çok

güzel ama gerçekten sadece iyi aritmetik ve matematik bilmemek mi problem? Devamlı hikâye kendi kendini yalanlamak ve bildiğinin yanlış olduğunu görmekten geçiyor. Sürekli aynı yere takılıp kalıyoruz. Acaba yöntemimizde sorun yok mu bizim? Akademik tıpta olsun, bütün bilim dünyasında olsun devamlı bölüp bölüp bir takım şeyleri yeniden tanımlamaya çalışıyoruz. Bunun acaba bizim farkında olmadığımız bir sınırı ve sorunu olabilir mi? Tıp eğitimi yaparken devamlı bu zeminde mi gitmemiz lazım, yoksa kendimizi biraz daha geri çekip bazen ne olup bittiğini tepeden değerlendirmemiz daha doğru olmaz mı ya da daha az nefes tüketici olmaz mı?

Hasan Yazıcı: İzzet bunun cevabını verseydim ben din kurardım. Bunun cevabı merkezi bir otoriteye bağlanıp onun her dediğini dinlemek değil. Bak ECR/EULAR kriterlerine. Hazırlayanlar arasında çok iyi bilim insanları var. Ancak çok yanlış bir şey yaptılar. Bu kriterler ortaya çıktığında, kriterlerin yazarlarından iki tanesi editoryal yazdı. Bu kriterlerin ne kadar iyi olduğunu yazdılar. Derginin bilmem kaçınıcı sayfasında kendi yazarı olduğu yazı var, öbür tarafta kendi yazdığı ne kadar iyi diye editoryal var. Böyle bilimsellik olur mu? Gençlere bunu mu öğreteceğiz?

Ahmet Gül: Hocam sadece eksiklik aritmetik bilip bilmemekte mi? Dil edebiyat bilip bilmemekte mi? Bu işi sadece matematiğe bağlamamak lazım. Onu da şu nedenle söylüyorum, aslında eğer biz yapılan tüm çalışmaları, bilimsel verileri kendi sınırları içerisinde, objektif olarak anlatır ve o anlatılanı doğru algıyorsak aslında ortada kavga edecek tartışacak bir şey de yok. Çünkü her veri kendi yapıldığı şartlar içerisinde geçerliliği olan bir veri, ama bizler bunu okumayı beceremiyoruz. Bence asıl sorun burada. Sadece sonuç kısmından okuyarak verinin detaylarını bilmeden, okuryazarlık eksikliğinden dolayı o sınırları içinde algılamayı düşünemiyoruz. Bir yazı çıktığı zaman, tamam bu yazı çıktı ama bu yazı bu kadar sınırlıdır, bu İncil'den, Kuran'dan gelme bir şey değildir. O limiti okumak için aritmetiğin yanı sıra bence biraz daha yoğun dil eğitimi ve edebiyata da yer bırakmak lazım.

Yasemin Kabasakal: Hocam çok teşekkür ediyorum çok yararlandım. Sonuçta, sınıflama kriterleri, tanı kriterleri yapmak gerçekten zor ama burada esas görüş, altın standart ne? Burada herhalde altın standart uzman görüşü. Bir çalışma yaptığınız zaman hasta grubunu toplamaya kalktığınız zaman, merkezlerden hasta istediğiniz zaman, size gönderilen hastalar, oradaki uzmanların uzman görüşüyle gelen hastalar. Kriterlere koyduğunuz zaman, hepsi tam kriterlere girmiyor, ama onlar o merkezlerde o kriterlerle, tanıyla izleniyorlar. Sonuçta burada bir yazı yazarken ne diyoruz, uzman görüşüne göre diyoruz. Herhalde gene altın standart uzman görüşü.

Hasan Yazıcı: Ama önemli olan uzman görüşü derken, o uzman görüşünün güvenilirlik sınırında olduğu. Yani Hasan 100 tane tanı koyuyorsa mesela, bunlardan 5 tanesi yanlıştır. Hasta hekime geldiği zaman bunu bilmesi lazım. Topluma bizim bunu öğretmemiz lazım.

Yasemin Kabasakal: Yani burada altın standart ne oluyor size göre hocam. Sınıflama kriterleri, tanı kriterleri dediğimiz zaman bir hastaya tanı koyarken altın standart nedir?

Hasan Yazıcı: Onu demek istiyorum, eğer elinde erken romatoid artrit olan insana anlatıp, sana ben metotreksat vereceğim, büyük ihtimalle senin romatoid artrit olma şansın %80, ama başka bir şey de olabilirsin, yakından takip edeceğim seni, diyerek şu dürüstlüğü yaptığımız anda iş bitiyor. Ama yapmıyoruz, problem burada.

Bedirhan Üstün: Dünya Sağlık Örgütü ICD'nin yeniden gözden geçirilmesi ve yazılmasıyla ilgili dün de konuşmuştum. O yüzden bu konuşma, az önceki tartışma sırasında Nurullah'ın da söylediği gibi ilginç oldu. Sizin söylediğiniz 10 şeyden 8'ine katılıyorum ama 2 tane katılmadığım şey var. Önce katıldıklarımdan birkaç tane söyleyim. Bu söylemiş olduğunuz çığırkanlık yöntemi. Ben ona dün desibel yöntemi demiştim. Çok geçerli, günü kurtaran bir şey ama, mantık açısından ya da sınıflandırma/tanı açısından bizi yanılgılara sokan bir şey. Ben özünde psikiyatristim, psikiyatride de bu tür tanı kriterleri var. Biz onlara

BOGSAAT Bunch Of Guys Sitting Around A Table, yani bir masanın etrafına oturmuş bir takım uzmanların onaylı görüşleri tarzında görüyoruz. Bu psikiyatride 30-40 yıl içinde yarar getirdiği gibi zarar da getirdi. Aynı şey romatolojide de oluyor, sizde çok güzel örneklerini verdiniz. Katıldığım ikinci nokta da aritmetik konusunda. Mesela biz Dünya Sağlık Örgütü olarak hastalıkların uluslararası yükünü tanımlayan çok önemli bir meta-sentez tarzında çalışma yaptık. Bunu Londra'da çok önemli bir üniversite tekrar gözden geçirdi. Ve söyledi ki hijyen düzeyi arttıkça gelişmiş ülkelerde Alzheimer sıklığı da artıyor, o zaman bu immünoloji ile alakalı bir şey. Alzheimer bir şekilde immünolojik düzenin bozulmasından ortaya çıkan bir şey. Tabii sizin verdiğiniz örnek gibi, yaz gelince boğulmalar artıyor, çünkü dondurmalarla bağlı tarzında bir açıklama. Ve bu da çok önemli dergilerde yayımlandı, basına gitti. Size katılmadığım noktayı son olarak söyleyeyim, o da tanı kriterleri ile sınıflama kriterlerini, bir şekilde, aynı değil diye algıladım. Onu açıkça söylüyorsunuz, bunları bir şekilde bağdaşık hale getirmek, birleşik hale getirmek de mümkün diye düşünüyoruz.

Hasan Yazıcı: Aynı şeyi söylüyoruz, yüzde yüz aynı şeyi söylüyoruz. Tanı kriteri esasında, hasta başında bir hastaya uygulanan sınıflandırma kriteridir.

Bedirhan Üstün: Orada çok önemli bir şey söylediniz, mantık. Yani asolan orada bir şekilde tanı mantığı ya da descriptif logic dediğimiz tarz. Bir bilgisayarın uygulayabileceği, aptal bir bilgisayarın uygulayabileceği, dolayısıyla herhangi bir akıllı insanın, bilgisayar mantığını uygular tarzda kopyalayabileceği bir mantık düzeyine getirmek lazım. Şu anda, bir hafta sonra psikiyatride önemli bir kongre olacak, orada konuşacağım. Mesela DSM ve ICD'de yapılan şeyler, bir şekilde, bu BOGSAAT dediğimiz çığırkanlık yöntemine geliyor ama altında çok önemli fizyolojik, genomik, fenotipik kriterlerle yeni bir bilim oluşabiliyor. Örneğin, eskiden bir sınıflama kriteri olarak şu örneği verip bitiriyorum, meme kanseri ile ovaryum kanserini iki ayrı kanser gibi görüyorduk, şimdi en azından bir kısım alt tiplerinin aynı kanserin iki

ayrı tezahürü olduğunu gösterebilen, sınıflandırma mantıklarına sahibiz. O açıdan belki bu tarzda yürürsek daha akılcı sınıflama kriterleri yapabiliriz. Teşekkür ederim.

Nurullah Akkoç: Hocam her zamanki gibi çok güzeldi ve sınıflama kriterleri konusunda siz %100 düşünebiliyorsunuz. Tanı kriterleri ve sınıflandırma kriterleri için. Yalnız bir tek şeyin altını çizmek lazım. Yasemin'in söylediği. Klasifikasyon kriterleri zaten klinik çalışmalarda o uzman görüşüne göre tanıyı almış hastaların klasifikasyon kriterlerine uyanlarını almak. Yani çalışma popülasyonunu homojen hale getirmek için yapıldığını hatırlatmak lazım.

Kaynaklar

1. Stolley PD, Lasky T. Health care services: What Works? In: Investigating Disease Patterns. Scientific American Library: New York, 1995, pp. 199-201.
2. Howick J. The Philosophy of Evidence-Based Medicine. Oxford: Wiley-Blackwell;2011.
3. Reid MC, Lachs MS et al. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA. 1995;274:645-51.
4. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992;268:2420-5.
5. Yazici H. Thought barriers to understanding rheumatic diseases--Halstead R. Holman revisited. Bull NYU Hosp Jt Dis. 2010;68:143-6.
6. Castrejón I, McCollum L, Durusu et al. Importance of patient history and physical examination in rheumatoid arthritis compared to other chronic diseases: Results of a physician survey. Arthritis Care Res 2012; 64:1250-5.
7. Funovits J, Aletaha D, Bykerk V, et al. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: methodological report phase I. Ann Rheum Dis. 2010; 69:1589-95.
8. Fries JF. Disease criteria for systemic lupus erythematosus. Arch Intern Med. 1984;144:252-3.
9. van der Linden MM, Knevel R, Huizinga TWJ et al. Classification of rheumatoid arthritis. Comparison of the 1987 America College of Rheumatology criteria and the 2010 American College of Rheumatology/American League Against Rheumatism criteria. Arthritis Rheum. 2011;63:37-42.
10. Schneps L, Colmez C. Math On Trial How Numbers Get Used and Abused In The Courtroom. Basic Books, 2013.
11. Chen YJ, Chang YT, Wang CB et al. The risk of cancer in patients with rheumatoid arthritis: a nationwide cohort study in Taiwan. Arthritis Rheum. 2011;63(2):352-8.
12. Yazici H, Tascilar K, Kiroglu G, et al. Mortality bias in reporting cancer incidence in disease registries: comment on the article by Chen et al. Arthritis Rheum. 2011;63:2543-4.
13. Lopez-Olivo MA, Tayar JH, Martinez-Lopez JA et al. Risk of malignancies in patients with rheumatoid arthritis treated with biologic therapy: a meta-analysis. JAMA 2012;308:898-908.
14. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57-74.
15. Yazici H, Gogus F, Esen F et al. There was less self-critique among basic than in clinical science articles in three rheumatology journals. J Clin Epidemiol. 2014;67:654-7.
16. Hickey S, Roberts H. Tarnished Gold: The Sickness of Evidence Based Medicine. CreateSpace;2011.

DÜNYA SAĞLIK ÖRGÜTÜ ÖZÜRLÜLÜK DEĞERLENDİRME PROGRAMI 2.0

T. Bedirhan Üstün

*Psikiyatri Bölümü
Dünya Sağlık Örgütü*

Dünya Sağlık Örgütü olarak, özürün nasıl ölçüle-
bileceğine ilişkin bir cetvel geliştirmeye çalıştık.
Bu anlamda aslında tümüyle burada cetvelde
simgesel olarak gösterdiğimiz gibi bir ölçüm aracı,
bir cetvel geliştirmek istedik. Sizin romatolojide
kullandığınız bu şekilde pek çok ölçek de var. HAQ
var, bir sürü değişik şeyler de var. Ve bunlar da
aynı şekilde santimetreyle, yani metrik sistemle
farklılıklar olabiliyor ama kimi cetveller bir şekilde
elastikten yapılmış oluyor, uzuyor. Kimi cetvelle-
rin bazı yerlerinde boşlukları var, söylemek iste-
diğim şey bu olacak. Şimdi hastaların resimlerini
göstermek Dünya Sağlık Örgütü'nde pek muteber
bir pratik değil maalesef, ama bunu Dünya Sağlık
Örgütü'nde bir önemli direktör gösterdi, şu anda
da Dünya Bankasını yönetiyor. Jim kim? Bu bir HIV
hastası. Mart 2003'teki resmi (1. slayt).

Aynı hastanın yaklaşık 8 ay sonraki resmi
(2. slayt). Siz romatolojide de bu tür drastik
iyileşmeler olan resimler gösterebilirsiniz. Ben
romatolog olmadığım için ona cesaret edemedim
ama bir açıdan bakacak olursanız, bu hastanın 1.
dönemdeki tanısı ICD'de konulan b24, hani bu-
nun CDC ve Dünya Sağlık Örgütü'nün devreleme,
stage'leme verileri de var ama tanı 8 ay sonra
gene aynı. İki tanıyı birbiriyle baktığınız vakit, bu
hasta iyileşti mi, iyileşmedi mi, ne oldu onu bil-
meniz mümkün değil. Ama baktığımızda bizim bir
ek başka sınıflamamız var, az sonra ondan kısaca
bahsedeceğim. Bu hasta bu haliyle yürüyemiyor,
kendine bakamıyor, okula gidemiyordu. Sekiz
ay sonra bütünüyle iyi olmuştu. Yalnızca yüksek
öğrenime girmesi açısından HIV'si olduğu için
bir engellenmesi söz konusuydu. Daha ayrıntıya
girmiyorum. Siz klinisyenlere, aslında burada
söyleyeceğim her şeyi daha önceden bildiğinizi
kabulle başlıyorum. Moliere'nin Kibarlık Buda-
lası'nda, sonradan görme bir zengin olan bir
kişi kendisine iki tane profesör tutar, matematik
profesörü, dans profesörü, müzik profesörü vs.
ve onlardan edebiyat profesörü der ki, "İki şekil-
de konuşulur, eğer kafiyeyle konuşursanız bu-
nun adı şiirdir, düz olarak konuşur ve yazarsanız
bunun adı nesirdir." Adam da der ki, "Aman Tan-
rım, 40 yıldır nesir konuşuyormuşum da bundan
haberim yoktu." Siz aslında ICF konuşuyorsunuz.
Ben tereciye tere satar gibi ICF anlatacağım,



1. Slayt



2. Slayt



**International
Classification
of
Functioning
and
Disability**



WORLD HEALTH
ORGANIZATION
GENEVA
2001

ICF

1. Ana Sınıflandırma
2. Klinik tanımlama ve değerlendirme kılavuzu
3. Araştırma için ölçütler
4. Özgül Değerlendirme Araçları (testler)

3. Slayt

ama ‘bunu nasıl ölçeriz’ kısmını anlatacağım. Bu bir şekilde böyle, ama bunun altında önemli bir kuram var. Nasıl ki edebiyat kuramı olduğu gibi, araçları var ve uygulaması var. Bu araçlardan bahsedeceğim. Yapmış olduğumuz sınıflandırma bu (3. Slayt).

Bir şekilde bir ana sınıflandırma var, onun klinik tanımlama değerlendirme kılavuzu var, tanı kriterleri gibi. Ölçütlerimiz var ve onun ötesinde de özgül değerlendirme açısından bir takım araçlarımız ve testlerimiz var. Burada önemli bir ayrıma da değineyim, sizin kafanızda baştan söyleyim, hastalık veya bozukluk dediğimiz şeyi biz IC'lere sınıflandırıyoruz. Sizin romatoid artrit dediğiniz gibi, artı hastalığın veya bozukluğun doğurguları var. Bu durumda olan hasta ne yapamıyor? Gerek bedensel olarak, gerek kişi olarak, gerek toplum içerisinde bir birey olarak. Bir de yaşam kalitesi dediğimiz şeyi aslında bunu John F. Kennedy sokmuş bizim bilincimize, 1960'larda. O insanın öznel olarak yaşamdan aldığı doyum ve ona ilişkin bir düşünce. Onu farklı tutuyoruz. Yani kişinin öznel olarak duyguları onunla ilgili, doyum farklı olarak alıyoruz.

Eğer bakacak olursak yetiyitimi dediğimiz, özür dediğimiz şey ya da disabled dediğimiz şey İngilizce'de kabaca, insanın sağ beyinle düşünenler için söylüyorum bunu, bu ise, bunun altında değişik şeyler de olabilir. İlla tekerlekli sandalye olması gerekmiyor, mental, işitsel ya da başka tür eksiklikler olabilir. Bunun altında yatan duruma, biz bir şekilde sakatlık ya da impairment diyoruz. O da bu kişinin romatoid artrit olması ya da ADHD olması olabilir. Bundan dolayı toplum içerisinde bir takım şeyleri yapamıyorsa, ehliyet alamıyorsa, okulda özel eğitime gerek duyuyorsa bu da 3. şey ‘engellenme’ diyoruz. Ve bunlar Türkçe karşılıkları her ne kadar çok öz Türkçe olduğu için anlamakta zorluk çekiyor olabilirsiniz ama bir kişinin bedninde ne var, bu bir kişi ne yapamıyor, bu da topluma neden katılamıyor tarzında düşünebilirsiniz. Bunların anahtar kavramlarını bir şekilde beden, birey ve toplum olarak ele alıyoruz. Mesela benim gözüme travma gelir, bir gözümü kaybedersem bu açıdan bir sakatlığım var. Bundan dolayı evde iş yapmamda sınırlanmam olabilir, ama araba

Kavramsal Düzlemler

Hastalık - Doğurgular - Yaşam Kalitesi

Hastalık / bozukluk
Hastalığın / bozukluğun doğurguları
Yaşam Kalitesi

4. Slayt

Yetiyitimi

Anahtar Kavramlar

İşlev /Bütünlük	Etkinlik	Katılım
BEDEN İşlev / Yapı (yitim) şiddet, yer süre	BİREY Etkinlikler (sınırlanma) güçlük, süre yardım gereksinimi	TOPLUM Katılım (kısıtlama) kerte, yardım çevresel engeller

5. Slayt

Travma

İŞLEV / YAPI	ETKİNLİK	KATILIM
Görme yitimi	Evde iş yapma sınırlanması	Ehliyet almada kısıt koyma (engel)

6. Slayt

kullanabileceğim halde toplumsal düzenlemeler tek gözüm olduğu için ehliyet vermezse, buna da bir katılım engellenmesi deniyor. Bu açıdan da bunların hepsini bir süreklilik olarak görüyoruz. Mesela görme işlevlerini verdiğimde bunu 0'dan 100'e kadar bir spektrum olarak ele alabilirsiniz ve hafif orta görme kaybında gözlük, kontak lens-ten tutun, ağır görme kaybında ameliyat gereken durumlar, orta görme kaybında ise bir takım yardımlar olabilir.

Söylemek istediğim şey, burada bir tek görmeyi örnek verdim ama bu sadece vizyonla ilgili değil, aynı zamanda hastaların kendine bakımı, çevreyi anlaması, eğitim görmesi, toplum içerisinde bir takım etkinliklere katılması, hareketi, işitmesi, görmesi gibi, çok değişik boyutları değerlendirmemiz söz konusu olabilir. Olmalıdır da. Bunları üç boyutlu uzayda birer vektör olarak düşünürseniz (7. slayt), bu vektörlerin hepsinin bileşkesine bir anlamda biz işlevsellik ve işlevsellikteki kısıtlama ya da 'disability' diyoruz. Ve bu sadece kişinin sağlık durumuyla değil, aynı zamanda kişinin kaynakları ve yaşadığı çevreyle de etkili diyoruz ve özetle, beden işlevlerinde yitim, bireysel etkinliklerde sınırlanma ve toplumsal katılımda kısıtlanmaya 'engellenme' deniyor. Ve bunların böyle bir karmaşık modeli var, illa eskiden burada kişide bir romatoid artrit var, onun etkinliği dolayısıyla, etkinliklerde bir takım kısıtlamalar oluyor, o da toplumsal bir takım şeylere katılmıyor diye tek düze, tek yönlü ilişkiler vardı, şimdi romatoid artriti buraya koyuyoruz onun yapmış olduğu işlevleri, mesela eklemlerdeki kısıtlılık, ondan dolayı yürüyememe, ondan dolayı toplum işlemlerine katılamama, mesela camiye gidememe vs gibi şeylerden tutun da çevresel etkenler, kişisel etkenlere varana kadar değişik bir model alınıyor.

Şimdi bunu neden söyledim, arka planda bunu düşünerek, neyi ölçeceğimizi bilmemiz gerek. Bu ölçümü yapmak için, dünyada yaklaşık 350-400 kadar var olan ölçme aracını alıp bir bakıma bir veri tabanına tümüyle koyduk. Elimizde yaklaşık 9000 tane soru oluştu. Her sorunun neyi ölçtüğünü, ne kadar sensitivitesi, spesifitesi olduğunu bildiren bir takım verileri de, yanına



7. Slayt

kanıtı da koyduk. Ondan sonra gördük ki pek çok şey, aslında bunların içinde öne çıkarlar tabii SF-36, London handicap scale, HAQ gibi şeyler olabiliyor ama bunlar genel ölçüm araçları. Yani bütün hastalıklara aynı anda uygulanabilen ölçüm araçları. Veya özgül ölçüm araçları. İşte film gibi, belli alanlarda öncelik gösteren diye sınıflandırılabilir. Onlardan bakarak bir şekilde kanıta dayalı bir analiz değil ama human fracture analyses, insan zihniyle yapılan commonsensical factor analysis yaptıktan sonra da bunu 6 ana boyuta indirdik.

1. Cognition dediğimiz bir insanın çevresini anlaması ve iletişimde bulunması.
2. Hareketlilik. Yani insanın bir ortamda hareket etmesi.
3. Kendine bakması. Hijyenine, giyinmesine, yemesine-içmesine ve yalnız kalabilmesi.
4. Başka insanlarla ilişkileri (interpersonal relationships).
5. Hayat içerisindeki etkinlikleri. Ev işlerini yapması, boş zamanlarını değerlendirmesi. İşe gitmesi, okula gitmesi.

ve

6. Boyut da 'participation' dediğimiz toplum içinde etkinliklere katılabilmesi.

Bütün bunları yaptığımız vakit, bütünlüklü, insanların tüm işlevselliğini kapsayan edimleri, etkinlikleri bir şekilde, uzaya oturtmuş olduğumuzu düşündük. Ve buradan çıkarak, WHODAS geliştirdik (8. slayt). Sonuçta bunun, kişilerin bu alanda, 6 boyutta özetlenebilecek bir yetiştirimi profili veren ve özet olarak da az önce göstermiş olduğum özet vektörde gösterildiği gibi bir 'summary measure' dediğimiz bir özet ölçüt olması lazım getirdiğini gördük. Ve bunun geçerliliğini bir şekilde istatistiksel yöntemlerle göstermekten başka değişik uygulamacıların elinde aynı sonucu veren reliable yani güvenilir bir ölçek olduğunu gösterdik. En önemli şeylerinden bir tanesi de bunun gerek Türkiye'de gerek Hindistan'da gerek Amerika'da aynı şekilde uygulanabilir ve bütün hastalıklarda, gerek psikiyatrik, gerek romatolojik hastalıklarda, gerekse bildiğimiz diğer cerrahi

WHODAS 2.0 nedir?

- Sağlık ve özürllülük ölçümünün genel değerlendirilmesi
 - klinikte
 - toplumda
- hayatın 6 alanındaki işlevsellik düzeyinin ele alınması
 - 1: **Kognisyon** anlama ve iletişim kurma
 - 2: **Hareketlilik** hareket ediyor olma, gezinme
 - 3: **Özbakım** hijyen, giyinme, yeme, yalnız kalabilme
 - 4: **Geçinmek** diğer insanlarla ilişki kurma
 - 5: **Yaşamsal aktiviteler** ev içi sorumluluklar, boş zamanları değerlendirmek, iş ve okul
 - 6: **Katılım** toplu aktivitelere katılım, toplumun bir parçası olma
- özürllülük profilini ve özet ölçütü veriyor.
 - güvenilir
 - farklı kültürler ve yetişkin insan topluluklarında aynı şekilde uygulanabilir

8. Slayt

hastalıklarda benzer sonuçlar verdiğini gördük. Size onun özetini anlatacağım kısaca ilerleyen dakikalarda.

Ve bu az önce göstermiş olduğum şekilde ICF'ye bağlı, o yüzden ICF'yi size önden tanıttım. Ve çok önemli bir şekilde bütün değişik ülkelerde aynı sonuçları veren iyi psikometrik özellikleri olan ve kolay kullanılabilen bir ölçek olmasına çalıştık. Az önce söylemiş olduğum gibi bizim ölçmek istediğimiz şey hastalığın tanısı, yani hastalık tanı kriterleri ya da sınıflandırma kriterleri neyi kullanırsanız kullanın, biz onu IC'lerde sınıflandırıyoruz. ICF'ye geldiğimizde az önce söylediğim gibi, bu kişinin beden düzeyindeki etkinliklerini WHODAS (9. slayt) ölçmüyor. Yani tam anlamıyla onun mesela romatoid artritte olan değişik T hücre özellikleri veya eklem kısıtlılığını orada ölçmüyor, ölçtüğü iki şey var. Kişi düzeyinde aktiviteleri, yani yürüyebiliyor mu, evden çıktığında yardımsız dışarıda dolaşabiliyor mu vs gibi. Ve sosyal düzeyde bunları ne yapabiliyor. Bunları ölçüyor. Ve ayrıca bu kişinin subjective well being dediğimiz yani öznel doyumunu bir şekilde ölçmüyor. Onunla bağlantılı, onunla paralelizmi var ama ondan değişik bir şey olduğunu vurgulamak isterim. Onu da biz ayrı bir ölçeğimiz var WHOQOL diye WHO'nun quality of life diye scalası. Onunla ölçmeye çalışıyoruz. Onunla da aynı domainlerde bir işi yapabiliyor musun, yapıyorsan bundan yeterince memnun musun tarzı değerlendirmek mümkün paralel olarak.

