

Tıbbi Verilerin Kalitesini Anlamak

Çeviri: Gülay YILMAZEL

BÖLÜM İÇERİĞİ

1. Veri Toplamanın Hedefleri ve Analiz

- 1.1 Doğruluğu ve Kesinliği Teşvik Etmek
- 1.2 Hataları Azaltmak
- 1.3 Değişkenliği Azaltmak

2. Tıbbi Testlerin Doğruluğunu ve Kullanışlılığını İncelemek

- 2.1 Yanlış-Pozitif ve Yanlış-Negatif sonuçlar
- 2.2 Duyarlılık ve Özgüllük
- 2.3 Prediktif Değerler

2.4. Olabilirlik Oranları (Likelihood Ratios),

Olasılık Oranları (Tahmini Rölatif Risk=Odds Ratios) Ve Kesim Noktaları (Cutoff Points)

2.5. Alıcı İşlem Karakteristik Eğrileri

3. Ölçüm Uzlaşması

- 3.1. Genel Uzlaşma Yüzdesi
- 3.2. Kappa Test Oranı

4. Özet

Değerlendirme Soruları, Cevaplar ve Açıklamalar

"Hatalar sanatta değil, zanaatkarlardadır."

Isaac Newton

VAKA GİRİŞ

46 yaşında bir kadın hastaya nadir görünen bir lezyon için cilt biyopsisi yapılmıştır. İlk patoloji raporu "invaziv melanom şüphesi" yönündedir. Biyopsi materyali ikinci bir görüş için gönderildiğinde, yeni yorum "hafif displastik nevus" (örneğin iyi huylu) şeklinde idi. Bir doktor ve bu ders kitabının (JGE) ortak yazarı olan hasta, cilt biyopsi materyalini üçüncü bir patoloğa -onlarca yıllık klinik deneyime sahip melanom üzerine ders kitapları yazar birisi- göndermeye karar vermiştir. Bu üçüncü biyopsi yorumu, tanı spektrumunun ortasına karşılık gelmiştir: Bir "atipik Spitz lezyonu"-taklit melanom olabilir ancak iyi huylu (yani, invaziv olmayan melanom). Özetle, aynı cilt biyopsisi örneği üç farklı patoloğa gönderilmiştir ve çok farklı üç yorum almıştır.

1. VERİ TOPLAMANIN HEDEFLERİ VE ANALİZ

Tıp uygulamaları, nicel ve nitel verilerin sürekli olarak toplanmasını, değerlendirilmesini, analiz edilmesini ve kullanılmasını gerektirir. Veriler hastalıkların tanısı, seyri, tedavilerin belirlenmesi ve değerlendirilmesi için kullanılır. Benzer şekilde, veriler halk sağlığı uzmanları tarafından programları değerlendirmek ve ileriye dönük önceliklere karar vermek için kullanılır. Verilerin doğruluğu açıkça önemlidir. Verilerin doğru olmaması durumunda verilerde "hata" var denir. Tıp alanındaki hatalardan bahsetmek endişe verici olabilir, ancak verilerde hatalar meydana gelir ve hataların ortadan kaldırılması zordur.

"Hata" terimi birden fazla anlamda kullanılmaktadır. Hastaların teşhis ve tedavisinde yapılan hataları

veya yanlış bir organın alınması gibi belirgin ihmal ile daha dikkat çekici hataları ifade etmek anlamında kullanılabilir. Hatanın bu anlamı, Amerikan IOM (NAM) tarafından yayınlanan *Hatasız Kul Olmaz: Güvenli Bir Sağlık Sistemi İnşa Etmek (To Err Is Human: Building a Safe Health System)* adlı raporda vurgulanmıştır. Rapor ülke çapında büyük bir heyecan yaratmıştır.