Bunlar bizim bu işi yaparken 1999'da başladı, 2010'da bitti, çalıştığımız takım ve bu da dünya yüzeyinde hangi alanlarda nerelerde çalışıldığı (10. slayt). Türkiye'de Ankara Hacettepe Üniversitesinde denendi bu. Ama onun dışında gördüğünüz gibi pek çok yerde gelişimi var, sadece hani bunun yaygınlığını göstermek açısından söyledim. Ve bunların hepsi bir şekilde WHODAS gelişip, onu tercüme edip uygulamış kişiler değil, tam tersine biz başından beri bu görmüş olduğunuz 23-24 merkezle birlikte yaptık. Söylemiş olduğum gibi 300'den fazla araştırma ölçeğini bir araya getirdik. Bunları ICF ve onun tanımlamalarına paralel hale getirdik ve bunun içerisinde şöyle bir sistem geliştirdik. Mesela bazı kişiler çok kısa,

WHODAS 2.0'ı neden kullanalım?



- Kavramsal olarak doğrudan ICF'e (International Classification of Functioning, Disability and Health) bağlıdır.
- Kültürler arası kıyaslanabilirlik
- Üstün psikometrik özellikler
- Kolay erişebilirlik ve kullanım

9. Slayt

WHODAS 2.0 Geliştirme TAKİMİ			
WHO DASA MERKEZLERİ		İŞBİRLİĞİ TAKİM ÜYELERİ:	
T. Kugener	<i>Avusturya</i>	M. von Korff	<i>ABD</i>
K. Hourn	<i>Kambocya</i>	C. Pulli	<i>Lüksemburg</i>
G. Yao	<i>Çin</i>	E. Badley	<i>Kanada</i>
J. Saliz	<i>Küba</i>	K. Ritchie	<i>Fransa</i>
V. Mayreas	<i>Yunanistan</i>	D. Wiersma	<i>Hollanda</i>
S. Multry	<i>Hindistan</i>	M. Prince	<i>Birleşik Krallık</i>
H. Pal	<i>Hindistan</i>	R. Kessler	<i>ABD</i>
R. Thara	<i>Hindistan</i>	R. Trotter	<i>ABD</i>
U. Nocentini	<i>İtalya</i>		
M. Tazaki	<i>Japonya</i>		
E. Karam	<i>Lübnan</i>		
C. Pulli	<i>Lüksemburg</i>		
H. Hoek	<i>Hollanda</i>		
B. Odejide	<i>Nijerya</i>		
J. Segura Garcia	<i>Peru</i>		
R. Vrástl	<i>Romanya</i>		
D. Veltishev	<i>Rusya</i>		
J.-L. Vazquez-Barquero	<i>İspanya</i>		
N. Glozier	<i>Birleşik Krallık</i>		
P. Doyle	<i>ABD</i>		
D. Haisin	<i>ABD</i>		

10. Slayt

epidemiyolojik araştırmalarda kullanmak için kısa bir ölçek istiyorlar. Ama klinisyenler genel raporlarında kullanmak için çok ayrıntılı bir sistem istiyorlar. O yüzden, kısa, orta ve uzun formlarını geliştirdik. Ve görmüş olduğumuz başka bir şey de, önemli çalışmalar yaptıktan sonra son bir aya mı, son bir yıla mı, yoksa tüm bir yaşam boyuna mı bakalım diye. Özellikle son bir aya baktık ama uzamış versiyonlarında bunu son bir yıla veya tüm yaşam boyuna da aktarmak mümkün.

Size, Nurullah'a vermek üzere manuelini getirdim, bunu WHO'nun sitesinden ücretsiz PDF indirebilirsiniz. Bunun içerisinde yapılandırılmış ve yarı yapılandırılmış şekilleri var ve kişilerin kendilerine uygulayabilecekleri, hastaya verip hastanın doldurabileceği, klinisyen olarak sizin kullanabileceğiniz ya da hastaların yakınlarına soarak yapabileceğiniz sistemler var. Ve bunun da bir skrolama sistemi var. Bu skrolama sistemin-den ortaya çıkabilecek bir takım algoritmalar söz konusu. İlk yapmış olduğumuz şey birinci evrede bunu geliştirirken, cross-cultural applicability research idi. Bunu şu şekilde bir kitapta yaptık. Yani yetiyitimi dediğimiz şeyi değişik ülkelerde değişik dillerde nasıl ele alıyoruz. Bu konudaki normlar, değerler, inanç sistemleri neler? Değişik ülkelerde yetiyitimi nasıl sınıflandırıyoruz. Ve gelişmiş ülkelerde örneğin bir hastaya giyinme zorluğunuz var mı dediğinizde Hindistan'da saree giyen bir kadınla Amerika'da tişört giyen bir kadının aynı soruya aynı cevabı vermesini beklemek anlamsız olur. Aynı şekilde Nurullah'la konuştuğumuz gibi Türkiye'de bazı hastaların özel bir takım istekleri var onları nasıl koyabiliriz. Yani sıra bu nasıl tercüme edilebilir veya dile getirilebilir, nasıl kullanılabilir ve bu az önce sözünü ettiğim gibi Hindistan'daki kadınla, Amerika'daki kadının giyinme konusundaki zorlukları nasıl eşlenik hale getirilebilir diye. Ve yapmış olduğumuz 2. fazda da bu 300 item'i, 6 domain'e ve 89 item'lik bir uzun forma getirdik. Burada classical test theory ve item response test theory denen iki tane eğitimcilerin ve psychometrician'ların kullanmış olduğu yöntemleri kullandık. En önemli şeylerden klasik test teori, bir sorunun bütün diğer sorularla olan iç tutarlılığına bakıyor. Ama modern item response theory, mesela Türkiye'de matematikten

on üzerinden dokuz alan bir çocuk lise çağında, Amerika’da A, B, C, D diye değerlendirilen bir sistemde A mı alır, B mi alır veya Japonya’da 20 üzerinden değerlendirilen bir sistemde ne alır şeklinde eğitimcilerin geliştirmiş olduğu sample independent bir takım değerlendirmeler var. Onlara baktık. Ve görmüş olduğumuz şey kültürler arasında stabil olan itemleri seçtik ve bunlardan hangi ölçek soruları daha sensitif ve daha spesifik onu seçtik ve sorduğumuz şeylerde insanlara, bunu yapabiliyor musun dediğimizde iki şey ortaya çıktı. Bir tanesi ne kadar zorluk yaşadığı. Yani son bir ay içerisinde bunu yaparken ne kadar zorluk yaşadığı. İkincisi de 30 günün içerisinde kaç gün bunun olduğu. Bunu ikiye böldük. Bir tanesi difficulty dediğimiz zorluk derecesi. İkincisi zaman ölçeği şeklinde. Üçüncü fazda da bu işin yeterli, güvenilirliğine baktık. Dördüncü fazda da size söyleyeceğim health services research dedimiz, yani bunu ölçmek bize sağlık sistemlerinde ne işe yarıyor, ona baktık.

Şu anda bizim kullanıma sunmuş olduğumuz 36 item’lik uzun versiyon var. Bir de 12 item’lik kısa versiyonumuz var. Hibrit versiyon dediğimiz 12 item’lik başlayıp pozitif bulduğunuz şeyleri 24 diye ekleyerek 36’ya götürebileceğiniz versiyon da var. 12 item’lik versiyonun Türkçe çevirisini Hacettepe Üniversitesi yaptı. Oradan edinebilirsiniz. Uzun sürelerle, 30 dakika kadar ayakta durmakta ne kadar zorluk yaşıyorsunuz son bir ay içerisinde? Ev işlerini yerine getirmekte ne kadar zorluk yaşıyorsunuz? Yeni bir işi öğrenmekte, özellikle yeni bir yeri, adresi öğrenmekte ne kadar zorluk çekiyorsunuz? Toplumdaki bir takım etkinliklere katılmakta mesela, dinsel ya da başka etkinliklere katılmakta, kutlamalara katılmakta ne kadar zorluk çekiyorsunuz? Ve kendi sağlık sorunlarınızdan dolayı, ne kadar duygusal yönden etkilendiniz? İlk beş soruya “hayır” diye cevap veren bir kişinin bundan sonraki sorulara “evet” deme şansı %5’den az, o yüzden burada kesebilirsiniz yani %5 yanılma payını düşünerek burada kesebilirsiniz. Bunun dışında konsantrasyon, yürüme, yıkanma, giyinme, diğer insanlarla geçinme, arkadaşlıklarını sürdürme ve gündelik işlerini yapma soruları var. Bu da 12’lik sorular halinde geliyor. Bu ilk baştaki 5 soru, söylemiş olduğum gibi %88

varyasyonu var ve genel yetiyitiminin %90’ından fazlasını yapıyor ve ondan sonraki herhangi bir domainde, 6 domain olduğunu söylemişim, %90’dan fazla sensitivitesi var. Ondan sonraki 7 item ise, bizim her alandaki 6 parametrede özel skorlar almamızı sağlıyor. Bu göstermiş olduğumuz principal component analysis sonuçları. 11. Slaytta gördüğünüz gibi her 6 parametrede yüksek derecede yüklemeler var ve onların alt sorularına bakarak, anlamlı derecede yüklemeler olduğunu görebiliyorsunuz. Ve diyelim ki bunun bir değerlendirmesini Ankara’da bir değerlendirmesini İzmir’de yapıyorsunuz aynı hastaların.

Sonuçta bunların intraclass correlation coefficient’ları yani şansa bağlı uyumun düzeltildiği coefficient’lar yaklaşık 0.9 düzeyinde, tabii değişik alanlarda daha düşük oluyor ama her biri %69’dan daha fazla ki o da önemli derecede güvenilir olduğunu gösteren bir sonuç. En önemli şeylerden bir tanesi, biz bunu yaklaşık bu ülkelerde 3000’den fazla klinik popülasyon hastası üzerinde test ettik ve bunu yaparken özellikle National Institute of Mental Health, National Institute on Alcohol Abuse and Alcoholism, National Cancer Institute’den fon aldığımız için bu alanlardaki hastalara bakmıştık. Değişik hastalara, mesela alkol kullanan hastalar veya drug kullanan hastalar, özellikle, participation society’de sorun oluyolar ama mesela fizik sorunu olan hastalarda ise, özellikle mobilite alanında daha yüksek skorlar gösteriyorlar. Yani bu aslında çok common-sensical, bilinenin gösterilebildiği bu aletin bizim istediğimiz şeyleri ölçebildiğini gösterecek basit bir test. Ama bunu başka şeylerle baktığımızda bu 6 parametrenin SF quality of life, London handicap scale ve filmle olan, concurrent valid denen eş anlı geçerlik skorlarına baktık. Ve gördüğünüz gibi, hemen hemen çoğu oldukça yüksek, her biri kendi alanında özellikle mobilite alanında, sizin açınızdan baktığınızda, diğer ölçme araçlarıyla yüksek derecede bir korelasyonu var. Düşük olanlar mesela quality of life’da çıkabiliyor çünkü onlar özellikle subjektif şeyleri ölçtüp, objektif şeyleri ölçmediği için (12. slayt). Boş olan yerlerde ise bu alanlarda bizim bilebildiğimiz kadarıyla bu ölçeklerin geçerli soruları yok karşılaştırılabilecek. O açıdan bir bakıma bu alanlardaki boşlukları

WHODAS 2.0

Geliştirme çalışmaları

- **Faz 1: Kültürlerarası Uygulanabilirlik Araştırması**
 - Etki alanı, öğeler (item), ifade
- **Faz 2: Araştırma (Soru) yapılandırması ve İtem indirgeme**
 - 300 item → ALTI BOYUT - 89 item (CTT ve IRT)
 - Kültürlerarası stabilite
 - Taramalarda bulma yeteneği (exploration on screens)
 - YOĞUNLUK (zorluk derecesi) ve SIKLIK (zaman ölçeği) araştırmaları
- **Faz 3: Güvenirlilik**
- **Faz 4: Sağlık Sistemi Araştırmaları**

11a. Slayt

kapamak için bunu kullanmanın yararı olabilir diyebiliriz. Bir de bağımsız olarak bu hastalara başka sorular sormuş ya da bu kişilerin işten güçten son bir ay içerisinde kaldığını öğrenmiştik. Bunu öğrendiğimizde gördüğünüz gibi şu alanlarda, 6 parametrede, bu kişilerin işten güçten kalmaları ile bu WHODAS skorlarının korelasyonu ortalama olarak 0.63 ve yarım gün ya da daha fazla işten güçten kalma, işe gidememe oranı ise %52. Bunu da bir bakıma independent bir verification olarak düşünebilirsiniz.

Depresyondaki hastalarla karşılaştığımızda Meksika'da, Nijerya'da, Londra'da SF-36 London handicap scale'la bakıldığında buna 'effect size' dediğimiz yani hastaların önceki ortalamalarıyla sonraki ortalamalarının grup içindeki varyansa oranına özetleyen, öncesi ve sonrası arasındaki farkın ne kadar anlamlı olduğunu gösteren bir indikatör. Diğer ölçüm araçlarıyla daha iyi olduğunu görüyorsunuz. Gördüğünüz gibi Seattle'da SF-36'nın onlar sadece mental health alt boyutunu kullanmışlar. Mental health subcomponentin SF-36'sı WHODAS'tan çok daha iyi. Neden diye araştırdığımızda orada mental health subcomponent'in "Do you feel blue or depressed" diye direkt bir sorusu var. O da bildiğiniz gibi depresyonun ana semptom sorusu yani, az önce Hasan hocamın söz etmiş olduğu sirkülarite ben aslında o anlamda SF-36'nın sadece bir fonksiyon ölçme aracı değil, bir yandan da semptom ölçme aracı olduğu için böyle bir sonuç doğurduğunu görüyoruz, yani sonuçta bir yerden yakalanıyorsunuz. Bu şizofrenide çok daha iyi ortaya çıkıyor. Çünkü şizofrenide SF'nin ya da London handicap scale'sının yakalamadığı şeyleri yakalıyor, ama alcohol dependence'da da öyle mesela Londra'da, ki onların bir anlamda flagships ölçeği London handicap scale idi. Hip arthroplasty, ki sizin alanınızla çok daha yakın olabilecek bir ölçek, burada 72 hasta da WHODAS'ın effect size'ı 0.77, London handicap scale'in effect size'ı 0.42. O anlamda girişim öncesi ve sonrası ortaya çıkan farkı, konuşmanın başlangıcında HIV olan çocuğun tedavi öncesi ve sonrası farklarını ne denli iyi ortaya çıkarabileceğini gösteren bir şey. Tekrar o slaytı hatırlatmışken şuna döneyim, Dünya Sağlık Örgütü olarak HIV'in sınıflandırması için 4 evremiz var ve bunu CDC'iy-

WHODAS 2.0 yapısı

- **Uzun versiyon (36-item)**
 - çoğu detayı verir
 - toplam ve 6 ana alandaki özel işleyiş skorunu verir
 - yüz yüze görüşmeyle, kendisinin yapabileceği ve vekaleten yapabilecek formları var
 - görüşme süresi, ortalama: 20 dak.
- **Kısa versiyon (12-item)**
 - Toplu anketler ya da sağlık-sonuç çalışmalarının kısa değerlendirmeleri için faydalı
 - genel işleyiş puanı hesaplamaya olanak sağlar
 - 36-itemlik versiyonun %81lik varyansını açıklar
 - yüz yüze görüşmeyle, kendisinin yapabileceği ve vekaleten yapabilecek formları var
 - görüşme süresi, ortalama: 5 min.
- **Hibrid Versiyon (12+24-item)**
 - problemli domainler için 12 maddeyle başlar
 - 12 maddeye pozitif sonuçlar eklenerek 24 ek soruya çıkabilir
 - Sadece yüz yüze ya da bilgisayar uyumlu test (CAT) ile yönetilir

11b. Slayt

WHODAS 2.0

çalışma engeli ile ilişkisi

WHODAS 2.0 Skoru	Evdeki sorumlulukların tam olarak yerine getirilemediği günler	Yarım gün ya da daha fazla işe gidilemeyen günler
<i>Kognisyon</i>	.28	.15
<i>Hareketlilik</i>	.42	.31
<i>Özbakım</i>	.48	.40
<i>Toplumsal</i>	.33	.28
<i>İş & Ev işleri</i>	.68	.58
<i>Katılım</i>	.53	.49
TOPLAM	.63	.52

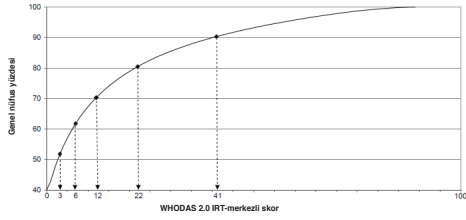
12. Slayt

le ortak çalışıyoruz, çok sayıda uzman geliyor, tanıya bağlı ölçütler için bir şeyler yapıyor ama sahada kullanılan, özellikle Afrika'da, Tayland'da kullanılan en önemli outcome ölçütü ne biliyor musunuz? Hastanın kilosu. Veri tabanlarına baktığınızda, CD4 countlarını görüyoruz ama outcome açısından baktığımızda tek şey hastanın kilosu. Öldü mü, yaşıyor mu, yaşıyorsa kilosu ne? Ve bu kadar parasını harcadığımız ABD, 15 milyar dolar ve bu parayı her yıl Afrika'ya yatırıyorlar ama ölçtükleri outcome kriteri, kilo, ölüm ya da yaşıyor. Yani o anlamda diyorum ki daha anlamlı ölçme, daha duyarlı ölçme sistemleri varken WHODAS olabilir, başka bir şey olabilir hiç önemli değil, herhangi bir ölçme, hiç ölçmemekten daha iyi. Öte yandan o kadar para yatırıyoruz, karşılığında ne alıyoruz, ölçmeyi bilmiyoruz. O yüzden belki kanıta dayalı tıbbın bize bu yönden yol göstermesi söz konusu. Skor açısından, iki tür skor var. Özellikle bu alerti yaptığımız için basit bir şekilde aritmetik bir toplam yapabilirsiniz ama daha duyarlı bir şey yapmak istiyorsanız bir bilgisayar programı var. O bilgisayar programı da alert based score'la bir toplam eğri veriyor. Bu uzun sitemin, bu da kısa sorulunun hemen hemen aynı özellikleri gösterdiğini görüyorsunuz (13. ve 14. slaytlar).

Ve bu birçok araştırmada kullanıldı, giderek de artıyor. Özellikle Amerika'da şu anda VA bunu bütün tümör register'inde kullanıyor. Amerika'da özellikle Afganistan'dan, Irak'tan gelenlerin fonksiyonel durumlarını değerlendirmek için standart bir şey haline geldi. Onun dışında dünyanın değişik yerlerinde de kullanılıyor. Ve en önemli olarak da değişik alanlarda örnekler. Sizin alanınızda artritte, belki sistemik sklerozda, bel ağrısında, ankilozan spondilitde kullanıldığı çalışmalar var, onların hepsi bu kitapta ve web sitesinde var. Diğerlerinden farklı bir ölçme aracı değil, yani genel anlamda SF'den vs ama bir bakıma kültürel uyumluluğu olduğu için, daha geniş bir kapsamı olduğu için, ICF'ye daha yüksek bir bağdaşıklığı olduğu için, Dünya Sağlık Örgütü bunu, parasız, özel, bütün araştırma ve kullanım açısından kullanılmasını öneriyor. Bu anlamda bir software de var, onu da download edebilirsiniz.

Yaşar Karaaslan: Çok teşekkür ederim. Benim

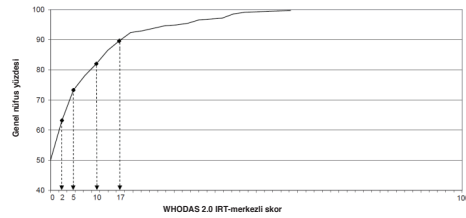
WHODAS 2.0 için IRT-merkezli skorların nüfus dağılımı- Tam (uzun) versiyon



Source: WHO multi-country survey study on health and responsiveness 2000-2001

13a. Slayt

WHODAS 2.0 için IRT-merkezli skorların nüfus dağılımı- Kısa versiyon



Source: WHO multi-country survey study on health and responsiveness 2000-2001

13b. Slayt

WHODAS 2.0 software



- Farklı WHO-DAS versiyonlarını yapılandırma.
- CAPI araştırma hastalarını şu yollarla minimize eder:
 - giriş değerlerini çek ederek
 - örnek atlama yöntemini kontrol ederek
- Kağıt kalemle bilgi girişini destekler.
- Masaüstü ve dizüstü bilgisayarlarda çalışır.

14. Slayt

konuyla belki çok ilgili değil ama tanımlama açısından, çünkü Dünya Sağlık Örgütü'nün bir hastalık tanımına çok itirazım var. Semptomsuz hastalık olur mu?

Bedirhan Üstün: Semptomsuz hastalık olabilir. Çünkü semptom dediğimiz şey, adı üzerinde bir belirti. Bir yüzey belirtisi. Altta yatan hastalık süreci çok önceden başlamış ama asemptomatik olarak o hastalığın sürdüğü durumlar söz konusu. Yani bu HIV'de var, Parkinson'da var, demansda var, Huntington chorea'de var, yani sizin genomik olarak hastalık fizyolojisi altta daha önceden başlamış olabilir ama semptom vermemiş olabilir.

Yaşar Karaaslan: Özellikle osteoporoz veya hiperlipidemi, bunlar risk faktörüdür, hastalık değildir. Ama bize hastalık diye öğretiliyor. Osteoporoz bir hastalık değildir, kırık hastalığıdır, osteoporoz onun risk faktörüdür. Hiperlipidemi de hastalık değildir, kalp krizi geçirmek hastalığıdır, hiperlipidemi onun risk faktörüdür. Ama bunları hastalık olarak tanımlayıp sizin dediğiniz tanımla hani belki disorder ve başka şeyler daha uygun olabilir. Böyle olunca modern tıp hastalıklar buluyor, tamamen sağlıklı sizin dediğiniz ICF tamamen normal insanı götürüp testler yapıp sağlam adamı hasta yapıyoruz. Buna da Dünya Sağlık Örgütü aracılık ediyor. Mesela osteoporoz tanımı Dünya Sağlık Örgütü'nün tanımına göre şu şu olan ve diyor ki kırık riskini arttıran bir sistemik hastalıktır, diyor. O zaman bu bir risk faktörüdür, adı üstünde yani şunlarla karakterize kırık riskini arttıran bir hastalıktır. Bu hastalık olmaması lazım, risk faktörü olması lazım.

Bedirhan Üstün: Tartışma bence sağlıklı bir tartışma, biz bu tartışmayı özellikle, gerek bilim adamları, gerek filozoflar bir sürü kişiyle mantıkçılarla da yaptık. Bir takım hastalıklar diğerleri için risk faktörü olabilir. Yani bir şeyin risk faktörü olması hastalık olmaması anlamına gelmez. Mesela hipertansiyon da bir risk faktörüdür, stroke için veya başka bozukluklar için, ama hipertansiyonun altında, hastalık olduğu durumlar söz konusu. Yani sonuç olarak bir hastalık başkası için risk faktörü olabilir. Osteoporoz o anlamda risk faktörüysse hastalık mıdır değil midir, onun tartışmasını ben yapabileceğim

kadar uzman deęilim. Ama gerek uzmanlar gerek dięerleri hiperlipidemi, hipertansiyon iin, bunların altında yatan bir takım fiyolojik bozuklukların olduęu, o yzden onları hastalık olarak grdęmz ve kabul etmemiz gerektięini sylyorlar. Osteoporoz

kıyıda bir durum, haklısınız ama osteoporozun altında da bir bakıma onu tetikleyen, destekleyen şeyler olabilir. Ayrıntıları aısından tartıřmaya hazırım ama genel ilke aısından ben size bildięim ilkeyi aıklamaya alıřtım.

KLİNİK ÇALIŞMALARDA BAYES METOTLARI

Emmanuel Lesaffre

*Biyoistatistik Bölümü
Leuven Katolik Üniversitesi
Belçika*

Tam 27 yıl önce yine Çeşme’de, farklı bir otelde ilk konuşmam için davet edilmiştim. O zaman da çok iyi zaman geçirmiştım, şimdi de öyle. Organizatörlere ve Hasan Yazıcı’ya bu toplantıya beni davet ettikleri için teşekkür ediyorum.

İlk olarak Bayes klinik çalışmalarından, ardından istatistik yorumlamaya yönelik farklı bir yaklaşım olan Bayes teorisinden bahsedeceğim. Dediklerimi izlemek için matematik de gerekli, ancak konuşmamda, tek bir integral dışında çarpma ve bölmeden öte bir şey olmayacak. Soru şu; neden zaten kullandığımız bir yöntem varken başka bir bakış açısına ihtiyaç duyuyoruz? Var olan bakış açımızdaki sorun ne? Bu konuşmanın amaçlarından biri p değerinin kullanıldığı klasik yaklaşımda neyin eksik olduğunu göstermek.

İlk olarak bazı Bayes randomize kontrollü çalışma (RKÇ) örnekleriyle işe başlayalım. İlk çalışma; amfizemli hastalarda akciğer volümünün bronkoskopik yöntemlerle düşürülmesinde, hava yolu stenti uygulamasıyla plasebonun karşılaştırılması. İstatistik analizde, yazarların örnekleme boyutunu Bayes adaptif yaklaşımıyla saptadığı belirtilmiş. Tedavi etkisini değerlendirmek üzere, sonsal olasılık (tedavi sonrası, posterior probability) kullanılmış. İleride, bu sonsal olasılığın çok daha popüler p değerinden daha işe yarar olduğunu göreceğiz.

İkinci çalışma ise antiaritmik ilaç tedavisi ile kateter ablasyonu denen girişimsel tedavinin karşılaştırılması. Aritmiyi tetiklemek ve odağını saptayınca da yakmak için belli noktalara kateter yerleştirilmiş. Yani, bu çalışmada ilaç tedavisi girişimsel bir yöntemle karşılaştırılmış. Çalışma süresince Bayes kuramının öngören olasılıklarına (predictive probability) dayanan analizler tasarlanmış. Bu analizler bizi çalışmanın sonucunda, nihai analize götürüyor. Demek bu RKÇ de Bayes tasarımına dayalı. Bayes ara analizlerine dayanan bu yöntemin klasik yaklaşımlara göre daha içgüdüsel olarak daha anlaşılır olduğunu ileride göreceğiz.

Üçüncü örnek, akut miyokard enfarktüsü hastalarında Perkütan Translüminal Koroner Anjiyoplasti

Bayes teoremi

Duyarlılık, özgüllük ve öngörü değerleri

Klasik tıbbi örnek:

- pred+ ve pred- değerlerin duyarlılık (sens), özgüllük (spec) ve prevalans (prev) ile ilişkisi

$$\text{pred+} = \frac{\text{sens} \cdot \text{prev}}{\text{sens} \cdot \text{prev} + (1 - \text{spec}) \cdot (1 - \text{prev})}$$

- $\text{pred+} = P(D+|T+)$
- $\text{sens} = P(T+|D+)$ and $\text{spec} = P(T-|D-)$
- $\text{prev} = P(D+)$

1. Slayt

(PCTA) sonrası süpersatüre oksijen yüklemesinin infarkt boyutu üzerine etkisi araştırılıyor. Çalışmanın kuvveti ve istatistik bölümünde çalışmanın ana son noktalarının (primary end points) analizinde daha önceden yapılmış AMIHOT-I çalışmasında gözlenenlere dayanan hiyerarşik bir Bayes modeli yaklaşımı kullanıldı deniliyor. Burada önemli olan nokta şu; önceki çalışma sonuçlarından edinilen deneyim güncel çalışmada daha küçük örneklem boyutu kullanarak aynı güce ulaşılmasını sağlıyor.

O halde, yukarıdaki RKÇ'nin Bayes bileşenleri sonsal olasılık ve sonsal öngörü olasılığı (posterior prediktif değer). Konuşmanın geri kalanında, bu kavramları açıklamaya çalışacağım.

Burada bahsedeceğim istatistik yaklaşım Bayes teoremi. Sanırım çoğunuz Bayes teoremini zaten biliyorsunuz. Pozitif öngörü değeri (pozitif prediktif değer) ve negatif öngörü değeri (negatif prediktif değer) ile duyarlılık, özgüllük ve prevalansın ilişkilendirildiği klasik tıp istatistiği kavramlarına yabancı değilsiniz. Aşağıdaki formül tıp araştırmacılarının iyi bildiği bir formül. Pozitif öngörü değeri testin uygulandığı popülasyonda hem hastalığı olup hem de testi pozitif olanların toplumda tüm testi pozitif olanlara oranıdır. Duyarlılık ise tam tersidir. Toplumda aranan hastalığı olanlar arasında testin ne kadar sık olduğunu gösterir. Formüle baktığımızda, toplumda bir hastalık ne kadar sık ve o hastalığı arayan testin duyarlılığı ne kadar yüksekse aynı testin söz konusu hastalık için pozitif öngörü değerinin yüksek olduğu anlaşılır. İlişkilerin eşdeğer bir formülasyonu sonraki slaytta olasılıklar bağlamında görülebilir (1. slayt). Bayes teoremi hasta olanda testin pozitif olma olasılığından testi pozitif olanda hastalık bulunma olasılığına ulaşma olanağı verdiğinden, teoreme Ters Olasılık teoremi de denir.

Bayes teoremindeki payda görünürde hasta olma olasılığı diye adlandırılır, bu da aslında pozitif teste ulaşma olasılığıdır. Testi pozitif olan kişilerin hasta olma olasılığı testin duyarlılığı ile prevalansın çarpımının standardize edilmiş şeklidir. Eğer duyarlılık yüksekse, birey yüksek olasılıkla hastadır. Formül Bayes teoreminin bir örneğidir (2. slayt) ve hastalık durumunu, testle ve hastalık

Bayes teoremi

Duyarlılık, özgüllük ve öngörü değerleri

Formül:

$$\begin{aligned} P(D+|T+) &= \frac{P(T+|D+) \cdot P(D+)}{P(T+|D+) \cdot P(D+) + P(T+|D-) \cdot P(D-)} \\ &= \frac{P(T+|D+) \cdot P(D+)}{P(T+)} \end{aligned}$$

Bu Bayes teoremi için bir örnektir

$$\begin{aligned} P(\text{Hyp}|\text{Data}) &= \frac{P(\text{Data}|\text{Hyp}) \times P(\text{Hyp})}{P(\text{Data})} \\ &\propto P(\text{Data}|\text{Hyp}) \times P(\text{Hyp}) \end{aligned}$$

Hyp = hipotez/hastalık

Şimdi hepiniz Bayesçisiniz!

2. Slayt

?



Bakın!

3. Slayt

Bayes teoremi

Duyarlılık, özgüllük ve öngörü değerleri

Bayes teoremi farklı prevalans/popülasyonlar ve **doktor ofisi** için pred+ değerini hesaplamamıza olanak verir.

65+ yaşta popülasyondaki **Tip II Diyabet**

- Gerçek Prevalans = 0.50 = $P(Hyp)$
- Duyarlılık = Test + Tip II Diyabet var = 0.80 = $P(Data = + | Hyp)$
- Özgüllük = Test - Tip II Diyabet yok = 0.90 = $P(Data = - | Hyp)$
- Gözlenen Prevalans = $P(Data = +)$ = testin pozitif olma olasılığı
- Pred+ = $P(Hyp | Data=+)$: Bayes teoreminden
- Pred- = $P(Hyp | Data=-)$: Bayes teoreminden

4. Slayt

Bayes teoremi

Duyarlılık, özgüllük ve öngörü değerleri

Doktor ofisi: 65+ yaş hasta **Tip II Diyabet** başvurusuyla gelir

- İlk Görüşme: (önsel) Tip II Diyabet olasılığı popülasyon prevalansında elde edilmiştir = 0.50 ($P(Hyp)$)**
- Vaka 1: pozitif test**

$$P(Tip\ II\ Diyabet|Test+) = \frac{0.80 \times 0.50}{0.80 \times 0.50 + 0.10 \times 0.50} = 0.89$$

- İkinci Görüşme: (Sonsal) Tip II Diyabet olasılığı prevalans ve pozitif teste göre = 0.89 ($P(Hyp | Data)$)**

- Vaka 2: negatif test**

$$P(Tip\ II\ Diyabet|Test-) = \frac{0.20 \times 0.50}{0.20 \times 0.50 + 0.90 \times 0.50} = 0.18$$

- İkinci Görüşme: (Sonsal) Tip II Diyabet olasılığı prevalans ve negatif teste göre = 0.18 ($P(Hyp | Data)$)**

5. Slayt

prevalansı ile desteklenen veya desteklenmeyen bir varsayımı ifade eder. Bu varsayım da uygulanan test sonucu, eldeki hastalık sıklığının da yardımıyla kanıtlanır veya kanıtlanmaz.

Bayes teoreminin verdiği mesaj şudur: Eğer gözlenen veri varsayıma uyuyorsa, varsayım muhtemelen doğrudur. Diğer bir deyişle elde edilen verilerin söz konusu varsayım altında elde edilme olasılıkları yüksektir ve inandırıcıdır ($p[hyp]$ muhtemelen yüksek). Bu mantıklı, değil mi? Dediklerimin varsayımları sınamak için uygun bir strateji olduğunu düşünüyorsanız, “Bayesçisiniz” demektir.

Başka bir test yapayım. Bu resimde ne görüyorsunuz? (3. slayt)

Bir Katılımcı: Bu açıdan bakıldığında, köpeğin bu tarafı siyah, diğer tarafının rengini bilmiyorum.

Bu klasik bir istatistikçi sözü. Bayesçi ise şöyle der: ‘Bir tarafı siyah ve diğer tarafı farklı renkte olan bir köpek hiç görmedim’. Yani bu açıklamayı yapmak için önceki bilgileri kullanır.

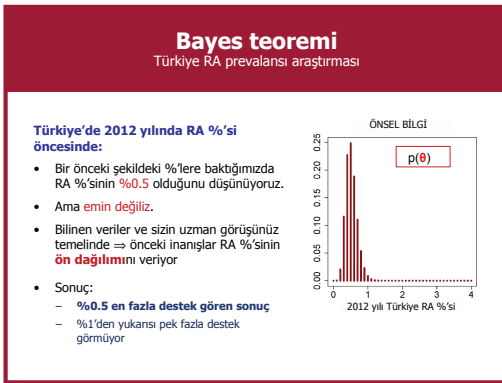
Şimdi pozitif öngörü değerini poliklinikte uygulayalım. Mesela, 65 yaş üstü popülasyonda tip II diyabet prevalansını araştırıyoruz. Gerçek prevalans 0.5. Duyarlılığı 0.8 ve özgüllüğü 0.9 olan bir tanı testi seçtik. Polikliniğe giren bir hastanın bir pozitif testle (vaka 1) ve bir negatif testle (vaka 2) diyabet olma olasılığını bilmek istiyoruz. Bayes teoremi uygulayarak, ilk vakada 0.89 ve ikinci vakada 0.18 olduğunu görüyoruz (4. ve 5. slaytlar).

Şimdi, bununla daha alakalı olan örnekleri inceleyelim. Ben küçük bir alıştırma yapacağım. 2012 yılında Türkiye’deki romatoid artrit prevalansını belirtmek için Yunanca sembol ‘ θ =teta’yı kullanacağım. Google taraması sonucu bulduğum değerler; prevalans Japonya’da %0.3, Çin’de %0.3, ABD’de bu oran %1’e çıkmış. Soru, 2012 yılında Türkiye’de RA yüzdesi ne? Araştırdığınızda Türkiye için bir istatistik bulamazsınız, peki bu oran %0.2, %0.3, %0.5 ya da %1’den hangisi? Şimdi, 500 denekten oluşan bir çalışma yaptığımı ve bu çalışmadaki 8 deneğin RA’sı olduğunu, yani

bu çalışma için RA prevalansının %1.6 olduğunu varsayalım. Biliyoruz ki, bu yüzdenin istatistiksel belirsizliği var ve %95 güven aralığı %0.82 ile %3.12 arasında, yani oldukça belirsiz. Soru şu ki artık Google’da bulduğumuz tarihi bilgileri göz önüne alabilir miyiz, bunları dikkate alabilir miyiz ve hatta uzman bilgisi ile birleştirebilir miyiz? Çünkü Google’da bulduğum tarihi bilgiler Türkiye ve 2012 yılı için geçerli değil, yani belki de Türkiye ve 2012 yılı açısından bunu bir rakama çevirmek üzere bazı uzman görüşlerini eklemem gerekebilir. Peki, bunu nasıl yapabileceğimiz konusunda hemfikir olmak mümkün mü?

Türkiye’de RA prevalansı hakkında romatologlardan bilgi aldığımızı varsayalım. Her bir dikey çizgi x ekseninde karşılık gelen değer test öncesi olasılığını (önselliğini) göstermekte. Bütün bu yükseltelerin toplamı 1’e eşit olmalı (6. slayt). %1’in biraz üzeri temelde 0 kabul edilir. Bu gösterilen grafik önsel dağılım olarak adlandırılır, dağılıma olan önceki güveni gösterir. Grafik bize ne diyor? Örneğin, %0.5 en çok desteklenen bilgi ve uzmanlar Türkiye’de RA’nın %1’den fazla olmadığını düşünüyorlar. Ön dağılım $p(\theta)$ ile gösterilir. Peki, bu 500 kişilik tarama bize ne söylüyor?

Taramaya göre, 500 kişiden 8’i RA hastası. O halde, prevalans tahminimiz %1.6, ancak bu sadece bir tahmin ve gerçek prevalansın emin değiliz. Her bir olası gerçek prevalans için, 500 kişiden 8’inin RA olduğu anlamına gelen L ile ifade edilen önsellik veya olasılık hesaplanabilir. Örneğin, 500 kişiden 8’inin RA olma olasılığı $\theta = \%1.6$ yani 0.071’dir. Bu değer tüm olası θ değerleri için en yüksek olasılıktır. $\theta = \%0.5$ olduğunda ise 0.0015’dir. Tüm yüzdelere için olasılık hesaplandığında ve yükseltinin olasılık anlamında kullanıldığı bu değerleri çubuklarla gösterip tüm yükseltileri birleştirdiğimizde, $L(\theta|D)$ ile gösterilen ve olasılık fonksiyonu denen bir eğri elde ederiz (burada D veri anlamında kullanılıyor) (7. slayt). Olasılık eğrisi hangi θ değerinin veri ile desteklendiğini grafik olarak tanımlar. %0.5’in altındaki değerler veri ile pek desteklenmez ve aynı şey %4’ün üzerindeki değerler için de geçerlidir.



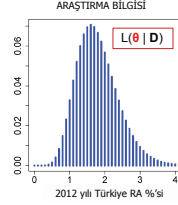
6. Slayt

Bayes teoremi

Türkiye RA prevalansını belirleme araştırması

2012 Türkiye araştırması: Olasılık (likelihood)

- Araştırma tarafından desteklenen RA %'leri neler?
- Sonuç:
 - %1.6 en fazla destek gören
 - %4'ten üzeri fazla destek görmüyor



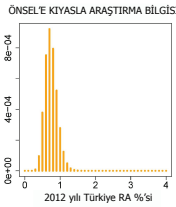
7. Slayt

Bayes teoremi

Türkiye RA prevalansını belirleme araştırması

Önsel bilgi ile araştırma bilgisinin birleştirilmesi:

- Her ikisi tarafından desteklenen değerler genel olarak iyi destek almış.
- Toplam** genel desteği verir
- Sonuç:
 - Önsel(%0.5) X Lik(%0.5) = 0.25 X 0.0015
 - Önsel(%1.6) X Lik(%1.6) = 5.32E-6 X 0.071
 - Önsel(%0.7) X Lik(%0.7) = 0.11 X 0.0084
 - %0.7 en fazla desteklenen**



Lik: Olasılık (likelihood)

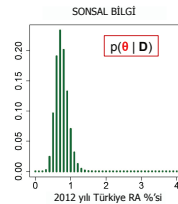
8. Slayt

Bayes teoremi

Türkiye RA prevalansını belirleme araştırması

Sonsal (Post):

- Standardize **toplam**
- Bu **Bayes teoremi**
- Sonuç:
 - Sonsal (%0.5) = 0.096
 - Sonsal (%1.6) = 9.51E-5
 - Sonsal (%0.7) = 0.23**
- Burada %0.1, %0.2, ..., %3.9, %4, ... gibi RA %'leri ile çalıştık.
- Ancak gerçekte RA %'si sürekli
- Dolayısıyla önsel, olasılık ve sonsal **sürekli**dir.



9. Slayt

O halde, bu çalışmada iki bilgi kaynağımız var (gözlenen veri D ve olasılık $L[\theta|D]$ ve önsel bilgi/önseziler $p[\theta]$). Olasılığın önceki formülde duyarlılığın bir genellemesi olduğu ve pozitif öngörü değeriyle duyarlılık, özgülük ve prevalansı ilişkilendirdiği unutulmamalı. Önsellik prevalansın bir genellemesidir. Bayes teoreminde olasılık (likelihood) ve önsellik alınıp sonsallık elde edilir ve ardından standardize edilir. O halde, eğer θ değeri hem olasılık hem de önsellikle desteklenirse, sonuç nispeten yüksek ve sonsal inandırıcılığı da güçlü olacaktır. Sonraki slaytta, sonsal dağılım denen ve $p(\theta | D)$ ile ifade edilen standardize edilmiş sonuç görülmektedir (8. slayt).

%0.5'in sonsalı 0.096'dır ancak %1.6 hiçbir θ değeri ile desteklenmemektedir. Önceki bilgi ile mevcut verinin kombinasyonu sonrası en inandırıcı θ değeri %0.7'dir. Altını çizmem gereken son ayrıntı, farklı değerleri %0.1'lik sıçramalarla dar bir skalaya sıkıştırmış olmam. Ancak, RA hastalarının yüzdesi sürekli, yani %0 ile %100 arasındaki tüm değerler mümkündür ve akıcı eğriler önsel, olası ve sonsal temsil etmelidir (9. slayt).

Şimdi bu alıştırmaı önsel, olası ve sonsal için sürekli eğrilerle birlikte tekrar edelim. θ olarak adlandırdığım RA yüzdesi üzerinde bazı ön bilgilerim var ve bu ön bilgiler artık bir sürekli (mavi) eğri halinde. %0.5 oranındaki değerler konusunda pek çok önsezim olduğunu ve %1'in üzerindeki veriler konusunda ise bir önsezim olmadığını görülmekte. Olasılık eğrisi (kırmızı), verilerin θ hakkında bana ne bilgiler sunduğunu göstermekte ve %0.5 ve %3.5 arasındaki θ değerlerinin verilerle desteklendiği anlaşılıyor. Yeşil eğri önsel bilginin standardize edilmiş sonucunu göstermekte ve buna sonsallık deniyor. Aynı sonsallık bana 2012 yılında Türkiye'de olduğuna inandığım RA yüzdesini söylemekte. Sonsallık eğrisi, önsellik ve olasılık eğrileri arasında ve sonsallık bilgisi önsel bilgi ile bu çalışmadan elde edilen bilgi arasındaki uzlaşmayı ortaya koyuyor.

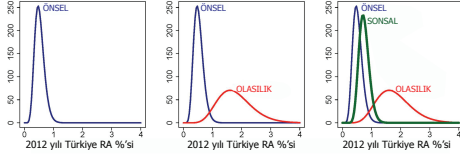
Bu alıştırmaıdan ne öğrendik? Öncelikle bu taramadan elde ettiğimiz bilgiye baktığımızda, gerçek θ değeri için %95 güven aralığına bakabiliriz. Önsel bilgilerle olasılığın birleştirilmesi

Bayes teoremi

Türkiye RA prevalansını belirleme araştırması

Araştırmadaki RA %'sine dair bilgi

- **Sonsal:** önsel ve veri dikkate alındığında RA %'sinin ($= \theta$) olasılığını ifade ediyor.



10. Slayt

sonsalı verir ve sonsaldan bir 'güvenilir aralığı' (credible interval) hesaplarız. Güvenilir aralık, önsel ve olasılık bilgilerinden elde edilen sonsal bilgi ışığında en güvenilir θ değerlerini içerir. %95 güvenilir ara %0.44 ile %1.12 arasında, yani %95 güven aralığından çok daha dardır. O halde sonuç olarak, sonsallık çok daha kesin bir θ değeri sağlamaktadır, ancak tabii önsel bilgi doğru olmak şartı ile. Pratikte, önsel bilginin bilimsel olarak anlamlı olduğundan emin olmak gerekir.

Verdiğimiz örnekte biz görece güçlü bir önsel bilgi ile küçük bir çalışmadan elde edilen verileri birleştirdik. O halde, önsel bilginin büyük bir etkisi olmalı. Genelde veri önsel bilgiye baskındır.

Örneğin büyük bir taramadan gelen verilerimiz ve buna karşın görece zayıf bir önselimiz varsa, olasılık en yüksek etkiyi gösterecektir. Bayes metodunda, ön bilgilerin önemi kasıtlı olarak azaltılabilir. Burada yaptığım şey de bu. 10. slaytta gördüğünüz gibi ön bilgilerim öncekinden daha düşük seviyede, olasılıklarım aynı, veriler aynı ancak artık sonsal olasılık ve ön bilgi arasında daha dengeli yerleştirilmişler. Ön bilgiler hâlâ son derece etkili ancak önceki kadar değil. Önsel bilgimiz olmasa veya onu kullanmak istemesek bile Bayes teoremini uygulayabiliriz. Bu durumda bilgilendirici olmayan bir önsel alabiliriz. 11. slaytta, olası önsel düz bir çizgi. Düz çizgi, tüm θ değerlerinin önsel olarak inandırıcı olduğunu gösteriyor. Yani ön plana çıkan bir θ değeri yok. Ardından, olasılık sonsalla karşı karşıya gelir ve önsel bilgiyi çıkarımımıza eklemediğimizi ortaya koyar.

Bayes teoremi

Türkiye RA prevalansını belirleme araştırması

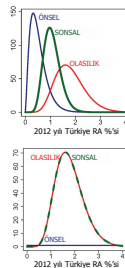
Tarihsel verinin etkisi çok mu büyük?

- Küçük araştırma, güçlü önsel $\Rightarrow$ önselin etkisi daha fazla
- Büyük araştırma, zayıf önsel $\Rightarrow$ olasılığın etkisi daha fazla
- Önsel bilgide azalma (discount)

Önsel bilgiyi kullanmayı istiyor muyuz?

- Bilgilendirici olmayan önselin kullanımı = düz önsel (flat prior)

Bayes teoremi sonsal bilginin hesaplanması için bir makine görevi görür.

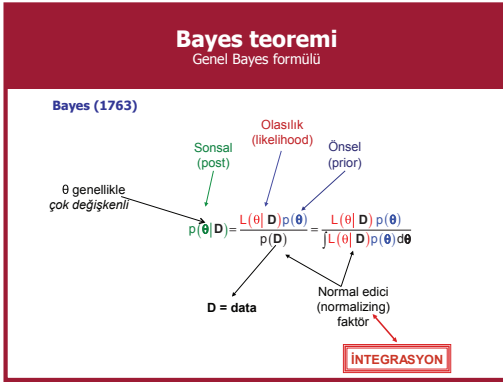


11. Slayt

Aslında, Bayes teoremi bir çıkarım aygıtıdır. Güncellenmiş bilgileri, yani veri toplayarak elde ettiğiniz bilgileri alıyor ve ön bilgilerle bunları birleştiriyor. Bu da bir Bayes teoremi, ancak şu şekilde okuyoruz: eğer olasılık olarak ifade edilen veriler elinizde ise, onu ön bilgilerle çarparsınız sonsalı elde edersiniz. “D” veri anlamında kullanılır (12. slayt). Paydada görülen integrale normalleştirme faktörü denir. Çok sayıda parametre için integral hesaplaması yapmak çok zor hatta imkânsız olabilir. Bu soruna sonradan tekrar döneceğiz.

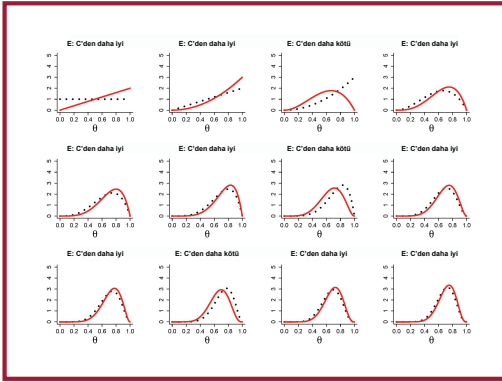
Bir Bayes analizinin sonuçları sonsal olasılıklarla formüle edilir. Örneğin, Türkiye’de 2012’de RA yüzdesinin %0.5’den yüksek olması sonsal olasılığını bilmek istersek, sonsal dağılım aralığı (%0.5, %100) eğrisinin altındaki alanı hesaplarız. Aritmatikle kolayca elde edilebilir ve 0.93’e eşittir. O halde, çalışmanın sonuçlarını gördükten ve önsel bilgi ile birleştirdikten sonra, Türkiye’de 2012’de RA yüzdesinin %0.5’i aşma sonsal olasılığı 0.93’dür. Klinik çalışmalarda iki tedavinin etkinliği arasında yüzde veya ortalama ile hesaplanan farkı sonsal olasılık, yani Δ ile ifade ederiz.

İki varsayıma da bakalım. Sıfır varsayımı deneysel kolun (E) etkinlik açısından kontrol kolundan (C) daha az etkili olduğunu gösterir. Alternatif varsayımsa deneysel kolun kontrol kolundan daha etkili olduğunu söyler. Bayes teoremi, araştırmacıya deney sonuçlarına (veri D ile gösterilir) dayanarak varsayımımızın (sıfır veya alternatif) doğru olma olasılığını verir. Bayes teoremi deneysel tedavinin gerçekten daha etkili olduğu varsayımının doğru olup olmadığı sorusuna da cevap verebilir. Bu ise sadece dolaylı bir cevabı, yani gözlenen sonucun sıfır hipotezinde gerçekleşmesi olasılığını veren p değerinin sağladığının tam tersidir. İki sonsal olasılık burada olduğu gibi ayrılabilir ve sonsal oranı denen sonsal olasılık oranını verir. Bu oran önsel oranla verinin sıfır varsayımında gerçekleşme olasılığının alternatif varsayımında gerçekleşme olasılığına bölümü ile elde edilen sonuçla çarpımına eşittir. O halde, başlangıçta alternatif varsayım hipotez için önsel bir tercih yoksa yani her iki varsayım z için önsel olasılık 0.5’e eşit ise, kırmızı renkli oran deneyi yaparak fikrimizi ne kadar değiştirebileceğimizi bize söyler. Yani, Bayes teoreminin, araştırma sorularına doğrudan cevap verdiğini ve verileri gördükten sonra fikirlerimizi nasıl değiştirdiğimizi de gösterdiğini söyleyebiliriz. Bayes teoremi hakkındaki bir diğer güzel şey de, doğal insan davranışını taklit etmesidir. Çünkü bugünün sonrası yarının öncesi olarak da kullanılabilir, yani sıralıdır. Her bir aşamada, bir önceki aşamanın sonrasına atıfta bulunduğunuzda, onu uzman bilgisi ile birleştirebilirsiniz.



12. Slayt

İşte bir alıştırma (13. slayt). Bir çapraz çalışma alalım. Hastalarda E'nin etkinliğini C ile karşılaştırıyoruz. θ , E'nin C'den iyi olma olasılığını gösteriyor. Tek bir olgu ile başlıyoruz. Başlangıçta çalışmada tek bir olgu var ve deneysel tedavi kontrolden daha iyi. Çalışmadan önce, E'nin C'ye göre performansı hakkında herhangi bir bilğimiz yok. Çalışmanın başlangıcında önsel düz bir çizgi. Gözlenen veriler ile birlikte, yüksek θ değerlerini tercih eden bir sonsal elde ediyoruz. Bu sonsal ilk hastanın sonuçları sonrası θ değerleri hakkındaki bilgileri içeriyor ve çalışmada ikinci hasta incelendiğinde önsel olarak kullanılabilir. İkinci hastanın sonucuna göre de E, C'den daha iyi ve sonsal yine daha yüksek θ değerlerini tercih ediyor. Bu sonsal da yine sonraki hastanın önseli olarak kullanılıyor.