Tıbbi anamnezler, fizik muayeneler, laboratuvar değerleri ve görüntüleme raporları, insanların dahil olduğu süreç sınırlamaları nedeniyle asla mükemmel değildir. Klinik tıpta ve araştırmalarda, verilerdeki hataların en aza indirilmesi önemlidir. Şöyle ki bu veriler yanlış yönlendirmesinden ziyade bakım veren bireylere rehber olması için kullanılabilir. Bu bölüm tıbbi verilerin kalitesini ölçmenin ve iyileştirmenin yolları üzerine vurgu yapmaktadır.

1.1 DOĞRULUĞUN VE KESİNLİĞİN GELİŞTİRİLMESİ

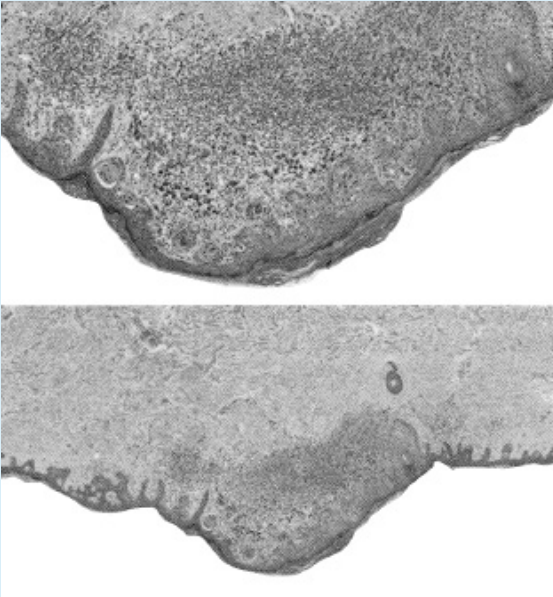
Veri toplamanın birbiriyle ilişkili iki farklı hedefi doğruluk ve hassasiyettir. **Doğruluk**, ölçümün ortalama doğru olma yeteneğini ifade eder. Eğer ölçüm doğru değilse, her iki yöne eşit olmak yerine gerçek değerden bir yöne saptığı için taraflıdır. Tekrarlanabilirlik veya güvenilirlik olarak da bilinen **kesinlik**, bir ölçümün aynı faktörün tekrarlanan ölçümleriyle aynı veya benzer bir sonucu verme yeteneğidir. **Rastgele hata**, verileri herhangi bir yönde kalıcı bir şekilde bozmadığı için farklılaştırmayan hatadır. Rastgele hata büyükse, taraflı olmayan yetersiz kesinlik ortaya çıkar, çünkü doğruluktan sapmalar her iki yönde de karşılaştırılabilir şekilde meydana gelebilir (bkz. Bölüm 4: taraflılık tartışması).

lir. Bununla birlikte, klinik tıpta altın standartlar nadiren bulunur ve altın standarttaki küçük bir hata bile bir testte önemli bir hata görünümüne neden olabilir. Eski tanı testlerinin çoğuna yönelik çalışmalara ek olarak, yeni tıbbi testlerin doğruluğu ve yararlılığı üzerine çalışmalara acilen ihtiyaç duyulmaktadır.

VAKA İNCELEMESİ

Bu cilt biyopsisi deneyiminden on yıl sonra, ders kitabının (JGE) yazarı ek cilt biyopsisi almamıştır. Aynı cilt biyopsisi üzerinde üç farklı yorum almanın kişisel deneyimi, bir bilim insanı olarak araştırmasının odağını değiştirdi. Bir hastanın kanser olup olmadığı her zaman basit bir soru gibi görünmüştür. O, her zaman patoloğun teşhisinin kesin olduğunu, klinik bakım ve araştırmada doğruluğu tanımlamak için altın standart olduğunu varsaymıştı. O, şimdi araştırmasını patoloji ile genişletti, meme ve cilt biyopsilerinin tanısında kapsamlı değişkenliği belgeledi (Şekil 7.7).