13. Slayt

Bayes yaklaşımıyla bugüne kadar bildiklerimizle gelecek sonuçların dağılımını belirleyebiliriz. Bu durum sonsal pozitif dağılımdır (PPD). PPD bir klinik çalışmanın muhtemelen pozitif mi negatif mi olacağını öngörmeye yararlı bir araçtır ve bir klinik çalışmanın faydasızlık analizinde, yani bir çalışmanın pozitif sonuçlanması şansını kontrolde işe yarar.

Bayes teoreminin mekaniklerinden bahsettikten sonra, artık tarihten söz etmeye sıra geldi. Bayes teoremi, matematik alanında saygın bir yeri olan Thomas Bayes'in adıyla anılmakta. Bayes, eldeki verilerden herhangi bir olasılıkla bir varsayım çıkartabilmek için önce ulaşılmak istenen varsayımın da bir olasılığının olması gerektiğini ilk söyleyen bilim insanı. Ünlü teoremi dâhil, çalışmaları 1763 yılında, kendisinin ölümünden 2 yıl sonra arkadaşı Richard Price tarafından yayımlanmış. Bayes'in tüm yazılarını araştırmış, toplamış ve kendi yorumlarını da ekleyerek şu ünlü kitabı hazırlamış: *An Essay toward a Problem in the Doctrine of Chances*.

Neden kesin bir yaklaşıma ihtiyaç duyuyoruz? İşte 2010 yılında yayımlanmış ve bütün p değerlerinden kurtulmamız ve Bayes analizini kullanmamız gerektiğini anlatan bir makale: *Demystify Statistical Significance – Time to Move on From the P Value to Bayesian Analysis* (JJ Lee, 2010).

Bu iddianın nedeni ne? Bunu anlamak için tekrar klasik çıkarıma dönelim.

Bir sıfır varsayımımız ve bir de bir alternatif varsayımımız var. Verileri topluyoruz ve sıfır varsayımımız gerçekse gözlenen verinin aşırılığını yansıtan p değerini elde ediyoruz. P değerinin bir sürpriz endeksi olduğunu söyleyebiliriz. Eğer sıfır varsayımında (null hypothesis) beklemediğimiz bir sonuç elde edersek, yani $p < 0.05$ (veya 0.01) ise, sıfır hipotezini reddediyoruz. Bunun nedeni de eğer sıfır hipotezi doğru ise, deney sonrası elde edilen sonuçlar gözlenen sonuçlara göre sadece %5 daha aşırı olacaktır. Günümüzde %95 güven aralığını da hesaplıyoruz. Ancak, beklenen aksine, p değeri bizim elde etmek istediğimizi vermez. Bir sıfır hipotezinin deney sonrası doğru ya da yanlış olma olasılığını açıklamaz. Bize sadece, “Hey! Gördüğüm şey biraz garip, sanki iki tedavi de eşit” der. Ancak, her bir araştırmacının “Benim alternatif hipotezim bu deney sonuçları ile nasıl doğru olabilir” sorusuna açık bir cevap veremez. Buna cevap sadece Bayes yaklaşımıyla bulunur. Aslında, p değeri zor bir kavramdır, tam olarak anlaşılammıştır, genelde kötüye kullanılmaktadır. Uzmanlar bile değeri hakkında ortak bir fikre sahip değildir. Örneğin az sayıda denekle yapılmış bir çalışmaya ait anlamlı bir sonucun ($p < 0.05$) değeriyle çok sayıda denekte yapılmış bir çalışmanın anlamlı p değerinin yorumu hâlâ tartışılmaktadır. Son olarak belirtiyim, klasik yaklaşımla yapılan analizlerde eski deneylerden elde edilen bilgi, Bayes analizinde kullanıldığı gibi, rahatça kullanılmaz.

Dediklerim farklı bir açıdan bakarsak istatistik çıkarımla tıbbi çıkarımları karşılaştırmak. Tıbbi çıkarım tıp fakültelerinde öğretilir. Örneğin, bir kişi A hastalığına yakalandığında, bu hastalıkta gözlenen semptomlar, B hastalığına daha özgü semptomlarla örtüşebilir ve aynı durum hastalık C için de geçerlidir. Uygulamada böyle semptomlar gösteren hastanın hastalığına tanı koymakta tıbbi çıkarım (endüksiyon) kullanılır. Klasik yaklaşımda bir dedüktif yapı vardır: Bir sıfır varsayımı kurulur, varsayım ışığında hesaplamalar yapılır ve olasılık elde edilir. Son adım ise sıfır varsayımını reddetmek veya reddetmemektir ve bu endüktif-

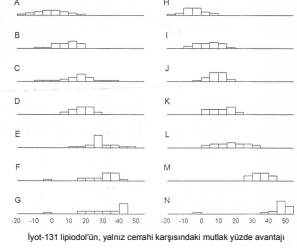
tir. Bayes yaklaşımında ise baştan itibaren endüktif yaklaşım hakimdir ve elde edilen verilerin hesaplanmasıyla varsayım doğrulanır veya yanlışlanır.

Peki, tüm bunlardan ne öğrendik? Klasik sıklıkçı yaklaşım dedüktif argümana dayalıdır. Çıkarımın %95 doğru olması isteniyor ve anlamlılık düzeyi 0.05 olarak kullanılıyor. Klasik yaklaşımın, kısmen endüktif Fisher yaklaşımı (sıfır hipotezi ve p değerinin mucidi) ile saf dedüktif Neyman ve Pearson yaklaşımının (alternatif hipotez ve tip I ve II hatanın mucitleri) bir alaşımı olduğu unutulmamalı. Diğer yandan, Bayes yaklaşımı sadece endüktif ve araştırmacıya istediğini veriyor. Yaklaşım aynı zamanda çıkarımdaki ön bilgiyi de ekliyor. Benim kişisel görüşüm bilim adamlarının doğal Bayesçi oldukları yönündedir.

Eğer Bayes yaklaşımı bu kadar albenili ise, popüler olması neden 250 yıl aldı? Aslında Bayes yaklaşımının yeşerip patlama yapmasını engelleyen bazı ciddi hesap sorunları ve görüş farklılıkları vardı. Bayes teoreminin paydasında bulunan integralin hesaplanmasındaki zorluklar yüzyıllarca Bayes teoremini gerçek problemlere uygulamanın önünde büyük engel oluşturdu. Gerçekten de çok yüksek boyutlu bir integralin numerik olarak değerlendirilmesi zor ve hatta bazen olanaksızdır. İkinci bir engel Bayes yaklaşımına düşmanca bakıştı. Fisher ve diğer pek çok önemli istatistikçi Bayes yaklaşımına karşıydılar. Böyle düşünenler bir yandan istatistik analizlere sübjektif elemanlar katılmasını istemiyorlar, diğer yandan ve daha da önemli, Bayes yöntemine bir bilinmeyen öğesinin katılamayacağı onları rahatsız ediyordu. Peki, Bayes yaklaşımı nasıl popüler hale geldi? 1990 yılında, örneklemeye dayalı yeni bir hesaplama yaklaşımı istatistik dünyasına girdi. MCMC (Markov Chain Monte Carlo) olarak adlandırılan bu yaklaşım integrali değerlendirme sorununu baypas etmemizi sağladı. Ayrıca, WinBUGS yazılımının geliştirilmesi Bayes yaklaşımını son derece yaygınlaştırdı. Yazılım Cambridge’de Dr. Spiegelhalter kılavuzluğunda geliştirilmiş olup paketin web sitesinde, 30.000’den fazla kullanıcısı olduğu ve bunun aslında daha fazla olması gerektiği belirtilmekte. Bu paket Bayes yaklaşımına müthiş bir ivme sağladı.

1. (Göreceli) bilgi kullanımı

• Örnek 1: Elde edilmiş önsel inanışlar/dağılımlar



14. Slayt

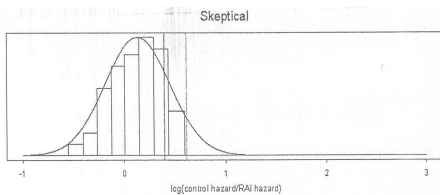
İşte bazı örnekler; 2003 yılında bitirilen bir faz I randomize kontrollü çalışmada, hepatosellüler karsinomunun cerrahi rezeksiyonu ile cerrahi rezeksiyon artı radyoterapi karşılaştırılmış. Bu çalışma 1999 yılında, ara analizdeki kombinasyon kolunun önemli ölçüde başarılı sonuçlar vermesinden sonra durdurulmuştu. Araştırmacıların elde ettiği $p < 0.01$ değeri daha önce tanımlanmış olan durdurma kriterinden düşüktü, ancak ara analiz 30 hastaya dayalı. Bu sonuç sürpriz olduğundan, karşılaştırmanın yeni bir klinik çalışma ile tekrarlanması isteniyordu. Aslında, 30 hastadan elde edilen sonuçlara inanmayan doktorlara bu şüphecilik yeni bir çalışmayı tetikleyen bir sonsal sağlamış.

Daha sonra 2002 yılında 300 hastayla yeni bir çalışma başlatıldı. Bu yeni çalışmada, çalışanlar küçük bir Bayes deneyi yapmış ve yeni çalışmada 14 klinik araştırmacının çalışma öncesi görüşleri alınmış. Amaç, ara ve nihai çalışma aşamalarında, olası Bayes analizi için kullanılacak olan önsel dağılımı elde etmek. Araştırmacıların önselliği bulmak amacıyla bir anket doldurmaları istenmiş, bu anket tek başına cerrahiyle kıyasladığımızda kombinasyon kolunun 2 yıllık nüksüz sağkalım için sağladığı olası avantajla ilgili. Daha ayrıntılı bakarsak; anketi dolduranlar bu kümelerden birine 100 dolar yatırıyorlar! Eğer araştırmacı bir kutuya 100 dolar koyarsa kombinasyon kolunun tek başına cerrahiden %50 olasılıkla iyi olduğu konusunda emin demektir. Öte yandan araştırmacılar genelde, bir çalışmaya başladıklarında çalışmanın sonucunun ne olacağından emin değildirlir. Buna ekipoz (equipose) denir. Yani, yeni tedavinin avantajlarının dezavantajlarına/risklerine üstün gelip gelemeyeceğini bilemezler. O nedenle yanıtlarını farklı kutular arasında dağıtırlar.

14. slaytta araştırmacıların çalışma öncesi ifade edilen görüşleri var. Burada hangilerinin kombinasyon koluna görece olumlu, hangilerinin görece olumsuz baktığını görebiliriz. En olumsuz görüşleri olan 7 araştırmacının görüşleri 'kuşku- lu önselliği' temsil eden ortalama, bir ortalama histogram olarak birleştirildi (15. slayt). Yöntem açısından çok önemli olmasa da, normal bir dağılıma uyarlandı. Ardından bu kuşku lu önsellik

1. (Göreceli) bilgi kullanımı

• Örnek 1: Kuşku lu önsellik



15. Slayt

erken sonlandırılan ilk çalışmanın ara sonuçları ile birleştirildi ve çalışmanın kombinasyon kolunda olumlu bir sonuç elde etmenin sonsal olasılığı ve bu olasılık 0.49 olarak bulundu. Hikâyenin özeti şudur: Ara sonuçlarla kombine edilen Bayes yaklaşımı ile araştırmacıların bir çalışmayı erken kesmesi ve başka bir çalışma başlatması engellenebilir. İstatistik literatüründe, ara analizde kullanılmak üzere gerek kuşkucu fakat aynı zamanda daha iyimser yöntemli modeller de oluşturulmuştur.

Bayes yöntemlerinin, ara sonuç analizleri dışında, klinik çalışmaların analizlerinde kullanımları hâlâ oldukça zor. Ara analizler iki sınıfa ayrılabilir. İlki ara etkinlik veya güvenlik analizidir. Bu tür analizle, çalışma daha planlanan toplam hasta sayısına ulaşmadan deneysel tedavinin kontrolden daha iyi (etkinlik) olması ya da daha fazla yan etkiye yol açabilmesi (güvenlik) olasılığı nedeniyle çalışmanın durdurulup durdurulmayacağı kontrol ediliyor. İkinci tür analizde ise, deneysel tedavinin pozitif sonuç verme şansı kontrol ediliyor. Buna nafilelik (futility) analizi diyoruz.

Klasik yaklaşımda ara analize yönelik klasik yaklaşımların amacı tekrarlayan istatistik test uygulamayı düzeltmek amacıyla yapılıyor. Şimdi bir tekrarlayan istatistik test örneği görelim. İki ortalama veya iki oran arasındaki standart farka eşit olan Z değeri çalışmaya katılan hasta sayısı arttıkça değişiyor. Eğer iki tedavi birbirinden farklı değilse (hipotez vakada olduğu gibi), Z değeri 0 etrafında dalgalanma gösteriyor. Arada bir de ‘istatistiksel anlamlılık’ dediğimiz – 2 ile 2 arasındaki sınırı aşıyor. Değerler – 2 ile 2 sınırının dışına zıplamalar yapmakta, bu da ara analizde tedavinin yanlılıkla daha iyi olduğu sonucuna varmamıza yol açabilir. Bu hatayı düzeltmek için, klasik yöntemde “Sıralı Grup Yaklaşımları” (Group Sequential Approaches) denen bir yaklaşım vardır. Bu yaklaşım tekrarlayan istatistik analizleri dinamik olarak düzeltmeye çalışır. Yapılan her ara ve son analizde p değerinin istatistiksel olarak anlamlı nitelendirilmesi için gerekli anlamlılık düzeyini ayarlamaya odaklanır.

Peki, Bayes dünyasında hangi yöntemler kullanılmakta? Bir örnekte akciğer hacminin

stent ve sözde (sham) stent ile bronkoskopik azaltılması karşılaştırmasının yapıldığı bir çalışma. 225, 270 ve üzeri hasta kayda alınarak adaptif Bayes yaklaşımı kullanılmış. Her bir ara analizde, deneysel kolun kontrol kolundan daha iyi olma sonsal olasılığı hesaplanmış. Eğer bu sonsal olasılık 0.965’den yüksekse, o zaman araştırmacılar anlamlı etkinliğe erişildi diye çalışma sonlandırılabilmiş ve deneysel kolun sözde stent kolundan üstün olduğu sonucuna varılmış. Diğer bir örnek 3 ara ve bir son analiz planlanan bir klinik çalışma örneği görülmekte. Her bir analizde p (deneysel kontrolden daha iyi | o ana kadar veri), ip ($\Delta > 0$ | veri) hesaplanmış. İlk ara analiz, olasılık p ($\Delta > 0$ | veri) 0.5 ve 0.965’den yüksek değil. Δ için %95 güvenilir ara 0’ı içeriyor. İkinci ara analizde (ek hasta kaydından sonra yapılmış), bu sonsal olasılık 0.67’ye çıkıyor, ancak hâlâ çalışmayı durdurabilecek düzeyde değil. Üçüncü ara analizde ise, sonsal olasılık 0.965’i aşmış ve %95 güvenilir ara da 0’ın üzerinde. Hatta eğer son analiz yapılmış olsaydı, varsayılan sonsal olasılık 1 olacaktı.

Bayes ara izlemi ile ilgili başka bir örneği inceleyelim. Bir faz 2 neoadjuvan kanser çalışması, 164 hasta alımı planlanmış, hastalar aktif ilaçla birlikte ya da tek başına bir kemoterapiye randomize edilmiş. Çalışmanın ana son noktası, tümörün patolojik incelemesinde tam yanıt olarak kabul edilmiş. Çalışmaya giriş beklenenden yavaş olmuş, ayda ortalama 1.5 hasta kayıt edilmiş. Otuz dört hastadan veri elde edildiğinde, DMC (data monitoring committee-veri izleme komitesi) tarafından sonuçlar değerlendirilmiş ve muhtemel sonucun ne olacağı sorusu kendilerine sorulmuş. Sonsal öngörü olasılıklarıyla hesaplama yapılmış. İlk ara analizin sonuçları şöyle; 16 kontrol hastada 4 pozitif sonuç ve deneysel tedavi alan 18 hastada 12 pozitif sonuç. DMC bu bilgiyi elde edilen sonuçlar ve PPD ile elde edilen olası gelecek sonuçlarla kombine ederek, tüm hastalar dâhil edilmiş olsaydı sonuç ne olurdu sorusunu cevaplamaya çalışmış. Hesaplamalara göre, çalışmanın deneysel kolda olumlu sonuç verme şansının %95 olması üzerine DMC çalışmanın durdurulmasını önermiş ve çalışma durdurulmuş.

Söz konusu çalışmayla ilgili 16. slaytta kontrollere (mavi histogram) ve deneysel kola (kırmızı histogram) ait kurgusal veriler var. Kurgusal veriler ilk ara analizde toplanmış gerçek verilerle alınan yanıtlarla birleştirilmiş. pCR (probability of complete response) tam yanıt alma olasılığı demek. Toplamda 10.000 kurgusal örnek simüle ettik (2 x 2 ki-kare tabloları) ve her bir tedavi koluna 82'şer hasta ayırdık. Sağ tarafta, soldan elde edilen tüm 2 x 2 tabloları üzerinde hesaplanan klasik ki kare testlerinin oluşturduğu istatistik p değeri histogramları görülüyor. Kırmızı dikey çizginin solundaki sonuçların tümü anlamlı (0.05'de). Bu sonuçlar simüle durumların yaklaşık %95'ini oluşturuyor ve bu nedenle DMC çalışmayı artık durdurma önerisinde bulundu.

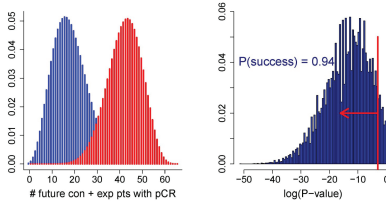
Bayes ara analiz tasarımları genelde Bayes adaptif tasarımlarının bir örneğidir. Faz 1 onkoloji çalışmalarında, maksimum tolere edilebilir dozu bulmaya yönelik Bayes yaklaşımı, ilaç endüstrisinde, tıp çalışmalarında hâlâ kullandığı eski 3+3 kuralından sonra en popüler yaklaşımdır. Bayes yaklaşımı 70'lerde kullanıma girmiş ve doz sınırlayıcı toksisite ile kullanılan kanser ilacının dozu arasındaki ilişkiyi tanımlayan istatistik modele dayanır.

Faz 1 ve faz 2 klinik çalışmalarda çalışmanın bir kolunu durdurmak, bir kol eklemek veya ilaç dozunu değiştirebilmek için adaptif karar kuralları önerilmiştir. Daha önce belirtildiği gibi, bir adaptif tasarım tipi de, süperiorite, noninferiorite ve yararsızlık nedeni ile çalışmayı durdurma imkânı veren Bayes ara-analizi yaklaşımıdır. Bir diğer karmaşık tip de faz 1'den 2'ye kolayca geçmek istendiği zaman kullanılan yaklaşımdır. Örneğin, 10-20 farklı ilacın (veya ilaç dozunun) tedavi adayı olduğu düşünülebilir. Söz konusu sofistike adaptif tasarım, çok sayıda ilaçtan vazgeçip sadece umut verenle devam etmeye yönelik bir ara analizdir. Klasik bir ilaç çalışmasında da genel anlamlılık düzeyinin 0.05 seviyesinde olması hâlâ geçerli bir kural.

Bayes adaptif tasarımda, tüm sonuçlar sonsal ve/veya sonsal öngörü dağılımlarına dayalıdır. Örnek olarak adaptif tasarımlı, komplike, bir tıbbi

2. Ara Bayes takibi

- Örnek 2: Elde edilmiş sonuçlara göre kurgusal gelecek çalışmalar üretilmesi



16. Slayt

4. Tarihsel veri kullanımı

• Örnek 1:
Karar tablosu

Table II. Confidence in dismissing a tocilizumab-type adverse effect on CV events* (assuming 16 events, $pT < 1.25^* pC$)

16 total events (based on 1.1% 24 wk-event rate)		Confidence ($pT < 1.25^* pC$)
Anacetrapib (pT)	Placebo (pC)	pT<pC
0	16	1
1	15	0.999
2	14	0.998
3	13	0.994
4	12	0.982
5	11	0.958
6	10	0.913
7	9	0.838
8	8	0.730
9	7	0.595
10	6	0.447
11	5	0.305
12	4	0.187
13	3	0.102
14	2	0.048
15	1	0.020
16	0	0.007

The green confidence values of $>20\%$ suggest study continuation. The red confidence values of $<20\%$ suggest study termination.
*Prespecified CV events include CV death, nonfatal MI, stroke, and hospitalization for unstable angina.

17. Slayt

cihaz çalışmasını inceleyelim. Çalışma yeni bir kalp pili tekniğinin etkisini bir kontrol grubuyla kıyaslayacak şekilde tasarlanmış. Başlangıçta bu cihazın etkinliğinin ne olması hakkında bilgi yok. O nedenle kaç kişi çalışılacağı, çalışma başladıktan sonra öngörü olasılıkları ile saptanmış ve etkinlik, yararsızlık ve zarar için durdurma kararı sonsal olasılıklarla planlanmış. Her ne kadar bir Bayes adaptif yaklaşımı olsa da FDA, çalışmacılardan çalışmanın operasyonel özelliklerini kontrol etmelerini istemiş. Yani, çalışmacılar simülasyonlar yardımıyla tasarımın genelde (çalışmaların %95’inde) istenen sonucu verdiğini göstermek zorunda bırakılmış.

Bayes yaklaşımında geçmişte yapılan çalışmalardan elde edilen bilgiler de kullanılıyor. Erasmus Tıp Merkezinde, 10 yıldır bazı yeni deneysel tedavileri, temelde aynı tür bir kontrol grubu ile karşılaştıran, AML çalışmaları yapan bir grup var. Her seferinde istenen sayıda hastayı çalışmaya almak yerine “çalışılacak hasta sayısını azaltmanın bir yolu var mı?” sorusu gündeme geliyor. Olası çözüm Bayes yaklaşımında ve bir önceki çalışmadan elde edilen önsel bilgileri çalışmaya açıkça dâhil etmek. Şimdi açıklayacağım çalışmada böyle yapıldı. Araştırmacılar ara analizde önceki çalışmadan elde ettikleri önsel bilgiyi kullanmışlar.

17. slaytta bir tablo görülüyor. Bu tablo DMC tarafından, deney kolunda AE (adverse event = yan etki) sayısının çok yüksek olması nedeni ile araştırmacılara bir uyarı yapılmasının gerekip gerekmediği kararını verebilmek üzere kullanılmış. Mevcut çalışma kesin bir sonuca erişecek kadar geniş çaplı olmadığından, önceki çalışmanın sonuçları şimdiki ile birleştirilmiş. İki çalışmadan elde edilen verilerin aynı olmasa da benzer tasarımlarla elde edildiği kabul edilmiş.

Son olarak, kan basıncı kontrol programının maliyet-etkinliğini değerlendirmede Markov Chain Monte Carlo (MCMC) örnekleme Bayes hesaplama tekniğinin kullanıldığı bir örnek vermek istiyorum: Cost-effectiveness of Community-Based Strategies for Blood Pressure Control in a Low – Income Developing Country. Bizce, Bayes yaklaşımı bu bağlamda ideal, çünkü WinBUGS gibi standart

Bayes yaklaşımları ile yapılabilen hesaplamalarda tanım üzere olan belirsizliği bu modellere güzel bir şekilde ekleyebilir.

Peki, randomize klinik çalışmalarda bu yöntemler ne kadar popüler? Bayes'in kullanımı hızla yaygınlaşıyor. Örneğin son 20 yıldır, bilimsel makalelerin ve Bayes yöntemleri ile ilgili kitapların sayısında muhteşem bir artış oldu. Pek çok Bayes kuruluştaki var ve tıp da dâhil olmak üzere neredeyse bütün bilimsel disiplinlere nüfuz etmişler. Bence hiçbir araştırmacı Bayes yaklaşımını kullanmaktan kaçınmaz.

Hâlâ şüphesi olanlar için: Neden randomize Bayes kontrol çalışması yaklaşımını seçmelisiniz? Analize tarihsel ve önsel bilgileri dâhil etmenizden dolayı, daha küçük hasta sayılarıyla çalışabilirsiniz. Daha anlaşılır analizler elde edebilirsiniz, çünkü sonsal olasılık ve sonsal öngörü olasılığı p değerinden daha kolay anlaşılabilirler. Ayrıca Bayes yaklaşımıyla her problem ele alınabilir. Aslında Bayes yaklaşımının bu kadar popüler olmasının nedeni klasik sıklıkçı yaklaşımların izin verebileceğinden çok daha kompleks modellemelere izin vermesi.

Peki, Bayes yaklaşımının eksikleri neler? Dendiği gibi, Bayes yaklaşımı MCMC yaklaşımı yardımı ile daha esnek modellemeye izin veriyor. Bir cevap almak için bazen haftalarca beklemek gerekse bile, bu yöntemle kompleks problemler de çözülebiliyor. O halde, bu kadar olumlu yanlarına rağmen hâlâ bu şüphelilik neden? Bir sorun, önsel bilgileri kullanma gerekliliği ve bunun da öznelliğe yol açması. Temelde Bayes yaklaşımında farklı çalışmalarda performansa dikkat edilmese de, genelde (çalışmaların %95'inde) doğru yanıtı vereceği kuralına riayet edilmeli. İşte bu nedenle, Bayes yaklaşımı sıklıkçı özellikler için de test edilmeli. Örneğin, farklı (hipotez) çalışmalarda tekrarlandığında, %95 güvenilir aranın gerçekten de gerçek değerlerin %95'ini kapsayıp kapsamadığı kontrol edilmeli. Bu nedenle FDA araştırmacıları, Bayes yaklaşımıyla simulasyon yaptıklarında bu yaklaşımın sıklıkçı yaklaşıma da uygun olduğunun gösterilmesini istemektedir.

Bayes yaklaşımının FDA örneği düzenleyici kurumlar, otoriteler tarafından faz 3 çalışmalarda

kullanımının kabulü biraz zaman alabilir, çünkü öznellik içeren istatistik analizler kuşku uyandırabiliyor.

Güher Direskeneli: Yaklaşık 50 sorum var, ilk üçü ile başlayacağım! Ara analiz ile ilgili olarak, bunun etkinlik için değil, sadece güvenlik için kullanıldığını düşünmüştüm. Bence Bayesçi yaklaşım bu fikre uymaz. İkinci ilgili bir soru da şu; çoklu testteki tesadüfi farkı görme olasılığının üstesinden nasıl geldiğinizi hâlâ anlayamıyorum.

Emmanuel Lesaffre: İlk soru güvenlik ile ilgili, ben güvenlik ya da etkinlik için test yapmanın bir farkını görmüyorum, bu sadece başka bir yönden bakmakla alakalı.

Güher Direskeneli: Peki, normalde gücü ölçerken belirli bir alfa düzeyinde etkinliği göstermek için ölçüm yapmıyor muyuz ve pek çok hastaya erişmeniz gerekiyor mu?

Emmanuel Lesaffre: Bayes yaklaşımında numune boyutunu hesaplamak zor, çünkü Bayes yaklaşımı ancak sonuçların alınması ile birlikte bilinebilen sonsal olasılığa dayalı. Diğer sorunuz tesadüfi bulma ile ilgili. Bence de bu bir problem, ve bu yüzden FDA önerilen Bayes tasarımının operasyonel karakteristiklerini göstermek için simülasyon egzersizleri yapılmasını istiyor. O halde, bunun sıklıkçı yaklaşımla farkı ne sorusu akla gelebilir. Fark bence p değerine göre daha anlaşılır sonsal olasılık araçları ve MCMC hesaplarının hesaplama çerçevesinde gizli.

Güher Direskeneli: Metaanalizde Bayes yöntemi nasıl kullandığınız hakkında birkaç kelime söyleyebilir misiniz?

Emmanuel Lesaffre: Metaanaliz bir çeşit Bayes yaklaşımıdır, ancak Bayes yaklaşımı çok daha esnektir. Çünkü asemptomatik argümanlara dayalı değildir. Ayrıca Bayes, metaanalizi ek tarihi bilgi ile kaynaştırılabilir.

Güher Direskeneli: Bu tür bir metaanalizde, çalışmaları düzenlediğiniz şekilde sonuçları değiştirebilir misiniz?

Emmanuel Lesaffre: Klasik bir metaanalizde, her bir çalışmayı boyutuna göre tartabilirsiniz, ancak bu tartıyı bir Bayes analizinde önsel dağılımı değiştirerek yapabilirsiniz.

Hasan Yazıcı: Benim temel sorum Güher'in meta-analiz hakkındaki sorusu ile ilgili. Eğer randomize kontrollü çalışmaların Bayes yaklaşımı ile bir metaanalizi yapılırsa, önsel olasılıkları önemsemeyen ya da normalleştiren yöntemlerin olması ve bu yöntemlerin sonsal olasılığa odaklanması ve metaanalizin bunun üzerinde hesaplanması mümkün müdür? Mekanik olarak bu nasıl yürür? Birileri, eğer önsel olasılığı görmezden gelirse ve sonsal olasılığın ilk değerini alırsa, bu nasıl halledilir? Romatolojide, hiç Bayes çalışmalarının bir metaanaliziyle karşılaştınız mı?

Yusuf Yazıcı: Bence Bayesçi ağ analizi, Bayesçi metaanalizden farklı.

Emmanuel Lesaffre: Ama aynı zamanda Bayesçi ağ metaanalizi de var. Kısaca, ağ metaanalizi şu demek oluyor: Sizin A, B ve C tedavileriniz var. A ve C'nin karşılaştırılması ile ilgilisiniz ancak elinizde A ile B'yi ve B ile C'yi karşılaştıran çalışmalarınız var. İşte bu ağ metaanalizle bu iki çalışmayı birleştirerek A ve C'yi karşılaştırabilirsiniz. Bu karşılaştırma hem klasik hem de Bayesçi yaklaşımla gerçekleştirilebilir.

Yusuf Yazıcı: Veri analizi ile ilgili verdiğiniz örneklerin, eğer yanılmıyorsa, Bayesçilerden bahsederken, dikotomik sonuçları vardı. Tıpkı "Eğer bir şey bundan iyiyse, bir şeyin bundan daha üstün olma ihtimali vs. vs.". Eğer doğru anladıysam, bizler genellikle önsel olasılıkları ve sonsal olasılıkları normal dağılımlar olarak varsayıyoruz. Gösterdiğiniz örnekler tıpkı böyleydi.

Emmanuel Lesaffre: Kısaca Bayes yaklaşımı herhangi bir veri ile sınırlı değildir. Kolaylık olsun diye bu örnekleri kullandım.

Yusuf Yazıcı: Bilgilendirici olmayan önsellikler hakkında bir sorum var. Eğer yanılmıyorsa, bilgilendirici olmayan önselliklerle yaptığımız şey,

belirsizliklerimizi daraltmak için önceden topladığımız önsel verileri kullanmak.

Emmanuel Lesaffre: Bayes yaklaşımında açık veri kullanma zorunluluğu yoktur. Bilgilendirici olmayan önsel kullanılabilir. Önsel veri yoksa bile, Bayes yaklaşımı kullanılabilir. Klasik sonuçlar tekrarlanırsa da, ilgilendiğiniz parametreler hakkında bilgi veren sonsal dağılım Bayes paradigmasına uymayı ilginç yapar.

Yusuf Yazıcı: Belki sorumu değiştirebilirim. Örneğin, belirli bir dizi veriden elde ettiğimiz olasılık dağılımımız olduğunda, verilerin bu şekilde dağıldığı hususunda belirsizlik yaşıyoruz. Bu bir şekilde kendi başına önsel bir bilgi, çünkü benim bir bilgim var ve bilginin sınırlar içinde dağılım göstermesi olası. Şimdi kendi bilgilerimi yeniden düşünebileceğimi bilerek, hâlâ verilerimi yeniden analiz edebilir miyim ve bulduğum veriler aslında önsel dağılım mıdır?

Emmanuel Lesaffre: Önsellik için verilerinizi asla kullanamazsınız, çünkü bu bir çatışma olur.

Yusuf Yazıcı: Tamam. Zamanınızı alıyorsam üzgünüm ama son bir soru. Artık her yerde Bayes yöntemlerinin kullanımını görüyoruz. Google reklamları ve arama tercihlerinin hepsi Bayes yaklaşımları ile belirleniyor. Araştırma yapmak için yeni bir yöntem var, sanırım "CAS" (Bilgisayar Destekli Araştırma) olarak adlandırılıyor, burada bir takım sorular ayarlayabiliyorsunuz ve siz cevap verdikçe, sistem sizin cevap vereceğiniz yeni sorular seçiyor. Hem CAS'ta hem de hesaplamalı bilimde ilgimi çeken tek şey, siz sisteme belirli bir dizi veri girmeye başladıkça, genellikle sizi gittiğiniz yere yönlendirmesi. Belki de orası sizin gitmek istediğiniz son yer değil, bu Bayes analizinde problem olabilir mi? Ya önselliklerim yanlışsa?

Emmanuel Lesaffre: Bu iyi bir soru. Eğer önsellikleriniz yanlışsa, o zaman sizin yanlış bir analiziniz olduğu açıktır. Eğer yeterli veriniz yoksa yanlış cevap alırsınız. Bayes önerisi nedir? Her zaman duyarlılık analizi yapmalısınız, farklı önsellikleri birleştirmelisiniz ve sonuçlarınızın nasıl değiştiğini görmelisiniz, eğer çok fazla

değişiyorlarsa, bu bütün çıkarımlarınızın önsel bilgilerinize dayandığı ve verilerinizin çok katkı sağlamadığı anlamına gelir. Genellikle veri diziniz nispeten büyüktür, hatta öncelliğiniz yanlış

olsa bile, oldukça istikrarlıdır, ancak en önemli şey şu; Bayes yaklaşımı bir sonraki aşamada ve daha sonrasında bilgiyi kullanmanız için size yardımcı olur.

NADİR GÖRÜLEN HASTALIKLARDA KLİNİK ÇALIŞMA TASARIMLARI

Brian Feldman

*Çocuk ve İç Hastalıkları Bölümü
Toronto Üniversitesi
Kanada*

Hepimiz görece nadir hastalıkları olanları tedavi etmeye çalışıyoruz, ancak ben bir pediatrik romatoloğum, bu yüzden, çok nadir hastalığı olan çocuklara bakıyorum. Pediatrik romatolojide en sık görülen hastalık olan juvenil idiyopatik artrit'in prevalansı 1/1000'dir, oysa duyduğumuz gibi erişkinde romatoid artrit'in sıklığı %1 ile %0.07 arasında değişir. Romatolojide nadir görülen bazı hastalıklar için elimizde çok fazla kanıt bulunmuyor. Burada kalp hastalığı ve kanıta dayalı tıp konularında Medline'de yaptığım bir taramayı belirteceğim. Kalp hastalıkları alanında tedaviyi yönlendirmek için yapılmış 24.000 randomize kontrollü çalışma ve ayrıca 3000 metaanaliz var ve bilindiği gibi metaanalizler kanıta dayalı tıbbın yapı taşlarıdır. Oysa çocukluk çağı aritri için yapılmış 102 randomize kontrollü çalışma var ve bunların sadece 21'i metaanaliz. Yine Medline'da görülen 102 sitasyonun çoğu randomize çalışmalar değil. Bütün bunlar, nadir bir hastalıkla uğraşıyorsak, elimizde görece az, tıp uygulamamıza sağlam kanıt oluşturabilecek kaynak bulunduğunu gösteriyor.

Kanıttan bahsederken, kanıtın niteliği ve randomize kontrollü çalışmalarla ilgili zorluklardan da bahsetmek gerek. Doğru düzgün randomize kontrollü çalışmaları gerçekleştirmenin zor olduğu haller için bir çözüm bulmamız gerektiği kanısındayım. Bu amaçla bazı esnek çözümler geliştirilebilir ve bu çözümler bizlere hastalıklarımızın tedavisi için bazı kanıtlar sağlayabilir.

Öncelikle bu çalışmalarımızı nasıl kabul edilebilir hale getireceğimize değineyim. Randomize Plasebo Fazı Tasarımı olarak adlandırdığımız şeyi tanımlamaya çalışacağım. Bu kapsamda, az sayıda olgu temelinde nasıl daha fazla bilgi toplayabileceğimizden bahsedeceğim. Son olarak da tıbbi kayıtlarda mevcut verileri tedavi için kesin kanıt oluşturmak üzere nasıl kullanabileceğimizi anlatacağım.

İşe size bir soru yönelterek başlayayım. Üniversitede kimler felsefe okudu? Ben okudum ve epistemoloji dersini çok severdim. Epistemoloji bilginin felsefe teorisi, bildiklerimizi nasıl bildiğimizi öğretir. Diğer bir deyişle bir bilgi teorisi, bilimi, yöntemi

mi ya da alt yapısıdır. 1500'lü yıllara kadar, epistemolojideki baskın paradigma bir tür tümdengelim (dedüksiyon) yöntemi idi. Aristo dedüktif yöntemi öne süren ve ilkelerini ilk savunan kişi. Verilen öncülden erkeklerin kadınlardan sayıca daha fazla diş sahibi olduğunu ilk ilkelerle çıkarmış, erkekler daha güçlüdür, daha fazla yerler vb.

Sir Francis Bacon'a gelene kadar herkes buna inanıyordu. Bacon tümdengelim yönteminden memnun değildi ve bunun yerine tümevarım (indüksiyon) öneriyordu. Tümevarım yönteminde, belirli bir deneyim ya da gözlem ile başladığımız süreci mantıksallaştırır, daha sonra bunlardan varsayımlar üretir ve son olarak da bu varsayımları sınarız. Bilimde yöntem budur. Örneğin erkeklerin ve kadınların aslında aynı sayıda diş olduğu bir gerçektir, sadece Aristo bunun bir önemi olmadığını düşünüyordu.

Bilim ve tıp tarihinde tıbbi tedavi bağlamında aslında Pierre Louis'e çok şey borçluyuz. Pierre Louis 1800'lerin başında yaşamış bir Fransız doktor. Bence tüm zamanların en ünlü doktoru olmaliydi, ancak pek çok insan onun kim olduğunu bilmez. Louis sayısal yöntem denen bir şey önermişti. Bu sayısal yöntem aynen şöyle çalışıyordu; siz hastanızı kendi yönteminizle tedavi ediyorsunuz, ben de kendi yöntemimle tedavi ediyorum, daha sonra bir araya gelip kaç kişinin öldüğüne bakıyor ve hangi yöntemin daha iyi olduğunu bulmaya çalışıyoruz. 1800'lü yılların başına kadar, bu tamamen sapkın bir yöntem olarak kabul edilirdi, her hastanın çok farklı ve kendine özgü olduğu düşünülürdü, onları aynı kefeye koyup sayamazdınız. 1800'lerin başında hastalıkların patofizyolojisi hakkında pek bir şey bilinmiyor, hala filojiston ve hümorlerden bahsediliyordu. Lober pnömoninin en iyi tedavisinin sülükle kan alınması olduğu düşünülürdü ve bu dünyanın her yerinde böyleydi. Ancak Pierre Louis bundan hoşnut değildi, bunun yerine "Bekle-Gör Yöntemi"ni önermişti. Aslında hastalığın kendi kendine iyileşmesinin beklenmesini, hastalara hacamat yapılmamasını ve sadece hastalığın kendi sürecinde seyretmesini öneriyordu. Kendi önerdiği sayısal yöntemle bir deney yaptı ve bekle-gör yöntemi ile tedavi edilen hastaların mortalitesi, sülükle tedavi

edilenlere kıyasla daha düşüktü. Bulguları önce Fransızca, bir yıl sonra da Amerikan Tıp Dergisi'nde yayımlandı. İzleyen 10 yıl içinde tüm dünyada tıbbi sülük satışları durdu. Pierre Louis, hastaların tıbbi tedavisinde tümevarım yöntemini kullanmanın önemini kanıtlamıştı ve aslında o günden beri tek yaptığımız Pierre Louis'in başarısını geliştirmek yolunda çaba göstermek.

Kanıtı neden ihtiyacımız var? Bence buradaki herkes belli ki bu konuyla ilgileniyor ve bu çabamızın güzel bir örneği pediatrik onkolojide var. 1960-70'lerde, ben büyürken, lösemili bir çocuk idiyse, hastalık %100 kesinlikle ölümcüldü. Eğer kansersensiz, öleceğinizi bilirdiniz. Bugün ise lösemi gibi çocukluk çağı kanserleri %90 oranında tedavi edilebilmekte. Amerika Birleşik Devletleri'nde (ABD) yılda 100 klinik çalışma yürütülüyor, kanser olan çocukların %90'ı tedavi ediliyor. Bu, dünyanın her yerindeki hastanelerde, her çocuğun ve tedavi ekibinin, dünya çapında tedavilerin araştırıldığı en son çalışmalara tamamen erişebilmesi anlamına geliyor. Son olarak, şunun gerçek olduğuna inanıyorum; çocuk onkologlarının yaptığı araştırmalar sayesinde 40 yıl önce çocukluk çağı kanserleri neredeyse tedavisz bir hastalık olarak tanımlanırken, bugün tedavi oranı en az %78. Bu, Pierre Louis'in başlattığı ve böyle çalışmaların da önünü açtığını gösteren bir kanıt.

Yine de bir sorun var. Bence çoğumuz tedavi çalışmalarında randomize kontrollü çalışmanın altın standart olduğuna inanıyoruz. Bunun sebebi elbette ki deney ve kontrol kolunda yapılan rastgele örneklemenin, daha iyi cevap vereceğini önceden bildiğimiz hastaları alıp bilerek ya da bilmeyerek deneysel tedaviye dâhil etmemizi önlemesi. Körleme değerlendirmeye de aynı şekilde deneysel tedavi alan hastaları, bilerek ya da bilmeyerek, daha iyi yanıt veriyormuş gibi değerlendirmemizi engellemekte, istatistik analiz ve çalışmaya katılacak hasta sayısı hesapları da şansa bağlı hata olasılığını titizlikle kontrol etmemizi sağlamakta.

Öte yandan, randomize kontrollü çalışmalar fazla kabul görmez ve sayıları öyle pek hızla artmaz. Kalitatif ve kantitatif literatürde bu konuda

hayli çalışma var. Ancak doktorlar hastalarını böyle klinik deneylere pek kolay katmazlar. Burada bir ikilem vardır. Doktor için öncelik bilimden ziyade karşısındaki hastadır. Hastalarını böyle çalışmalara dâhil etmelerinin tek nedeni, onların en iyi ve yeni tedavileri denemesini ve faydalanmasını sağlamaktır. Ancak sadece hastalarının plasebo grubuna düşmemelerini dilerler. Hastalar açısından da aynı şey geçerlidir; klinik çalışmalara girmelerinin tek sebebi en iyi tedavi yönteminin kendilerine uygulanması şansını kullanmaktır, bilime yardımcı olma arzusu değil. Biliyoruz ki aydınlatılmış onam sürecinde; gençseniz, iyi bilgilendirildiyse ve çalışmayı iyice anladıysanız çalışmaya katılma ihtimalinizin o derece düşük olduğu anlamına gelir. Çalışmalara katılan hastaların pek çoğu böyle çalışmalara doktorları istediği için katılmışlardır.

Sonuç olarak, nadir görülen hastalıklarda doğru tedaviyi saptamada elimizde fazla bilimsel kanıtımız olamaz. Söz konusu bu nadir hastalığa yakalanmış az sayıda hastanız varsa, onlar arasında randomize kontrollü bir çalışma yapamaz veya ancak bilimsel kalitesi düşük çalışmalar yapabilirsiniz. Doktora öğrencilerimden biri, çocukluk çağı artriti üzerine yapılmış randomize kontrollü çalışmalara bakmıştı. Geçerli ve sağlam bir cevap için gerekli olan randomizasyon gizliliği, seçim sırası ve uygun maskeleme yöntemlerine ne kadar uyulduğuna baktığımızda, her ne kadar yıllar içinde bir düzleme gözledikse de tasarımı yeterli ve titiz çalışmaların az sayıda olduğunu anladık. Örneğin, taranan çalışmaların ancak %37'sinde deneklerin hangi gruba girdikleri düzgün bir şekilde maskeleyebilmişti. Hâlbuki bu maskelenme randomize kontrollü çalışmanın çok önemli bir unsuru.

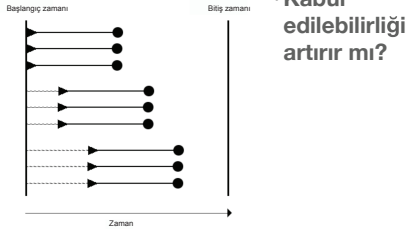
Öyleyse bir sorunumuz var. Peki, potansiyel çözümler neler? Hastalar için daha kabul edilebilir çalışmaları nasıl oluşturabileceğimizden başlayalım. Genelde çalışmalarımızın kalitesini olgu sayısını artırarak ya da her bir olgudan daha fazla bilgi elde ederek yükseltebiliriz. Protokollerimizin kabul edilebilirliğini geliştirip insanların çalışmalara katılmaya istekli olmasını sağlayabiliriz. Böylece çalışmadaki olgu sayısını artırabiliriz. Diğer bir yol ise çok merkezli çalışmaları

artırmak. Ancak bu konu üzerinde fazla durmak istemiyorum. Eminim ki benim sahip olduğum bütün deneyime sizler de sahipsiniz. Çok merkezli çalışmaların önünde Avrupa Birliği, Kuzey Amerika Konsensüsü gibi bir sürü bürokratik engel olduğunu hepimiz biliyoruz.

Kabul edilebilirlikten söz etmek gerekirse, plasebo alan hasta sayısının mümkün olabildiğince az veya plasebo uygulama süresinin mümkün olabildiğince kısa tutulmasıyla klinik araştırma protokolünün kabul edilebilirliğini arttırabileceğinizi öneren çok sayıda kalitatif çalışma var. Burada 'randomize plasebo faz tasarımı' (RPFT) olarak adlandırdığımız yöntemden bahsedeceğim. Burada araştırma soruları biraz farklı. Geleneksel paralel gruplu randomize kontrollü çalışma esas olarak şu soruya yanıt arar: Çalışmanın sonunda deneysel ilacı alan grubun sonuçları mı, yoksa plasebo veya standart tedaviyi alan grubun sonuçları mı daha iyi? RPFT'de sorulan soru ise şudur: Deneysel ilacı erken alan hastalar, ilacı daha geç alan hastalara göre ilaca daha erken yanıt veriyor mu? Yöntemden beklenen, kabul edilebilirliğin artması. Bunu biraz daha açacağım.

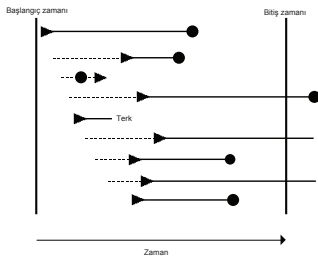
1. slaytta RPFT'nin ana özelliklerini şematik olarak görüyorsunuz. Solda görülen dikey çizgi çalışmanın başını, sağda görülen ise sonunu gösteriyor. Alttaki yatay çizgi ise zaman. Her bir yatay çizgi bir hastayı, çizgilerin üzerindeki üçgenler tedaviye başlama, yuvarlaklar ise tedaviye yanıtın ne zaman olduğunu gösteriyor. Birinci dikey çizgi ile üçgen arasında kalan zaman dilimi ise hastanın plasebo veya geleneksel tedavi aldığı zaman dilimini belirtiyor. Tabii uygulamada araştırıcı hangi hastanın etkili olup olmadığını sınıadığımız yeni ilacı ne zaman almaya başladığını bilmiyor ve yine tabii, deney ilacının herhangi bir hastada ne zaman başlanacağı randomizasyon ile belirleniyor. Üçünün biraz daha sonra katıldığını ve diğer üçüne de ilacın en son verildiğini görüyoruz. Tabii ki gerçek bir çalışmada okların dağılımı böyle olmaz. 2. slaytta görüleceği gibi bazı oklar kısa, bazıları uzun olur, bazılarında ise çalışma süresi içinde hiçbir yuvarlak nokta (yani tedaviye yanıt) izlenmez.

Randomize plasebo faz tasarımı



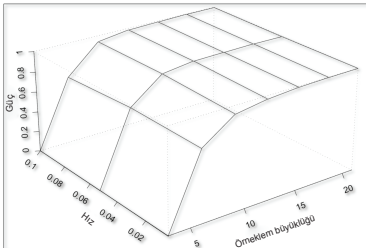
1. Slayt

Gerçek bir deneysel çalışmada olaylar çeşitli şekillerde meydana gelir



2. Slayt

Randomize plasebo faz tasarımı (RPFT) için güç yüzeyi (power surface)



3. Slayt

Tasarımın çalışıp çalışmadığını sınamak amacıyla Monte Carlo simülasyonları yaptık. Bu terim nereden geliyor? 1930'ların sonunda atom bombası geliştirilip geliştirilmeyeceğinin olasılığını sınamak sürecinde çok sayıda deney yapmak gerekiyordu. Ancak bu deneylerin hepsini gerçekleştirmek çok zor olduğundan, çeşitli olasılıkları matematik yöntemlerle irdelemek gerekmiş ve bu gizli projenin adı Monte Carlo simülasyonları olmuş.

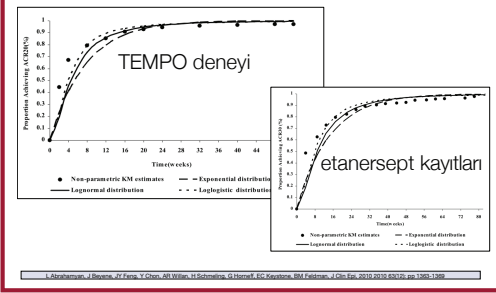
Biz ise şöyle bir şey yaptık. Yeni bir ilacın denendiği ve geleneksel tedavi veya plaseboya üstünlüğü kanıtlanmış çalışmaları aldık. Bu yayımlanmış çalışmada elde edilen verilere, aralarında RPFT'nin de olduğu çeşitli kontrollü klinik çalışma yöntemleri uygulandığında, söz konusu yöntemlerin çalışmanın bilinen sonucunu (yani sınanan ilacın, plasebo veya geleneksel tedaviye üstünlüğünü) ortaya çıkarma gücünü inceledik. Özetle gördük ki RPFT gerçekten de çok güçlü bir çalışma tasarımı.

Size burada gösterdiğim şey (3. slayt), dikey eksen gücü temsil ediyor ve tabii ki gücün yüksek, yani 0.8'in üstünde olmasını istiyoruz. Yatay eksenlerin biri yanıt hızı (rate) diğeri ise denek sayısı. Gördüğümüz gibi, oldukça küçük sayılar olsa da bu tasarımla farklı tedavi yanıtlarına rağmen gücü yüksek göstermek teorik olarak mümkün. Simülasyon yapmak için, tedaviye verdiğimiz yanıtın aslında üstel dağılım olarak adlandırılan basit bir dağılımı izlediği varsayımını kabul ettik. Bunun doğru olup olmadığını test etmek için, geri döndük ve bazı verilere baktık. Çoğunuz TEMPO (Etanarsept/Metotreksat) çalışmasını bilirsiniz. Simülasyonlarımızda bir bunu kullandık, bir de pediatrik artrit için Alman etanarsept veri tabanını kullandık.

Burada gördüğünüz (4. slayt), ACR20 ve pediatrik ACR30'u birlikte başaran hasta oranı ve hafta olarak zaman.