Hekimler arasında değişkenliğe yol açan bir faktör, bu metinde açıklanan ilkeler kullanılarak dikkatli bir şekilde toplanmış ve analiz edilmiş verilerin eksikliğidir. Tıpta karşılaştığımız diğer bir zorluk da, "atipik" veya "anormal" tıbbi test sonuçlarına sahip hastaları, sıklıkla onlara net bir fayda sağlamaksızın genişleyen etiketlememizdir. Birçok hasta ve doktor tıbbi verilerin mükemmel olduğunu inanıyor olsa da tıbbi veriler hastalığı hastalık olmayandan ayırt eder, gri alanlar tıp pratiğinde sıklıkla bulunur. William Osler'in dediği gibi, "Tıp bir belirsizlik bilimi ve bir olasılık sanatıdır."



Şekil 7.7 36 ABD'li patoloğ tarafından bağımsız olarak yorumlanan bir cilt biyopsisinin görüntüleri: 7 patoloğ bu biyopsiyi normal/iyi huylu olarak tanımladı, 10 patoloğ bunu biraz atipik olarak nitelendirdi, 8'i ciddi derecede displastik nevüs tanısı koydu, 6'sı melanoma in situ olarak adlandırdı ve 5'i invaziv melanom olarak tanımladı (üstteki resim 5 kat büyütülmüştür, alttaki resim 10 kat büyütülmüştür). Elmore JG, ve diğerleri, BMJ 2017.

4. ÖZET

Veri toplama ve analizinin üç önemli hedefi, doğruluk ve kesinliğin desteklenmesi (bkz. Şekil 7.1 ve 7.2), diferansiyel ve diferansiyel olmayan hataların (rastgele ve rastgele olmayan hatalar) azaltılması, gözlemci içi ve gözlemciler arası değişkenliğin (iki gözlemcinin bulguları arasında veya iki durumda bir gözlemcinin bulguları arasında) azaltılmasıdır. Klinik tıpta tarama testleri ve tanısal (doğrulamalı) testlerin doğruluğunu ve yararlılığını incelemek için çeşitli istatistiksel yöntemler mevcuttur. Genel olarak, yüksek derecede duyarlılığa ve buna karşılık gelen düşük yanlış-negatif hata hızına sahip testler, hastaları taramaya (dışlamaya) yardımcı olmak için, yüksek derecede özgüllük ve buna karşılık gelen düşük yanlış-pozitif hata hızına sahip testler belirli bir hastalığı olduğundan şüphelenilen hastalarda tanıyı doğrulamak (karar vermek) için yararlıdır. Tablo 7.1, 7.2 ve 7.3 duyarlılık, özgüllük, hata hızları, prediktif değerler ve olasılık oranlarının hesaplanması için tanımları ve formülleri sağlar. Tablo 7.4 ve 7.5, gözlemci içi ve gözlemciler arası uzlaşma ölçümlerini tanımlar ve genel uzlaşma yüzdesini ve kappa test oranını hesaplamak için formüller sağlar.

KAYNAKLAR

1. Kohn LT, Corrigan IM, Donaldson MS, eds. *To Err is Human: Building a Safer Health System. Report of the Institute of Medicine*. Washington, DC: National Academies Press; 2000.
2. Brennan TA. The Institute of Medicine report on medical errors: could it do harm? *N Eng J Med*. 2000;342:1123-1125.
3. Dosemeci M, Wacholder S, Lubin JH 1. Does non-differential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol*. 1990;132:746-748.
4. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Eng J Med*. 1978;299:926-930.
5. Abbas AK, Lichtman AH, Pober IS. *Cellular and Molecular Immunology*. Philadelphia, PA: Saunders; 1991.
6. Rose DN, Schechter CB, Adler JJ. Interpretation of the tuberculin skin test. *J Gen Intern Med*. 1995;10:635-642.
7. Elmore JG, Barton MB, Mocer VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med*. 1998;338:1089-1096.
8. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. *N Engl J Med*. 1994;331:1493-1499.
9. Jekel JF, Greenberg RA, Drake BM. Influence of the prevalence of infection on tuberculin skin testing programs. *Public Health Rep*. 1969;84:883-886.
10. Stein PD. Diagnosis of pulmonary embolism. *Curr Opin Pulm Med*. 1996;22295-299.