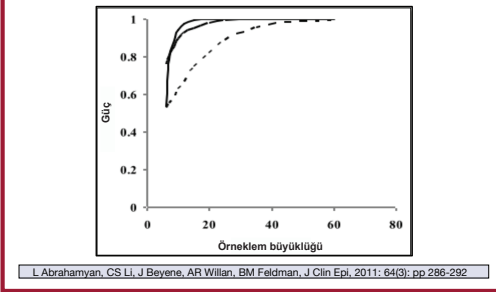
Gerçek yanıt oranları noktalarla temsil edilmekte. Uzun noktalı çizgiler üstel dağılımı göstermekte ve aslında çok iyi bir uyum gözlenmemekte. Log-normal, log lojistik ya da Weibull

Belki üstel olmayabilir



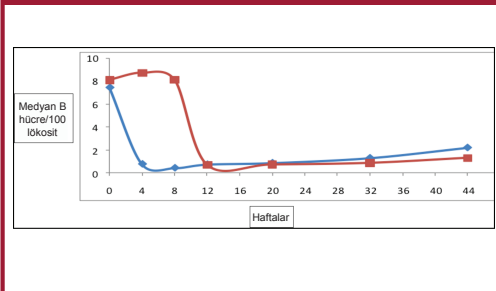
4. Slayt

RPFT ve RKD Weibull dağılımına benzerlik gösteriyor



5. Slayt

44 hafta boyunca B-hücre sayısı



6. Slayt

dağılımları gibi daha karmaşık dağılımlara dönüş-türdüğümüzde, asıl tedavi yanıtını gösteren dağılımı elde etmeye daha yakın olabiliyoruz. Peki, bu neden önemli? Geleneksel paralel grup randomize kontrollü çalışmanın, her ne kadar hastalar çalışma boyunca plasebo alacakları için çalışmaya çok katılmak istemeseler de, randomize plasebo faz tasarımından daha güçlü olacağı açıktır. Aslında, eğer üstel (exponential) bir dağılım kullanırsak (5. slayt), size burada gösterdiğim gibi, örneklem büyüklüğünün yatay ekseninde ve gücün de dikey ekseninde olmasından dolayı, bu paralel grup randomize kontrollü çalışmasıdır ve üstel dağılımdan çok daha güçlü olacaktır.

Eğer daha gerçekçi bir dağılım kullanırsak RPFT ile elde edeceğimiz sonuçlar geleneksel yöntemle elde edilen kadar güçlü olacaktır. Bu teoride iyi bir şey, ancak nasıl işe yarıyor? Bu tasarımı kullanan sadece iki romatoloji çalışması var. İlki bu yıl yayımlandı, miyozit için rituksimab çalışması veya kısaca RIM çalışması, pek çoğunuz bu çalışmaya aşina olmalı. Refrakter miyoziti olan hastalarda plasebo kontrollü çalışma yapmak imkânsız gibi görülebilir. Bu nedenle biz bir RPFT çalışması önerdik. Hastalar başlangıçta randomize edilerek 2 kez 750 ml/m² rituksimab veya plasebo infüzyonlar başlanmak üzere seçildi. Dokuz hafta sonra plasebolu hastalara rituksimab ve çalışmanın körlüğünü bozmamak amacıyla da başlangıçta rituksimab alanlara da plasebo verildi. Yani hiçbir hasta 9 haftadan fazla bir süre plaseboda kalmadı. Beklenti rituksimaba erken başlamanın erken bir yanıt oluşturmaya olanak vermesiydi.

6. slaytta B-hücre sayıları ve hafta sayısı var, mavi ile gösterilen rituksimabı en başta alan grupta B-hücre sayısı oldukça hızlı düşmüş. Kırmızı eğri de rituksimabı 9 hafta sonra alan grup. Rituksimab hakkında çok şey bildiğimiz için bu sürpriz bir durum değil, istatistik olarak oldukça anlamlı.

7. slaytta tüm çalışma grubu ve 3 alt grup için Kaplan Meier eğrileri görülüyor. X eksenini zaman, y eksenini ise remisyon kriterlerini doldurmamış hastaların yüzdesi. Bu grafiklere bakarak hastaların büyük bir kısmının zaman içinde iyileştiğini söylemek mümkün. Ancak RPFT açısından temel

şoru olan hangi eğride (Rituksimab gri eğri) iyileşme daha önce, söylemek mümkün değil. Bunun iki anlamı var, eğer fark ettiyseniz, hastaların %80'den fazlası bir yılın sonunda yanıt vermiş. O halde, ya ağır hastalığı olan ve progresyon nedeni ile rituksimabın bir etkisi olmaksızın kendi kendine düzelen hastaları seçmişiz ya da rituksimabın etkili olmasına rağmen etkisi beklenenden daha geç çıkmış. Aslında, sonraki romatoid artrit deneyimleri, ikinci seçeneğin muhtemelen doğru olduğunu ve mevcut klinik çalışmalara göre romatoid artritli hastalarda iyileşme için 3-6 ay kadar beklenmesi gerektiğini göstermektedir. Ne var ki RIM çalışmasını planlarken bunu bilmiyorduk.

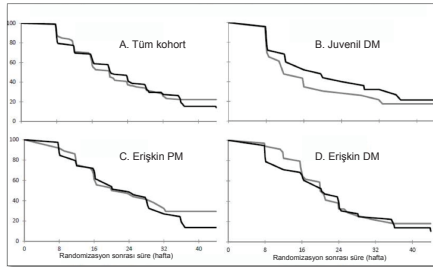
İkinci bir çalışmada ise sistemik başlangıçlı juvenil artrit tedavisinde bir interlökin-1 blokörü olan rilonacep RPFT yöntemiyle çalışılmış ve sonuçlar klinik olarak çok anlamlı çıkmıştır. Ancak ACR kongresinde sunulacağı için halen bu çalışma hakkında daha fazla bilgi veremiyorum. Yukarıdaki örneklerde görüldüğü gibi RPFT yöntemi ile çalışmaya katılmak isteyen hasta sayısı görece daha fazla olduğundan başarılı ilaç çalışmaları yapılabilir. RPFT yönteminde bunu sağlayan, böyle çalışmalarda herhangi bir hastaya uzun süreli placebo vermek gereksiniminin ortadan kalkmasıdır.

Şimdi biraz da hasta sayısı az olduğunda daha fazla bilgi nasıl toplayabiliriz, onu görelim. Burada n=1 çalışmalarından ve Bayes analizinden bahsedeceğim.

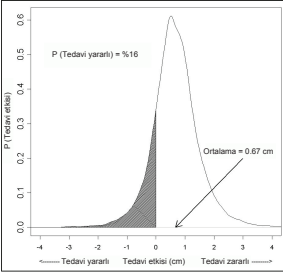
Çocuklar arasında dikkat eksikliği hiperaktivite bozukluğu (DEHB) sık görülür. Çocuk doktorları bu hastaların sürekli stimulan almasını isterler ancak bu öneri anne ve babalar tarafından kuşkuyla karşılanır.

Diyelim ki, Jenny bir DEHB hastasıdır ve onun tedavisinde sitümulanların işe yarayıp yaramadığını araştıracağız. Bu amaçla Jenny'ye belirli zaman dilimleri için placebo, belirli zaman dilimleri için de stimulan ilaç veririz. Ancak bu zaman dilimlerinin sırası rastgele örneklemeye yapılır ve bu devreler sırasında çocuğun dikkat eksikliği ve/veya hiperaktivitesinde bir değişiklik olup olmadığı ailesi ve/veya öğretmenlerinden öğrenilir.

Tüm kohort ve bireysel miyozit alt-grupları için DOI'siz sağkalım



7. Slayt

Ağrı azalmadı**8. Slayt**

Sonunda stimulan ilacın alındığı periyotlarda davranışların ve okul performansının plasebo alınan periyotlara kıyasla daha iyi olup olmadığına bakılır. Bu bize yalnızca tek bir kişi için tedavinin işe yarayıp yaramadığını kesin bir şekilde verir.

Ancak bütün bu gözlediğimiz sadece tek bir birey -Jenney- için geçerlidir. Peki, bizim Jenny’de gözlediğimiz, Jenny ile aynı hastalığı olanlarda da geçerli olur mu? Diğer yönden biz bu bilgiye olabildiğince az sayıda hasta üzerinde deney yaparak nasıl ulaşabiliriz? İşte burada Bayes analizine ihtiyaç vardır.

Bayes analizinin nasıl çalıştığını size hatırlatayım. Diyelim ki romatoid artritte daha önce çalışılmamış küçük moleküllü yeni bir ilaç olan Feldmacitanib ile çalışıyoruz. Bu çalışmadan önce, Feldmacitanibin nasıl çalıştığı hakkında en ufak bir fikrimiz yok, her hasta cevap verebilir, hiçbir hasta cevap vermeyebilir, insanların yarısı cevap verebilir. Yanıt veren bir hasta ile çalıştıktan sonra artık hiç kimsenin cevap vermemesi imkânsız, çünkü birisi vermiş, ancak hâlâ bu aralıkta yaygın olarak dağılmış bir olasılık var. Dokuz hasta üzerinde çalıştığımızı farz edelim, tedavi 8’inde başarılı ve 1’inde başarısız. Hiç kimsenin yanıt vermemesi olasılığı imkânsız, çünkü 8 kişi yanıt vermiş, herkesin yanıt vermesi olasılığı da imkânsız, çünkü 1 kişi yanıt vermemiş, işte bu durumsal (conditional) olasılık. Ancak, çok yüksek olasılıkla hastaların ancak %30-40’ından daha azı bu tedaviye yanıtızsız kalacak. Yani, hassasiyetle çalışılan az sayıdaki olgu ile bu tür Bayes analizi kullanarak çok fazla bilgi elde edebiliriz.

Bu amaçla üç küçük klinik çalışma yaptık. Biri çocukluk çağındaki artritlerde ağrı kesici olarak amitriptilin kullanımına aitti. Bir onkologla beraber yaptığımız ikinci çalışmada kemoterapinin neden olduğu mukoziti önlemek için E vitaminini deniyorduk. Üçüncü çalışma ise kemoterapiye bağlı kusmaları engellemek için metopimazin kullanımına yönelikti.

Artritleri oldukça iyi bir şekilde kontrol altında olan ufak bir grup hastamız vardı. Ancak aynı hastalar hâlâ ağrılardan yakınıyorlardı. Hastaların ağrısının fibromiyaljiye ait olabileceğini düşündük

ve bu amaçla bir $n=1$ amitriptilin çalışması düzenledik. Hastalara 6 hafta boyunca amitriptilin verildi ve detaylı bir ağrı günlüğü dolduruldu. Daha sonra, 6 haftalık bir plasebodan sonra, 3 dönem amitriptilin ve 3 dönem plasebo uygulandı. Çalışma sadece 6 hasta üzerinde ve randomize olarak yapıldı.

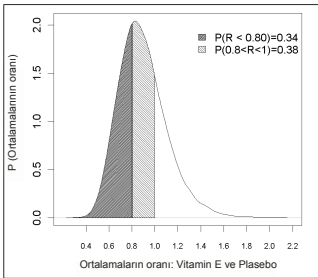
8. slayttaki diyagram Emmanuel Lesaffre'nin önceden gösterdiği diyagramlara çok benziyor. Bu olasılığa göre orantılandı, dolayısıyla eğrinin altındaki alan 1'e eşit. Sıfırın sağında görülen oranların anlamı, tedavinin aslında ağrıyı daha kötü yaptığı. Sıfırın solundakiler de tedavinin aslında ağrıyı daha iyi yaptığını düşündürüyor, bu 10 sonometrelik bir ağrı skalasında sonometre cinsinden ortalama tedavi etkisidir. Burada gördüğümüz şey durumsal olasılığın 0.7 sonometre kadar yükseldiği, yani amitriptilinin aslında ağrıyı biraz kötüleştirdiği ki bu 10 sonometrelik skalada klinik olarak önemli. Tedavinin yararlı olma şansı ise sadece %16. Bu sonuçla, artık daha fazla olgu üzerinde çalışmamız gerekmediği sonucuna vardık, 6 olgu bunun yararlı bir tedavi olmadığını ve bunu bırakmamız gerektiğini bize işaret etti.

Bir sonraki çalışma, kemik kanseri için kemoterapi alan çocuklarla ilgiliydi, güçlü ama bir o kadar da toksik kemoterapötik ajanlarla çocuklarda ağır bir mukozit ortaya çıkmıştı. Bu durum çocukların beslenmelerine engel oluyor, morallerini de bozuyordu. Bu durumda onkolog da ilaçlarını azaltmak zorunda kalıyor, ama bu kez hastaların esas derdi olan kanserlerinin tedavisi düzgün olarak yapılamıyordu.

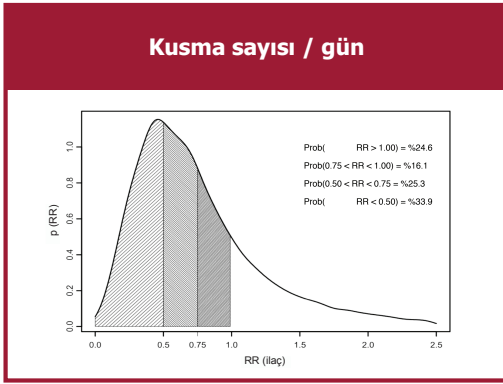
Çalışmamızın varsayımı, topikal E vitamini kullanımının çocuklarda kemoterapi sırasında görülen ağız mukozası iltihaplanmasını önleyecekti. Bu amaçla ilk iş olarak E vitamini gibi yağlı bir maddeyi kullanarak ağızda çiğnenecek ve sonra da tükürülecek iyi bir plasebo hazırladık. Planımız kural olarak sikluslar halinde uygulanan kemoterapi sırasında, bazı sikluslarda E vitamini ve bazı sikluslarda da plasebo vermektir. Çalışmanın ana çıktısı ise ağızdaki ülser sayısını esas alan bir dental skorlamaydı.

Tüm bu klinik çalışma sadece 16 olguyla yayımlandı. 9. slaytta sonuçları görüyorsunuz.

16 kişiden sonra



9. Slayt



Eğride görülenler ortalama mukozit skorları dikey eksen ve gerçek iyileşme oranları da diğer eksen. X ekseninde 1'in sağındaki her şey mukozitin kötüye gittiğini, solundaki her şey işlerin iyiye gittiğini gösteriyor. Gördüğünüz gibi tam olarak burada 3-1 olasılığı (odds) var, eğer tamamen zararsız bir madde olan E vitamini alırsanız biraz daha iyi olursunuz ki bu da iyi bir şey. Demek E vitamini mukoziti bir miktar önüyor. Ancak bu yazının ilk yazarı başka bir sonuca vardı. Ona göre E vitaminin plaseboya karşı daha iyi olduğunu söyleyebilmek için en azından %20'lik bir iyileşme görülmesi gerekti. Hâlbuki bu çalışmanın sonuçlarına göre bu büyüklükte bir fark için sadece %34'lük bir olasılık var. Bu nedenle ona göre E vitaminini bırakıp başka, daha etkili ajanlar bulmalıydık.

Üçüncü araştırmamız beyin kanserli çocuklarla yapılmıştı. Herkesin bildiği gibi kemoterapiye bağlı olarak mide bulantısı ve kusma oldukça sık görülür. Çoğu hasta kusmayı azaltmak amacıyla deksametazon ve ondansetron kullanmaktadır. Fakat beyin kanseri varsa deksametazon kullanamazsın, çünkü kan beyin bariyerine etki ederek kemoterapinin beyne geçişine izin vermemektedir. Deksametazonun yerine alternatif olarak Avrupa'da metopimazin isimli bir bileşik bulunmuştu. Bu bileşik Kuzey Amerika'da artık üretilmiyordu ve zaten Avrupa'daki üretici firma da ilacın üretimini durdurmuş ve başka şirket üretim haklarını satın almıştı. Son derece karmaşık bir dizi müzakere sonrasında ilaç bizlere, Kanada'ya gönderildi ve yetkili makamlar da kullanma izni verdiler. İlacın intravenöz (iv) ve oral şekilleri vardı. Çocuklarda kolay kullanımını sağlamak amacıyla oral kullanımı uygun bulduk. Çalışma protokolü bir önceki ile aynıydı; farklı kemoterapi sikluslarında hastalar ya metopimazin ya da plasebo alacak şekilde randomize edildiler ve ilacın kusmayı kontrol edip etmediğine bakıldı.

10. slaytta çocuklarda günlük kusma sayısı, dikey eksendeki olasılık ve yatayda görece risk görülüyor. 1'in üstündeki her şey işlerin kötüye, altındakiler de her şeyin iyiye gittiğini gösterir. Yine, plasebo yerine metopimazin alınması halinde daha iyi olma olasılığı (odds) 3-1'di, bu başarılı

olarak değerlendirilebilirdi. Eğer intravenöz formunu kullanabilseydik çocuklar daha az tükürecek ve tedavi muhtemelen daha başarılı olacaktı.

Bu örnekler birkaç olgudan nasıl oldukça fazla bilgi alabileceğimizi ve bu bilginin popülasyon bazında tüm hastalara uygulanabileceği hakkında ip uçları vermektedir. Çok nadir görülen nörosarkoidozda bile bu yöntemlerle yeni bir tedavi çalışması yapabilirsiniz.

Son olarak, randomize bir çalışma yapmanın oldukça zor olduğu bu tür durumlarda tedavi tahminleri ile mevcut verileri nasıl karşı karşıya getirip kullanabileceğimizi açıklayayım ve bu konuda birkaç örnek sunayım.

Yale Üniversitesinden Alvan Feinstein, klinik epidemiyolojinin kurucu üç babasından biri olarak kabul edilir. Kendisini çeşitli yerlerde dinledim ve ısrarla söylediği şeylerden bir tanesi randomize kontrollü çalışmaların tıp bilimine fazla katkılarının olmadığıydı. Bu çalışmaların oldukça pahalı olduğunu, yürütülmesinin zor olduğunu ve çalışılan hastaların niteliklerinin gerçek hayatta olan biteni yansıtmadığını söylerdi. Onu dinleyenlere sürekli olarak daha iyi gözlemsel çalışma yöntemleri geliştirmelerini öğütlerdi. Ona göre tedavi etkisinin kesin ve tarafsız tahminlerini yapmak mümkün olacak ve randomize çalışmalara gerek kalmayacaktı. Belki günümüzde böyle yöntemler bulunmuştur.

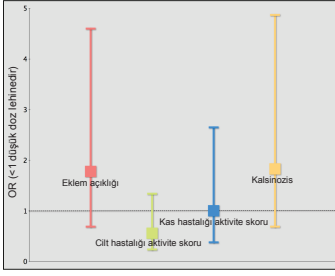
Gözlemsel çalışmalarda ana sorun endikasyonun doğurduğu karıştırıcılıktır (confounding by indication). Bununla ne demek istiyorum? Feldsinanibi romatoid artrit için kullanıp etkilerini metotreksat ya da etanarsept gibi bilinen ilaçlar ile kıyaslamak istediğimde ilk hatırlamam gereken, neden ilk başta bir kısım hastaya metotreksat yerine feldsinanib başladığımı hatırlamak olmalı. Hastaların durumu ya daha ağırdır ya da daha fazla komorbiditeleri vardır. Özetle onlara özel bir şey, onları en başından diğerlerinden farklı kılmıştır. Genelde bu sebep tedavinin kendisinden çok sonuçla alakalıdır. Elmaları portakallarla kıyaslamış oluruz ve buna endikasyonun doğurduğu karıştırıcılık denir. Ancak bunu engellemek için, geniş veri tabanlarında uygulanan oldukça basit çözümler vardır. Bu çö-

zümlerden biri eğilim skorudur (propensity score). Önce hastayı farklı kılabilcek her bilgi, hastalıklarının şiddeti, demografik bilgileri, önceki ilaç kullanımları gibi sonuçları önceden belirleyebilecek bütün bilgileri bir araya getiririz ve feldsinanib ile ilgisi olup olmadığına bakarız. Diskriminant fonksiyon analizi ya da lojistik regresyon gibi oldukça basit yöntemler kullanarak feldsinanib ile metotreksat karşılaştıran verilere dayanarak bir olasılık skoru elde ederiz. Eğer feldsinanib alma olasılığı aynı olan ancak biri feldsinanib alıp diğeri almayan 2 hastam olsa, bu tedavi kararı aslında rastgele bir karar olur. Aynı eğilimi, yani aynı olasılık skoru olan hastaları eşleştirdiğimizde, elde ettiğimiz sonuçların randomize kontrollü çalışma ile aynı derecede yansız olduğu gösterilmiştir. Bunu nasıl uyguladığımız konusunda size bir örnek vereyim. Burada çok geniş bir veri tabanı yerine sayısı görece sınırlı bir veri tabanında gerçekleştirdiğimiz bir çalışmadan söz edeceğim.

Çocukluk çağı miyozitinde elimizde ABD ve Kanada Ulusal Sağlık Enstitüsü destekli 400 hastanın kaydı bulunmakta idi ve bu hastaların 139'unun 3 yıllık ayrıntılı verileri mevcuttu. Çalışmamızın ana amacı, bu hastalara ilk tanı konduğunda glukokortikoid uygulamasının ne kadar agresif olduğunu saptamaktır. Güncel uygulamada bazı doktorlar günde 1 gram yüksek doz (pulse) metilprednisolon kullanıyor, bazıları ise günde 1-2 mg/kg prednison öneriyorlardı. Bu grupta, ilk başta yüksek doz metilprednisolon alanlarla oral prednison alan hastaları karşılaştırsaydık, yüksek doz metilprednisolon alan hastaların durumunun ağırlaşmaya eğilimli olduğunu görmek şaşırtıcı olmayacaktı. Ancak eğilim skorunu kullanarak, ağırlık derecesi eşit 42 tane çok iyi eşleşen çift bulabildik ve 3 yıl içinde bu hastalara neler olduğunu inceledik. Sonuçlar 11. slaytta görülüyor.

Burada üçüncü yıldaki sonuçlar yer alıyor. %95 güven aralığı sınırlarından anlaşılacağı gibi eklem hareket açıklığı, deri hastalığı aktivitesi, kas hastalığı aktivitesi ve kalsinöz açısından iki grup arasında 3 yıl sonunda anlamlı fark yoktu. Eğilim skoru analizini uyguladıktan sonra dahi bunların hiçbirinde istatistiksel bir farklılık bulunmadı. Düşük doz prednison grubunun lehine açık bir

Daha agresif kortikosteroid uygulaması daha iyi sonuç vermedi



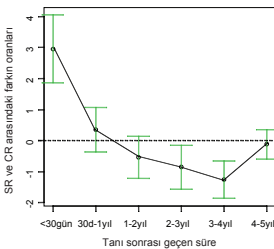
11. Slayt

endikasyona göre karıştırıcılık vardı ve sonuçta bunlar daha düşük seviyede ağırlaştı.

Dermatomiyoitte bu sefer intravenöz immunglobulin (IVIG) kullanımını aynı şekilde incelemek istedik. Ancak bu sefer işimiz çok daha zordu. IVIG uygulamasını kesin olarak daha ağır hastalara yapıyorduk. O nedenle incelemek istediğimiz hastalarda IVIG uygulaması herhangi bir zamanda başlamış olabilirdi. Burada analiz için başvurduğumuz yöntem zaman içinde ilerleyen bir eğilim skoruydu. Buna “marjinal yapısal model” denir, grupları ağırlık derecelerine göre dengeli hale getirir ve endikasyona göre karıştırıcılık olmaksızın bize grupları karşılaştırma olanağı verir.

Sonuçlar 12. slaytta görülüyor. Anlaşıyor ki zaman içinde steroide dirençli hastalar (SR, bunlar IVIG de kullanıyor) ve kontrol hastaları (bunlar IVIG almıyor) arasındaki hastalık aktivite skoru farkı hemen aynı derecede azalıyor. Bu hastalara tanı konduktan hemen sonra çalışmamız için uygun eşler bulamadık, gamma globulin alanlar sadece ağır hastalığı olanlardı, ancak zaman geçtikçe almayanlara göre tedavi alan hastalar daha düşük hastalık aktivitesi gösterdiler ve bu durum ilacın etkili olduğunu düşündürdü. 4-5 yılın sonunda, her iki tedavi kolunda yeterli sayıda hastamız kalmamıştı, sadece gamma globulin almaktaydılar. Ancak hiçbir zaman randomize kontrollü çalışma yapılmayacak olan bu alanda, bu çalışma sonunda gamma globulinin hastalık aktivitesini düşürdüğüne dair kanıtlar elde ettik.

IVIG ile hastalık aktivitesi daha düşüktür



Pullenayegum E, Lam C, Manthot C, Feldman BM: Ann Rheum Dis 2011; 2089-94

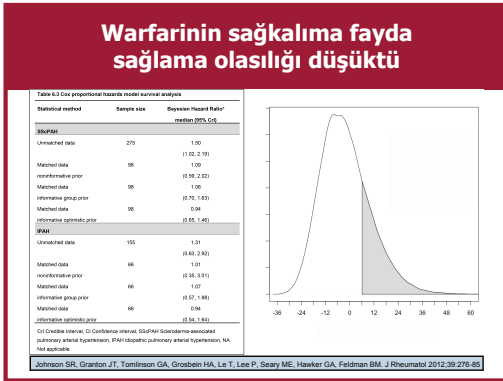
12. Slayt

Size son olarak, sizleri daha fazla ilgilendireceğini düşündüğüm, erişkin hastalar arasında yaptığımız bir çalışmadan bahsetmek istiyorum. Çalışmanın konusu sklerodermalı yetişkin hastalarda pulmoner hipertansiyon çalışması. Bu çalışma, doktorasını benimle birlikte Toronto’da yapan bir yetişkin romatoloğu Sindhu Johnson tarafından, Bayes analizi kuralları içinde yapıldı. Sindhu, sklerodermada pulmoner hipertansiyon tedavisi hakkında uzman görüşlerine dayalı en son gelişmeleri kapsayan bir yöntem bulmak istiyordu. Sklerodermada bazı hastalarda warfarin sıkça kullanılır, bazılarında da kullanılmaz ve şimdilerde daha etkili tedavilerimiz de var. Sindhu, çalışmaya önce uzman görüşle-

rini ortaya çıkarmaya yönelik bir yaklaşımla başladı. Bunu eğilim skorlarını oluşturmak için kullandı ve bir eşleştirme algoritması geliştirdi. Daha sonra birlikte, skleroderma ve pulmoner hipertansiyonu olan birkaç yüz hastanın yer aldığı tablo şeklindeki derlemelerden verileri sentezleyerek Bayes temelli bir sağkalım analizi yaptık.

13. slaytta sonuçları görüyorsunuz. Eğer analizler eğilim skoru ile düzeltilmezse endikasyona göre karıştırıcılık kaçınılmazdır. Bunun nedeni ise açıktır. Günlük uygulamada warfarin genellikle daha ağır hastalara verilir ve böyle hastalar daha kısa yaşarlar. Bunun yanında çalışmamız gösterdi ki warfarin kullanan hastaların bütün düzeltmelerden sonra bile sağkalımı aslında daha kısaydı ve verilerimize göre warfarinin yaşam süresini uzatma olasılığı sadece %24'tü. Diğer bir deyişle hasta warfarin kullanmaya devam ettiği takdirde, 3-1 olasılıkla (odds) prognozda ya değişiklik olmayacak ya da hasta daha kötüye gidecekti. O halde warfarin kullanmak en azından oldukça iyi eşleştirilmiş bu hasta kohortumuzda pulmoner arteriyel hipertansiyon tedavisinde daha pahalı ve etkili ilaçların yerine geçmiyordu. İdiyopatik pulmoner hipertansiyonu olanlarla da aynı analizi yaptık ve aynı sonuçları elde ettik.

Özetle, çocukluk çağı romatizmaları gibi oldukça nadir hastalıklar hakkında elimizdeki bilimsel kanıtlar da oldukça eksiktir. Bu kanıtların niteliğini, tümevarım ve tümdengelim paradigmalarını; özellikle hasta sayımızın az olduğu durumlarda randomize kontrollü çalışmalarda yaşanan problemleri; randomize plasebo faz tasarımını ve kimi durumda yaptığımız çalışmaların nasıl daha kabul edilebilir hale getirebileceğimizi tartıştık; Bayes analizi ve tekrarlanan n=1 çalışmaları kullanarak daha az olgudan nasıl daha fazla bilgi alabileceğimizi konuştuk. Gelecek için bir çözüm de farklı tedavilerin geçerliliğini nasıl arttırabileceğimizi gördük. Bu çalışmalarda hasta dosyalarından nasıl yararlanabileceğimizi gözden geçirdik. Önemli olan, hâlâ altın standart olan randomize kontrollü çalışma yapma olasılığı olmadığı durumlarda elimizde bir araç gereç çantamız olması. Böylelikle tedavi kanıtımızın ve olgu sayımızın çok az olduğu durumlarda da karşılaştığımız sorunları tartışıp irdeleyebiliriz.



13. Slayt

Çok teşekkürler. Soru sormak isteyen var mı?

Yusuf Yazıcı: Propensite skorları bir nevi istatistik oyun veya aldatmaca değil mi? Bunları hesaplar-ken risk faktörü olmayan bazı şeyleri risk faktörü olarak kabul ediyorsun.

Brian Feldman: Burada birkaç sorun var. Bunlardan bir tanesi kullandığın risk faktörü gerçek bir risk faktörü veya karıştırıcı (confounder) olmaya bilir. Bunun yanında verilerimizi etkileyen gerçek karıştırıcılardan da haberimiz olmayabilir. Böyle durumlarda kullanabileceğimiz başka yöntemler de var. Bunlardan benim sevdiğim bir yöntem. ‘Instrumental variable’ (işe yarayan değişken) yöntemi deniyor. Örneğin bir hastalık değişik yörelerde sadece o yöreyle ilgili değişik yöntemlerle tedavi edilmişse verileri değerlendirmekte bu yöntem işe yarıyor. Ancak hakkın var. Gerçek karıştırıcının ne olduğunu bilmiyorsan işler zor.

Bir Katılımcı: Özellikle nadir görülen hastalıklarda ‘withdrawal’ (uygulanan tedaviyi kesmek) çalışmaları kullanmaya ne dersiniz?

Brian Feldman: Bu güzel bir soru. Pediatride kullanılma izni alan birçok ilaç bu yöntemle çalışılmış. Bu etik olarak da önemli. Herkes birden bir ilaca başlayıp, daha sonra randomize bir grupta ilacı durduruyorsun. Bu yöntem ‘enrichment’ (zenginleştirme) yöntemi de deniyor. İlaç gerçekten işliyorsa, randomizasyona sıra geldiğinde daha az sayıda hastayla etkiyi gösterebilmek mümkün. Ancak bu yöntemin önemli bir sorunu gerçek etki büyüklüğü hakkında az fikir veriyor. Bunun nedeni de randomizasyondan evvel herkesin ilaca yanıt vermiş olması. Gerek Avrupa gerekse de Kuzey Amerika’da bu yöntemle kullanma ruhsatı almış çok sayıda çocukluk çağı romatizması ilaçları var.

Bir Katılımcı: Bu yöntemle yan etkilerin anlaşılması zor değil mi? Tanım üzere sadece ilaca devam edenler randomize oluyor.

Brian Feldman: Burada, yan etkiler açısından daha önemli sorun çalışmanın başında plasebo kolu olmaması. Yoksa, etanerseptte yapıldığı

gibi aktif ilaç alan grubu uzun süre izleyebilirsin. Özetle, tabii ki bu yöntemlerin hiçbirisi klasik randomize, kontrollü ilaç çalışması kadar yararlı bilgi vermez.

Bir Katılımcı: Benim 2 sorum var. Birincisi $n=1$ çalışmalarla ilgili. Böyle çalışmalarda ‘carry over’ (geçmişten gelen) etkiyle nasıl başa çıkıyorsunuz? Bazı ilaçlarla yapılan çalışmalarda bu bana kaçınılmaz gibi geliyor.

Brian Feldman: $n=1$ çalışmalarında bu tabii çok önemli bir konu. Tüm çalışmalar için işe yarayacak sihirli bir formül yok. Örneğin bizim Amitriptilin çalışmasında bu sorun pek yoktu çünkü tedavi değişimlerinde 4 haftalık bir ‘wash-out’ (ilaçtan arınma zamanı) vardı. Ancak, kemik iliği nakli gibi bazı tedavi yöntemlerinde tabii $n=1$ çalışmaları yapılamaz.

Aynı Katılımcı: Benim diğer sorum randomize plasebo faz çalışmalarına ait. Burada plasebo sürelerini nasıl saptıyorsunuz? Hastaların kliniğinize başvurma zamanları çok değişik olacağından bu iş oldukça zor olur gibime geliyor.

Brian Feldman: Bu çalışmaları gerçekleştirebilmek için elinizde oldukça fazla sayıda hasta olması gerek. Bu arada gördük ki planlanan plasebo fazınız ne kadar kısaysa çalışmaya o kadar fazla sayıda hasta almanız gerekiyor. Bunu kestirebilmek, çalışmanın kuvvetini saptayabilmek için simülasyonlar yapmak gerekiyor.

Bir Katılımcı: Böyle çalışmalarda bir ‘run in’ (alıştırma) zamanı kullanmaya ne dersiniz? Yani hastaları çalışmaya almadan, yeni bir ilaca başlamadan açık olarak izlemek.

Brian Feldman: Bence bu çok iyi bir fikir. Çünkü bizim uğraştığımız romatoid artrit örneği hastalıklar alevlenme ve iyileşmelerle seyreden hastalıklar ve hastaya deneysel ilacı bu dalgalanmanın neresinde başladığını değerlendirmede çok önem taşıyor. Eğer böyle bir izlemeyle seyirleri gerçekten kötü olan hastaları çalışmalara alabilirsek çok iyi olur.

* Özet olarak sunulan bu çalışma, orijinal yazı olarak daha sonra *Arthritis and Rheumatology* dergisinde yayımlanmıştır:

Yurdakul S, Mustafa BN, Fresko I, Seyahi E, Yazici H. Brief report: inadequate description and discussion of enrolled patient characteristics and potential inter-study site differences in reports of randomized controlled trials: a systematic survey in six rheumatology journals.

Arthritis Rheumatol. 2014 May;66(5):1395-9. doi: 10.1002/art.38349.

RANDOMİZE KONTROLLÜ ÇALIŞMALARDA HASTA ÖZELLİKLERİ VE MERKEZLER ARASI FARKLILIK*

Sebahattin Yurdakul

Romatoloji Bilim Dalı
İstanbul Üniversitesi
Cerrahpaşa Tıp Fakültesi

Son zamanlarda yaptığımız bir çalışmayı sizlerle paylaşmak istiyorum. Biliyorsunuz gerek ilaç gerekse de ilaç dışı çalışmaların etkinliğini gösteren en iyi metot randomize kontrollü çalışmalardır. Bundan dolayı randomize kontrollü çalışmalar altın standard olarak kabul edilir ve bu çalışmaların sunulmasının da çok iyi, doğru ve şeffaf olması gerekmektedir. Bir grup epidemiyolog, metodolog ve bilimsel dergi editörleri bir araya gelerek 1996'da CONSORT kurallarını (**CONSOLIDATED STANDARDS OF REPORTING TRIALS- Kontrollü ilaç çalışmaları standard bildirimi**) oluşturarak yayımladılar. Bu kurallar daha sonra 2001 ve 2010'da güncelleştirildi. Bu neye yarıyor? Bu kurallar, yazarlara, bir randomize kontrollü çalışmanın başından sonuna kadar nasıl yapılıp ve en iyi şekilde nasıl yazılacağına kılavuzluk ediyor. Ayrıca, CONSORT kuralları, randomize kontrollü çalışmaların, hem internal hem eksternal validitesini ile ilişkilidir. CONSORT kuralları yayımlandıktan sonra birçok dergi bunu kabul edip yazılarını bu şekilde istemeye başladı. Daha sonra da birçok hem romatoloji hem de genel dahiliye dergilerinde yazıların rapor edilme şeklinde kalite artmaya başladı. Ancak birçok araştırma, hâlâ bu kurallara yeterli derecede uyulmadığını ve bunların yeterli miktarda bildirilmediğini göstermektedirler.

Şu anda 25 tane CONSORT kuralı vardır. Bu kurallarından bir tanesi 4A ve 4B, çalışmaya katılan hastaların seçilme özellikleri, diğeri de çalışmanın yapıldığı, bilginin toplandığı yerin, şehrin, ülkenin ve konumunun özellikleriyle ilgilidir. Uzun yıllardır yaptığımız dergi kulüplerinde ve geçen sene de özellikle burada yaptığımız dergi kulübünde de çok merkezli çalışmalarda bu bilgilerin son derece eksik olduğunu fark ettik. Tabii bu bilgilerin eksik olması randomize kontrollü çalışmaların genellendirilebilirliğini veya tatbik edilebilirliğini son derece kısıtlamaktadır. Bu nedenlerle çalışmaya alınan hastaların seçim kriterlerini ve çalışma yerinin özelliklerini, buna ilaveten bir de çok merkezli çalışmalarda merkezler arasında çalışma sonuçları bakımından fark olup olmadığını araştırmak gayesiyle, iyi bilinen 6 tane romatoloji dergisinde 2011 ve 2012 senelerinde yayımlanan randomize kontrollü çalışmaları tek tek, elle taradık. Bu dergiler, *Annals of the Rheumatic Diseases*,

Arthritis and Rheumatism, Arthritis Care Research (Hoboken), Rheumatology (Oxford), The Journal of Rheumatology ve Clinical and Experimental Rheumatology idi. İki ayrı yazar çalışmaları standardize bir formla değerlendirdi. Yazının başlığından sonuna kadar bütün ilgili bilgileri topladılar. Sadece burada söylemek istediğim ise yazarların bağlı bulunduğu merkezler kısmı hariç tutuldu. CONSORT 4A'da hastaları doktor mu gönderdi? Yoksa hasta kendisi mi bu çalışmaya yöneldi: televizyon, gazete, internet, radyo gibi haber yoluyla mı? CONSORT 4B ise çalışmanın ülkesi ve şehri, ve bunun yanında çalışma merkezinin özelliklerini kapsar: bu bir toplum mu, muayenehane mi, hastane mi, yatan hasta mı, ilk basamak, ikinci basamak veya üniversite merkezi mi? Bunun yanında diğer faktörler var, iklim, fiziksel, ekonomi, coğrafya, sosyal ve kültürel, özellikle ırk ve eğitim ile hastaların ulaşım problemlerinin olup olmadığına baktık. Ayrıca, merkezler arasındaki çalışma sonuçlardaki potansiyel farkın bildirilip bildirilmediği kayıt edildi.

İyi bilinen 6 tane romatizma dergisinde 2011 ve 2012 senelerinde yayımlanan randomize kontrollü çalışmaların sayısı toplam 202 olarak bulundu. Ancak bu çalışmalardan ekstansiyon ve ikincil analizler (n=64), bir kaç çalışmanın beraber değerlendirildiği çalışmalar (n=9), kısa raporlar (n=3), mektuplar (n=2) ve faz 1 çalışmaları (n=6) olmak üzere toplam 84 tanesi çalışma dışı bırakıldı ve çalışmaya alınan toplam çalışma sayısı 118 tane kaldı. Şimdi teker teker hastaların nasıl refere edildiğine bakalım: burada bildirilen hastaların 23'ü (%19,5) doktor tarafından gönderiliyor, 3 hasta ise (%2.5) kendisinin çalışmadan haberi olduğu için çalışma içinde ve her ikisi birden refere olanın nasıl olduğunu söyleyen çalışma sayısı 118 çalışmanın içinde oldukça azdır (n=10, %8.5). Bunun yanında çalışma yerinin özelliklerini bildirilenler ise toplumda 3, hastanede polikliniğinde 33, yatan hasta 2, ikincil basamak 2 ve üniversite merkezi 16'dır. Hastaların prognozu ile son derece ilişkili olan, sosyoekonomi seviyesini bildiren ça-

alışma sayısı 11, eğitim seviyesini bildiren 10, ırk bildiren 48 çalışma vardı. Ülke bildiren 75 çalışma varken 1/3'i ise ülke bildirmiyor; şehir ise (n=42) 1/3'inde mevcuttur. Bu çalışmaları içinde çok merkezli çalışmalar 76 tanedir (%64.4) ve bunlar çok merkezli çalışma olmasına rağmen 76 çalışmanın ancak 13 tanesi merkezler arasında bir fark olup olmadığını bildiriyor. Bu konuda, CONSORT hasta seçimi ve merkezler arasındaki etkisini gösteren bu ilk çalışmada, bizim gözlemlerimiz bu bilgilerin oldukça az olarak bildirilmesidir. Tabii bu bilgilerin az olması, çalışmaların genellendirilebilirliğini son derece kısıtlıyor. Yani New York'taki yapılan çalışma sonuçları belki Türkiye'ye uymuyor veya Türkiye'deki bir sonuç Afrika'ya uymuyor demektir. Diğer taraftan çalışmaya alınmanın değişik şekilleri ise çalışmanın eksternal validitesi yanında burada internal validitesini de kısıtlıyor. Sonuç olarak her ne kadar böyle basit görünüyorsa da seçili hastaların referral şekli ve de çalışma merkezinin özelliklerinin bildirilmesi çok önemlidir. Çünkü ne derece ve nasıl bizim hastalarımıza uyuyor, gerçek hayata ne derece uygun bunların gerek yazarlar tarafından, gerek okur tarafından, gerek dergi editörleri tarafından bilinmesi çok önemlidir.

Vedat Hamuryudan: Appendiksleri de dikkate aldık dedin, o ne demek?

Sebahattin Yurdakul: Appendikte çalışan merkezlerin şehir ve ülkeleri varsa onları aldık.

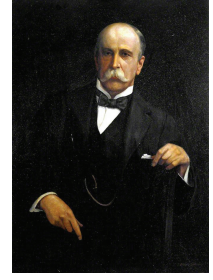
Vedat Hamuryudan: Neredeki appendiks?

Sebahattin Yurdakul: Metin içinde.

Vedat Hamuryudan: Şimdi birçok derginin internet baskısında makaleyi az detayla veriyorlar, ama diyorlar ki daha fazla detay için literatürün içerden bir yere gönderiyor seni, oradan bir sürü detayı alıyorsun. Ona bakmadınız mı?

Sebahattin Yurdakul: Web analizini almadık, biz sadece ana metni aldık.

Olağandışı olana her zaman
önem verin ve kayıt altına
alın... yayımlayın.
Bunları kısa ve öz bir
şekilde kalıcı olarak kayıt
altına alın.
Buna benzer yazılar her
zaman değerlidir.
— Sir William Osler



1. Slayt

ROMATOLOJİ LİTERATÜRÜNDE “OLGU SUNUMU” OLGUSU

İsmail Şimşek

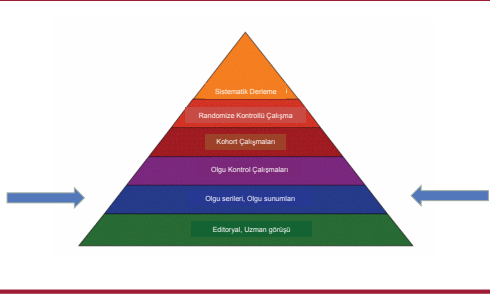
Geçen sene fellow kursunda demek ki bazı ilhamlar alınmış, Sebahattin Hocam kursta aynı CON-SORT’tan bahsetmişti, ben de olgu sunumu nasıl yazılır, diye sözlü sunum yapmıştım. Aslında bu fikir de oradan geldi. O sözlü sunumu yaparken fellowlara olgu sunumu nasıl yazılmalıyla ilgili literatür ararken bir şeyler buldum, şöyle yazılsın, böyle yazılsın ama romatolojide olgu sunumlarının durumu nedir, ne yazılsa daha iyi olur, böyle bir data bulamamıştım.

Elbette ki olgu sunumlarının kurucu babasının William Oster olduğunu biliyoruz (1. slayt). Ancak son yıllarda olgu sunumlarının cazibesini yitirmiş olduğunu görüyoruz. Editörden elbette ki bu şekilde yorumlar almasak da, olgu sunumlarını reject oranı daha fazla gibi. Bunun çeşitli nedenleri var. Baktığımızda olgu sunumları kanıt hiyerarşisinde en son sıralarda yer alıyor (2. slayt), en temel gösterilen nedenlerden biri bu, ama diğer nedenler de var. Genellenebilirliğinin az olması ve tabii dergilerin yoğun olarak randomize kontrollü çalışma yayımlaması nedeniyle bir şekilde yer sorunları da buna etken olarak gösteriliyor.

Buna o zaman bakmıştım hatta 3. slayta baktığımızda romatolojideki bazı dergilerin ki impact faktörü biraz daha yüksek diyebiliriz, hiçbir şe-

*Romatoloji Kliniği
Güven Hastanesi
(Ankara)*

Kanıt hiyerarşisi



2. Slayt

Dergi	Olgu	Editöre mektup	Diğer
Arthrit & Rheum	X	X	Kısa sunum
Rheumatology	X	✓	X
ARD	X	✓	X
J Rheumatol	X	✓	X
Clin Rheum	X	X	Olgu esaslı sunum
J. Clin Rheum	✓	X	X
Rheum Inter	✓	X	X
Modern Rheum	✓	X	X
Clin Exp Rheum	✓	X	X
Case Rep Rheum		300 \$	

3. Slayt

Kategoriler

- Tamamen orijinal durum / yeni tanımlanan hastalık
- Nadir ve önceden tek tük rapor edilmiş durum
- Sık görülen bir hastalığın nadir bir bulgusu
- Nadir görülen iki semptom / bulgunun beklenmedik beraberliği
- Bir hastalığın başka bir hastalık üzerine etkisi
- Hastanın tedavi sürecinde beklenmeyen komplikasyonu
- Tedavi yönteminin öngörülemezlik durumu
- Yeni ve benzer örneği olmayan tedavi

4. Slayt

kilde 'case report' kabul etmediğini görüyoruz. Bazılarının case'i yalnız letter formatında kabul ettiğini görüyoruz. Bir grup dergi halen devam ediyor. Bir de Case Report in Rheumatology diye bir dergi var, 300 dolar karşılığında sadece case report yayımlıyor. Arthritis and Rheumatism'in son derece katı kuralları var. 'No longer we submitted to the journal' diye kestirmiş, case report gönderilebilmesi için çok özel şartlar belirlemiş. Oxford en liberallerinden biri, yüksek impact faktörlü dergiler arasında. Peki biz ne yaptık? Yaptığımız şu, bir şekilde romatoloji dergilerindeki durumu görelim istedik.

Her bir dergi için, 6 tane dergi aldık, olgu sunumlarının o dergide yayımlanan, tüm yayınlara oranına baktık ve bir de o yayımlanan olgu sunumunun konusuna baktık. Bunun için de ekip olarak 4. sınıf öğrencileri, bizim baş asistan arkadaş ve ben vardım. İkişerli gruplar oluşturarak bakıldı, yani iki kere check edildi ve uyumsuzluklar senior olarak tarafımdan çözüldü.

Kategorilere baktığımızda şu kategorileri görmekteyiz. Tamamen yeni bir hastalık, nadir bir hastalığın, nadir ya da daha önce rapor edilmiş bir durum olabilir. Veya sık görülen bir hastalığın nadir bir prezantasyonu, 4. slaytta kategorileri görmekteyiz, bu kategorilere göre dergilerdeki olgu sunumlarını sınıflandırdık. Bunun için 2007 ile 2011 yılları arasındaki 5. slaytta gördüğünüz dergilere baktık. Annals Rheumatic Disease ile Clinical Rheumatology tamamen zaman yetersizliğinden eksik kaldı. Dolayısıyla dergilerin büyük bir kısmını kapsamış olduğumuzu düşünüyorum. Hemen sonuçlara geçiyorum. 6. slaytta, 2007 ile 2011 arasında 6 dergide toplam 1350 olgu sunumu yayımlanmış ki, bunların bir kısmı bizzat olgu sunumu başlığında, bir kısmı da letter to editor'un altında yayımlanmış olan. Ve bu süreçte de bu dergilerde 9232 tane makale yayımlanmış. Şimdi her bir dergi için olgu sunumlarının o dergide yayımlanan bu 5 yıllık süre zarfındaki yayın sayısına oranına baktığımızda burada en düşük oran Arthritis and Rheumatism'de. Ki o da açıkça bunu baştan koşul olarak belirtmiş durumda. Bu 5 yıllık süreçte, yanılmıyorsam, birazdan datayı vereceğim, 12 tane olgu sunumu var. En yüksek

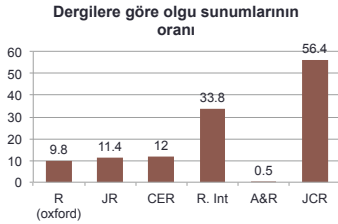
Dergiler (2007-2011)

- Arthritis & Rheumatism
- Rheumatology (Oxford)
- Journal of Rheumatology
- Clinical Experimental Rheumatology
- Rheumatology International
- Journal of Clinical Rheumatology
- Annals of Rheumatic Diseases
- Clinical Rheumatology

5. Slayt

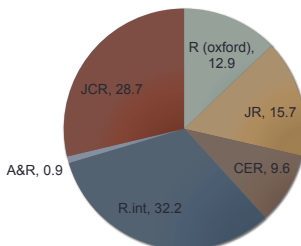
Sonuçlar

• 2007-2011 yılları arasında 6 dergide 1350 olgu sunumu / toplam 9232 makale



6. Slayt

Olguların dergilere göre dağılımı



7. Slayt

oran Journal of Clinical Rheumatology’de, %50 yani 5 yıl içinde yayımlananların yarısı olgu sunumu. Bunu Rheumatology International takip ediyor. Ve bir de yaklaşık %10-12 marjında Oxford Journal of Rheumatology ve Clinical Experimental Rheumatology görmektesiniz. Peki bu dergilerde yayımlanan olguların toplam romatoloji literatürüne katkısına baktığımızda en yüksek oran Rheumatology International’da, ondan sonra Journal of Clinical Rheumatology’de görünüyor.

7. slaytta ayrıca grafik olarak diğer her bir derginin toplam literatüre katkısını görmektesiniz. Peki yıllara göre herhangi bir farklılık var mı? Ne görüyoruz? Baktığımızda istatistik anlamda tek bir farklılık Journal of Clinical Rheumatology’de diğer yıllara göre 2010 yılında bir artış var. Diğerinde hafif iniş çıkışlar görünse de dergilerde yayımlanan olgu sunum oranlarının yıllar içinde değişmediğini görmektesiniz. Kategorileri tekrar gösteriyorum çünkü şu an kategoriler üzerinden gideceğim. 10 tane kategori vardı (8. Slayt). Şimdi bu kırmızıyla gördükleriniz, her bir dergi için ve de totalde en yüksek olan konu başlığı, öyle söyleyeyim. Totale baktığımızda B kategorisi yani nadir ve daha önce az rapor edilmiş kondüsyonlar ilk sırada yer alıyor. Bunun haricinde sık görülen bir hastalığın nadir prezantasyonu, örneğin Rheumatology International’da yayımlanmış. Kimi dergilerde ise gördüğünüzde en sık yer alan durumu yeni tedavi uygulamalarının case reportlarda konu başlığında ilk sırada yer aldığını görüyoruz. Yeşiller de ikinci sıklıkta her bir kategori için gördüğümüz değerler. Burada yeni eklenen nedir diye baktığımızda tamamen yeni bir hastalığın örneğin, rapor edilmesi, bu Clinical Experimental’de gelmiş. Bunun haricinde C ve D kategorileri sık görülen bir hastalığın nadir prezantasyonu ve de ikiden fazla durumun bir arada seyretmesi beklenmedik bir şekilde. Bunun yer aldığını görmektesiniz ve 3. kategoride de görüyoruz yani büyük oranda A, B, C, H grubunda toplanıyor. Bunları da 9. slaytta görmektesiniz.

Yalnız mesela hiç olmayan konular var, mesela bu J sütunu malpractice’i aslında yani tedavi sırasında yanlış giden bir olayın, rapor edilmesi tarzında, bakıyorsunuz kimse kimseyle aybını

Kategoriler

- A- Tamamen orijinal durum / yeni tanımlanan hastalık
- B- Nadir ve önceden tek tük rapor edilmiş durum
- C- Sık görülen bir hastalığın nadir bir bulgusu
- D- Nadir görülen iki semptom / bulgunun beklenmedik beraberliği
- E- Bir hastalığın başka bir hastalık üzerine etkisi
- F- Hastanın tedavi sürecinde beklenmeyen durum
- G- Tedavi yönteminin öngörülemeden komplikasyonu
- H- Yeni ve benzer örneği olmayan tedavi
- J- İşlerin nasıl kötüye gittiğinin itirafı
- K- İlaç etkileşimleri

8. Slayt

paylaşmıyor. Yani J kategorisi en düşük olan kategori gördüğünüz üzere. Sonuç olarak (10. slayt) yaklaşık olarak olgu sunumları şu an romatoloji literatüründe %15 kadar yer kapladığını görüyoruz ve oranın derginin impact faktörüyle ters orantılı olduğunu da görmekteyiz, bu beklediğimiz bir şey. Yıllara göre en azından, bu 5 yıllık süre içerisinde herhangi bir değişiklik olmadığını görüyoruz. Çıkış noktamıza gelince eğer daha çok kabul edilsin isteniyorsa case reportlar şu an için Journal of Clinical Rheumatology ve Rheumatology International dergileri en yüksek oranda yani total yayımlanan paper sayısına göre olgu sunumu basıyor gibi gözüküyorlar ve en popüler temaların olgu sunumları için, nadir veya az rapor edilmiş bir durum, yeni bir tedavi veya sık görülen bir hastalığın nadir prezantasyonu olduğunu söyleyebiliriz. Teşekkür ederim.

N / (%)	A	B	C	D	E	F	G	H	J	K
Rheumatology	28 (16)	31 (18)	11 (6)	17 (10)	8 (5)	5 (3)	16 (9)	36 (21)	0 (0)	22 (13)
J. Rheum	28 (13)	41 (19)	16 (8)	19 (9)	13 (6)	3 (1)	27 (13)	32 (15)	0 (0)	32 (15)
CER	24 (18)	25 (19)	15 (12)	9 (7)	1 (0)	0 (0)	12 (9)	25 (19)	0 (0)	18 (14)
Rheum. Int	1 (0)	67 (15)	121 (27)	89 (20)	55 (12)	5 (1)	31 (7)	50 (11)	0 (0)	22 (5)
Arth& Rheum	2 (17)	4 (33)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	5 (42)	0 (0)	1 (8)
J. Clin. Rheum.	3 (1)	85 (22)	75 (19)	29 (7)	64 (16)	13 (3)	45 (12)	42 (11)	3 (1)	29 (7)
Total	86 (11)	253 (21)	238 (12)	163 (9)	141 (7)	26 (1)	131 (8)	190 (20)	3 (0)	124 (10)

- A- Tamamen orijinal durum / yeni tanımlanan hastalık
- B- Nadir ve önceden tek tük rapor edilmiş durum
- C- Sık görülen bir hastalığın nadir bir bulgusu
- D- Nadir görülen iki semptom / bulgunun beklenmedik beraberliği
- H- Yeni ve benzer örneği olmayan tedavi

9. Slayt

Tartışma

- Olgu sunumları yayınların %15'ini oluşturuyor
- Olgu sunumlarının oranın derginin impact ile ters orantılı
- Yayımlanan olgu sunumu oranlarında azalma (yıllara göre) yok
- R. Int / J.C.R en çok olgu sunumu yayımlayan dergiler
- En popüler temalar
 - Nadir ve önceden tek tük rapor edilmiş durum
 - Yeni ve benzer örneği olmayan tedavi
 - Sık görülen bir hastalığın nadir bir bulgusu

10. Slayt

ROMATOİD ARTRİTTE KLİNİK ÇALIŞMA TASARIMLARI

Yusuf Yazıcı

Romatoloji Bölümü
New York Üniversitesi
Amerika Birleşik Devletleri

Bugün sizlere romatoid artrit (RA) klinik çalışma tasarımları ve bu tasarımların sorunlarından bahsetmek istiyorum. En çok da çalışmalara alınan hastaların özellikleri ve bu çalışmaların sonuçlarının nasıl analiz edilip bize sunulduğu üzerinde duracağım.

İlk 2010 yılında ACR ve EULAR'ın bir araya gelerek yaptıkları yeni RA tanı kriterlerinden bahsedeceğim. Burada başta gelen amaçları, erken dönemde olan RA hastalarına daha kolay tanı koyup, hem tedaviye erken başlama şansını arttırma hem de bu hastaların klinik çalışmalara girmelerini arttırmaydı. Eski 1987 kriterleri, 7 kriterin 4'ünü sağlamak ile RA tanısına yardımcı oluyordu. Ancak 7 kriterin 2'si, ekstraartiküler belirtiler, özellikle nodül oluşumu ve radyografide hasar son 15-20 senede artık az görüldükleri veya görüldükleri zaman ileri evrelerde görüldükleri için, erken gelen hastalarda pek kullanılmayan kriterler oluyorlardı. Romatologlar çoğu hasta için geriye kalan 5 tane kriter ile tanı yapmak zorunda oluyordu. ACR/EULAR grubu da çoğunluğu Avrupa'dan olan erken RA hastaları bulunan büyük veri tabanlarına bakıp, şu anda kullanılan 10 puanlık yeni kriterleri yayımladılar. Bu yeni kriterlere göre 6 puan ve üstü kesin RA olarak tanımlandı. Ancak, bu yeni kriterler herkes tarafından kabul bulmadan önce, birçok başka RA veri tabanında da test edilip, hepsinde doktorun ve eski kriterlerin RA dediği hastaları yeni kriterler ile de RA olarak tanımlandığı gösterildi. Ancak her yeni tanı kriterinin pratikte yapması gereken, benzer hastalar geldiklerinde onları doğru olarak sınıflandırıp sınıflandırmadığını da araştırmak olmalı. Bizim yaptığımız çalışmaya kadar da bu yapılmamıştı. Biz de kendi kliniğimizde ardışık gelen hastalarda kaçının hangi tanı kriterlerini doldurduğunu araştırdık. Doktorun verdiği tanıyı 'altın standart' olarak alıp baktık. Sonuçlarımız daha önceden RA tanısı almış hastaların yine neredeyse hepsinin yeni kriterler ile de RA tanısını aldığını gösterdi. Ancak bunun yanında, %60 lupus, %50 osteoartrit ve %40 sedef artritli hastalarının da yeni RA kriterlerini doldurduğunu gördük. Böylece yeni kriterler aslında eski kriterlere göre daha yararsızdırlar, demek mümkün. Spesifisitiye baktığımızda yeni kriterler %55 iken eski kriterler %76 olup, sensi-

tivite neredeyse ikisinde de aynı oluyordu. Yani kısaca RA olmayan bir sürü hasta yeni kriterler kullanıldığı zaman RA tanısı alabilecekti. Tabii bu da kriterlerin asıl yapılma sebebine tam zıt bir sonuca sebep olabilir. Yeni kriterler hem bilimsel klinik çalışmalarda hem de günlük pratikte yanlış tanı ve tedaviye neden oluyor. Ancak bu yanlışların daha başından önlenmesi mümkün eğer neyi neden yaptığımızı iyi anlayıp, ona uygun düzgün metotlar kullanarak, sadece bir iş yapıyor görünmek yerine gerçekten yararlı işler yapmaya çalışsak. Aynı prensipler ile yeni RA kriteri projesi de gerçekten işe yarar bir proje haline gelebilirdi. Bizim çalışmamızdan sonra yayımlanan başka bir yazıda da, 7 tane bizimkine benzer ancak sadece RA hastalarına bakan çalışma derlenince, biri dışında hepsinde eski kriterlerin yeni kriterlere göre spesifitesi daha iyi çıktı, bu da daha önce dediğim gibi eğer baştan zaten yapılması gereken bu analiz yapılmış olsa bu kadar gereksiz zaman ve emek kaybı olmayabilirdi.

İkinci değinmek istediğim nokta hangi tip hastaların klinik çalışmalara alındıkları ve bu konuya gereken önem verilmezse ne gibi yanlış sonuçlara yol açabileceği. Örnek olarak fostamatinib'in hikâyesini anlatacağım. 2010 senesinde metotreksata (MTX) yeterli yanıt vermeyen hastalara ya fostamatinib ya da plasebo verilen bir çalışma yayımlandı. Çalışmanın sonucu her iki fostamatinib dozunun da plasebodan daha etkili olduğunu gösterdi. Fostamatinib'den önce piyasaya çıkmış olan bütün biyolojik ajanlar da aynı şekilde MTX'a yeterli cevap vermeyen RA'lı hastaların alındığı çalışmalarda plasebodan daha etkili olmuşlardı. Bütün bu çalışmalar karşılaştırılacak olsa veya bir metaanaliz yapılsa, fostamatinib de öteki biyolojik ajanlar kadar etkilidir şeklinde bir sonuca varmak mümkün olacaktır. Ancak çalışmanın detaylarına baktığımızda, hastaların 6 tane ülkeden geldiğini ve bu ülkelerin 5 tanesinin ABD dışı ülkeler olduğunu görüyoruz (Bulgaristan, Kolombiya, Meksika, Polonya ve Romanya). Çalışmada ABD'den gelen hastaların ABD dışından gelen hastalara göre daha az fostamatinib'e yanıt verdiği görülüyor. Bunun sebebinin ne olduğunun ise bilinmediği ve daha önce yapılan çalışmalarda da aynı ABD dışı hastaların daha iyi yanıt vermesinin

görüldüğü anlatılıyor. Hastaların nasıl çalışmaya alındığına baktığımızda, en az 3 ay 7.5-25 mg arası MTX alan hastaların eğer doktorun gözünde yeterli yanıt vermiyorlar ise çalışmaya girebilecekleri belirtilmiş. Hastaların daha önceden kullandıkları ortalama MTX dozu ya da en son aldıkları dozun ne olduğu ise belirtilmemiş. Bu durumda teoride sadece 3 ay 7.5 mg MTX alan hastaların çalışmaya girmesi mümkün oluyor. Unutulmaması gereken birçok hastanın bu 7.5 mg'lık doza pek yanıt vermeyeceği. Kesin olarak bilmek mümkün değil ancak ABD'den gelen hastaların genelde 25 mg metotreksat kullandıkları, en azından daha yüksek doz kullandıkları göz önüne alınırsa, belki de ABD dışında bu kadar yüksek doz kullanılmamaktadır. Başka bir açıklama da, etik olarak tartışılabilir sebeplerden hastalar düşük dozda tedavi edilip çalışmaya girmeleri sağlanıyorsa, o zaman gerçekten MTX'a yetersiz yanıt verme şansı olmamış hastalar çalışmaya girip, yeni ilaca, bu çalışmada fostamatinib'e, daha iyi yanıt verebilirler. Böylece bu çalışma başarı ile tamamlanmış olup, diğer daha önce piyasaya çıkmış biyolojikler gibi etkiliymiş görüntüsü oluşabilmektedir. Bu olası açıklamayı ve neden metotreksat dozlarının detaylarının verilmediğini soran bir mektubu dergiye gönderdik ancak kabul edilmedi. Bu mektup ve reddedilmesi daha sonra "Letter to the Editor" adında yeni bir derginin ortaya çıkmasına sebep oldu. Mektubumuz orada yayımlandıktan yaklaşık 3 ay sonra aynı ilacın bu sefere TNF'ye yeterli cevap vermeyen hastalarda kullanıldığı çalışma yayımlandı ve burada plasebodan daha iyi olmadığı gösterildi. Yani diğer bütün biyolojik ajanların tersine, MTX'e yanıt vermeyen hastalarda başarılı, TNF'ye yanıt vermeyen hastalarda başarısız olduğu görüldü. Ancak bizim daha önce belirttiğimiz gibi MTX'e cevap vermeyen hastalarda da ne kadar başarılı olduğu tartışılır. Nitekim bu çalışmadan 3 sene sonra firma bu ilacın RA'da çalışmalarını durdurup piyasaya sürme planlarını şimdilik durdurduğunu açıkladı.

Kısaca eğer ilaçları karşılaştıracaksak, daha önce MTX kullanmamış erken hastaları çalışmalara almamız gerekiyor. MTX'e ya da başka bir ilaca yeterli cevap vermenin ne olduğunu bilmenin tam standart bir yolu olmadığı için bu çalışmalara

karşılaştırma açısından güvenmemek gerekiyor derim.

MTX kullanmamış hastaları karşılaştırdığımız zaman daha sağlıklı göreceli etki bilgisi almamız mümkün. Bunu göstermek için yaptığımız başka bir çalışmada, sadece daha önce MTX kullanmamış, erken RA'lı hastaları tedavi etmiş olan biyolojik ajan çalışmalarına baktık. Bunu yaptığımızda bu kriterlere uyan sadece 4 çalışma vardı. Etanersept, infliksimab, adalimumab ve abataseptin birer çalışmasını aldık. Hepsinde MTX alan bir kol ve MTX'le biyolojik alan bir kol vardı ve MTX cevaplarını hastalar birbirine çok benzedikleri için, sabit tutup üzerine biyolojik almanın “number needed to treat” (NNT), denen yani bir olumlu yanıt için tedavi edilmesi gereken hasta sayılarına ne kadar etkisi olduğunu hesapladık.

Böyle baktığımızda ACR 20/50/70 sonuçlarında bir fark göremiyoruz, özellikle de burada adalimumab ile abatasept arasında hiçbir fark yok. Sonra daha majör cevaplara baktığımızda yavaş yavaş bir farklılık ortaya çıkıyor. İnfliksimab diğerlerine göre daha az etkili ve adalimumab ile abatasept yine birbirine çok benzer cevaplar veriyorlar. Son olarak remisyona baktığımızda yine etanersept, abatasept, adalimumab aynı, infliksimab daha az etkili görünüyor.

Daha sonra AMPLE çalışmasında abatasept ile adalimumab kafa kafaya karşılaştırıldı ve bizim yaptığımız analizdeki aynı sonuca varıldı, ikisi arasında bir fark olmadığı görüldü.

Yine bizim kendi NYU veritabanına bakarak yaptığımız analizde de, ilk biyolojik kullanan hastalarda abatasept ve adalimumab karşılaştırılınca, AMPLE çalışmasındaki cevap oranlarına çok yakın olan %64-65 oranında yanıt görüldü. Yani hem yayımlanmış klinik çalışmaların tekrar analizi olsun hem de rutin hasta takibinde, eğer veriler doğru toplanmış ise ve doğru hastalarda bakılırsa ilaçların karşılaştırmalı etkinliklerinin ne olduğunu anlamak mümkün, daha fazla kafa kafaya randomize klinik çalışma yapmaya da pek gerek yok gibi görünüyor.

Sonuç olarak, doğru hastalar ve bu hastalardan toplanan veriler doğru şekilde analiz edilirse, hem ilaçların gerçek etkilerini bilmemiz ve karşılaştırmamız mümkün olacaktır hem de bazı çalışmaların gereksiz olarak yapılması engellenebilecektir.

Kaynaklar

1. Kennish L, et al. BMJ Open. 2012;2. pii: e001117. doi: 10.1136/bmjopen-2012-001117.
2. Van der Helm-van Mil AHM, Huzinga TWJ. Ann Rheum Dis 2012;71:1596-1598.
3. Weinblatt ME, Kavanaugh A, Genovese MC et al. N Eng J Med 2010;363:1303-12.
4. Genovese MC, Kavanaugh A, Weinblatt ME et al. Arthrit Rheum 2011;63:337-45.
5. Yazici Y, Swearingen CJ. ACR 2010.
6. Yazici Y, et al. ACR 2011.

A

A. Feinstein	156, 199
ACR/EULAR 2010 kriterleri	142
Adaptif Bayes yaklaşımı	184
Ağırlıklı GEE yaklaşımı	21
Allocation concealment	94
Alzheimer	53
AMPLE çalışması	102, 211
Anketör	8
Ankilozan Spondilit (AS)	25, 36, 41, 79, 171
ANOVA	17
Area under the curve (AUC) (eğrinin altındaki alan)	15
Aricept	53
Artan maliyet - etkinlik oranı (R-incremental cost effectiveness ratio)	60
AS	36
ASAS	36, 146
ASDAS	40
ASSERT çalışması	36
ASPIRE çalışması	25
Atfedilebilir risk	88

B

BASDAI50	36
BASFI	39
BASMI	38
Bayes ara (interim) analiz tasarımları	185
Bayes teoremi	179
Bekle-gör yöntemi	191
BeSt çalışması	25
Biyomedikal model	157
Biyopsikososyal model	157
BOGSGAAT (Bunch Of Guys Sitting Around A Table)	161
British Society of Rheumatology Biological Register (BSRBR)	74

C

'Carry over' (geçmişten taşınan) etki	202
CDC	163
Cetvel 1.de p değeri	66
Cevapsızlık oranı	3
Classical test theory	168
COBRA çalışması	26
COCHRANE Foundation	154
COCHRANE modeli metaanaliz	43

Cognition	166	Etki (efficacy)	57
Common effect (fixed effect)	47	Etki tahmini (effect estimate)	116
Confidence limit	156	Etkinlik (effectiveness)	57
CONSORT	66, 203	Evidence Based Medicine Group	156
CORRONA kayıt sistemi	75	External validity (dış geçerlilik)	204
Cross cultural applicability research	168		
		F	
Ç		F. Bacon	155
Çıkarım (deduction) (tümünden gelme)	182	Faydasızlık analizi (futility analysis)	18
Çoklu imputasyon yöntemi	20	Fixed effect (common effect)	47
		Forest plot (orman grafiği)	45
D		Fostamatinib	102
Data Monitoring Committee		Friction yaklaşımı	58
(veri izleme komitesi) (DMC)	184	Funnel grafiği (ters huni grafiği)	45
Dedüktif yöntem	191		
Değer	61	G	
Denmark Registry	74	Genelleştirilmiş tahmin denklemleri (GEE)	21
DETECT çalışması	138	Generalized estimating equations	
Dış geçerlilik (external validity)	94	analysis	151
Diagnostik (tanı) kriterleri	145, 146, 157	GO-RAISE çalışması	36
Disability	165	Görülme (intangible) maliyet	57
Doğrudan maliyet	56	Güven aralığı	116
Dolaylı maliyet	57	Güvenilir aralığı (credible interval)	179
Dutch Rheumatoid Arthritis Monitoring			
Register (DREAM)	75	H	
Duyarlılık analizi	19	H. Holman	157
Duyarlılık güven aralığı		Head to head (kafa kafaya)	101
Duyarlılık (sensitivity) analizi	58	HIV	163
Dünya Sağlık Örgütü	145	Hiperlipidemi	172
		Hipokrat	155
E		Hot decking	20
E. Garfield	65		
E. Pearson	182	I	
Effect estimate	47	I. Chalmers	155
Effect size	47	ICD-10	75, 145
Eğilim skoru (propensity analizi)	201, 202	ICF	167
Eğrinin altındaki alan (area under the		Inception kohort	140
curve, AUC)	15	Insidence rate	116
Ekipoz (Equipose)	183	Instrumental variable (işe yarayan	
Ekonomik sonuç	57	değişken yöntemi)	202
Eksik veri	17	Intention to treat (niyet ettim tedaviye)	93
Eksik verilerin imputasyonu	16	Internal validity (iç geçerlilik)	204
Eksternal validitesi	204	Item response test theory	168
Elmas şekli (COCHRANE metaanalizinde)	45		
Engellenme	165	i	
Enrichment (zenginleştirme) yöntemi	202	İç içe vaka kontrol çalışma yöntemi	
Eski ve Yeni Kanıta Dayalı Tıp	154	(nested case control)	81

İndirgeme (discounting) oranı	58	Maliyet-etkinlik analizi	57, 59
İndirgenme (duyarlılık) analizi		Maliyet-Etkinlik Oranı (MEO)	60
İngiliz Tıp Araştırma Grubu (MRC)	155	Maliyet-fayda analizi	57
İnsancıl sonuç	57	Maliyet-kar analizi	57
İnsidans	116	Manchester Template	74
İnsidans yoğunluğu örnekleme	82	Mantel Haenszel yöntemi	49
		MANOVA	17
J		MAR (Missing at random)	19
J. Fries	157	Marjinal yapısal model	200
J. Lister	155	Markov Chain Monte Carlo (MCMC)	182, 186
J. Neyman	182	<i>Math on Trial</i>	159
J. Neyman ve K. Pearson	182	MedRA kodlama sistemi	75
Jadad skorlama sistemi	95	Metaanaliz	43
		Meta-regresyon	49
K		MCAR (missing completely at random)	19
K. Pearson	182	MNAR (missing not at random)	19
K. Popper	67		
Kanıtı dayalı tıp	91	N	
Kanıt evi	92	N=1 çalışmaları	195
Kanıt kalitesi	92	Negatif öngörü değeri (negatif prediktif değeri)	175
Kanıt piramiti	92	Non-inferiorite çalışmaları	185
Kayıt kütüğü	73	Norwegian DMARD Register (NOR-DMSRD)	75
Kayıt tabanı (registry)	69	Nosebo	97
Kendini kanıtlama	159	Number Needed to Treat (NNT) (tedavi için gerekli sayı)	30
Kesinlik (precision)	47		
Kesintili eksiklik	15, 18	O	
Kesitsel araştırma	2	Odds ratio	116
Klinik sonuç	57	Olasılık fonksiyonu [L(θ D)]	177
Kohort çalışmaları	116	Orman grafiği (forest plot)	45
Kuşkulu önsellik	183	Osteoporoz	172
Kümülatif insidans	109		
Kümülatif metaanaliz	49	Ö	
		Ölümsüz (immortal) zaman yanlılığı	105
L		Öngören olasılıklar (predictive probability)	174
L. Pasteur	155	Örneklem	5
Last Observation Carried Forward (LOCF)		Özet istatistik yöntem (summary statistics method, SSM)	15
(son gözlemi ileriye taşıma)	16		
Letter to Editor Rheumatology	65	P	
Likelihood	178	p=1.0	159
Locus minoris resistentia	155	p değeri	116, 120
London handicap scale	166	p değerini düzeltmek	14
L(θ D)	177	P. Louis	191
		Patient years (hasta yılı)	103
M		PICO yaklaşımı	92
Maliyet analizi (cost analysis, cost of illness analysis)	56, 57, 59		
Maliyeti en aza indirme analizi	57, 59		

<i>Philosophy of Evidence Based Medicine</i>	155	Sıfır varsayımı (null hypothesis)	182
Popülasyona atfedilebilir risk fraksiyonu	87	Sıklık oranı (rate ratio)	105
Post marketing çalışmaları	126	Sıklık farkı (rate reduction)	105
Posterior probability	174	Sınıflama kriterleri	
Pozitif öngörü değeri (pozitif prediktif değeri)	175	Simpson's paradox	49
Precision (kesinlik)	47	Sistemik Skleroz (SSk)	135
Prevalans	1, 116	Sol sansürleme	159
Prognoz öngörücüleri	25	Son gözlemi ileriye taşıma (<i>last obseravatin carried forward, LOCF</i>)	16
Propensite skorları	202	Sonsal olasılık (posterior probability)	174
Publication bias	48	Sonsal pozitif dağılım (Positive Predictive Distribution) (PPD)	181
Pulmoner Arteriyel Hipertansiyon (PAH)	135	Sonsallık eğrisi	178
Q		Sonsuz zaman yanlılığı (immortal time bias)	111
QALY (Quality Adjusted Life Year) (Kalite Düzeltmeli Yaşam Yılı)	58, 61	Spanish Register	74
R		Subjective well being (öznel doyum)	167
R. Fisher	182	Summary Statistical Method (SSM) (Özet İstatistik Yöntem)	
RABBIT registry	74	Superiorite çalışmaları	185
Random effect	47	Survey	1
Randomize kontrollü çalışma	91	SUT	145
Randomize Plasebo Fazı Tasarımı (RPFT)	190, 192	Swedish Biological Register	74
Randomizasyon	93	SWEFOT çalışması	34
Rate difference (sıklık farkı)		Swiss Clinical Quality Management Database	
Rate ratio (sıklık oranı)	105	Ş	
Rate reduction (sıklık farkı)	105	Şansa bağlı eksik veri (missing at random) (MAR)	18
RAPID3	102	Şansa bağlı olmayan eksik veri (missing not at random) (MNAR)	18
Regression to the mean (ortalama değere kayma)	96	T	
Relatif risk	117	T. Bayes	181
RevMan	52	Tamamen şansa bağlı eksik veri (missing completely at random) (MCAR)	18
Rezidü karıştırıcı	89	Tedavi için gerekli sayı (number needed to treat) (NNT)	105
Ribavirin	51	Tekdüze eksiklik	18
Riley sıçanları	157	Tekdüze olmayan eksilik (kesintili eksiklik)	18
Risk seti örnekleme	82	Tekrarlanan anlamlılık testleri	14
Risk odds oranı		TEMPO çalışması	193
Romatoid artrit (Türkiye)	176	Tercih	61
Run in (alıştırma) zamanı	202	Ters huni grafiği (funnel grafiği)	45
S			
Sansürlü değişken	140		
SAVE çalışması	143		
Schamberg kremi	156		
Science Citation Index	65		
SF-36	166		

Ters Olasılık Teoremi	175	W	
Test öncesi olasılık (önsellik)	177	Wash out (ilaçtan arınma) zamanı	202
Tıbbi maliyet	57	Weibull dağılımı	193
Tıbbi olmayan maliyet	57	WHODAS	166
Tip 1 hata	89	WinBUGS yazılımı	182
Tip 2 (beta) hata	94	Withdrawal (uygulanan tedaviyi kesmek)	
Tonsillektomi	155	çalışmaları	202
Tümdengelim (dedüksiyon)	190	Women's Health Initiative	100
Tümevarım (indüksiyon)	191	WHOQOL	167
Türkiye İstatistik Kurumu	124	Y	
Türkiye Nüfus ve Sağlık Araştırmaları	124	Yarar	61
U		Yetiyitimi	164, 166
Ultrason	147	Yuvalanmış (nested) kohort	81
Unconditional lojistik regresyon	89	Z	
Utilities	58	Zarar vermek için gerekli sayı (number needed to harm) (NNH)	106
Uygunluk örneği (convenience sample)	6		
V			
Vaka-kontrol çalışmaları	85, 117		
Vaka (olgu) kontrol çalışması			
Verimlilik (efficiency)	57		

ROMATOLOJİDE KLİNİK ARAŞTIRMALAR VE NESNEL TİP VI, VII, VIII