11. Smith AP. Diagnostic value of serum-creatinine-kinase in coronary-care unit. *Lancet*. 1967;2:178-182.
12. Kinder B. Personal communication, 1994.
13. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, UK: Oxford University Press; 2003.
14. Sackett DL, Tugwell P, Haynes RB, Guyatt GH. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2nd ed. Boston, MA: Little, Brown; 1991.
15. Cicchetti DV, Shanna Y, Cotlier IE. Assessment of observer variability in the classification of human cataracts. *Yale J Biol Med*. 1982;55:81—88.
16. Greenberg RA, Jekel IF. Some problems in the determination of the false positive and false negative rates of tuberculin tests. *Am Rev Respir Dis*. 1969;100: 645—650.
17. Elmore JG, Barnhill RL, Elder DE, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *BMJ*. 2017;357:j2813.
18. Elmore JG, Longton GM, Carney PA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*. 2015;313(11):1122—1152.

OKUMA ÖNERİLERİ

Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Eng J Med*. 1978;299:926—930. [Use of diagnostic tests to rule in or rule out a disease.]

DEĞERLENDİRME SORULARI

1. Hangisi $c/(a + c)$ olarak hesaplanır?
 - B. Doğruluk
 - C. Alfa hatası
 - D. Beta hatası
 - E. Tarafılık
 - F. Kesme noktası
 - G. Yanlış-negatif hata hızı
 - H. Yanlış-pozitif hata hızı
 - I. Pozitif prediktif değer
 - J. Kesinlik
 - K. Rastgele hata
 - L. Duyarlılık
 - M. Özgüllük
2. Hangisi, bir testin bir hastalık varlığını tespit etme yeteneğidir?
 - C. Doğruluk
 - D. Alfa hatası
 - E. Beta hatası
 - F. Tarafılık
 - G. Kesme noktası
 - H. Yanlış-negatif hata hızı
 - I. Yanlış-pozitif hata hızı
 - J. Pozitif prediktif değer
3. Hangisi Tip 1 hatayı gösterir?
 - D. Doğruluk
 - E. Alfa hatası
 - F. Beta hatası
 - G. Tarafılık
 - H. Kesme noktası
 - I. Yanlış-negatif hata hızı
 - J. Yanlış-pozitif hata hızı
 - K. Pozitif prediktif değer
 - L. Kesinlik
 - M. Rastgele hata
 - N. Duyarlılık
 - O. Özgüllük
4. Hangisi normal ve anormal test sonuçlarını tanımlar?
 - E. Doğruluk
 - F. Alfa hatası
 - G. Beta hatası
 - H. Tarafılık
 - I. Kesme noktası
 - J. Yanlış-negatif hata hızı
 - K. Yanlış-pozitif hata hızı
 - L. Pozitif prediktif değer
 - M. Kesinlik
 - N. Rastgele hata
 - O. Duyarlılık
 - P. Özgüllük
5. Hangisi, bir ölçümün ortalama olarak doğru olma eğilimidir?
 - F. Doğruluk
 - G. Alfa hatası
 - H. Beta hatası
 - I. Tarafılık
 - J. Kesme noktası
 - K. Yanlış-negatif hata hızı
 - L. Yanlış-pozitif hata hızı
 - M. Pozitif prediktif değer
 - N. Kesinlik
 - O. Rastgele hata
 - P. Duyarlılık
 - Q. Özgüllük
6. Hangisi $a/(a + c)$ olarak hesaplanır?
 - G. Doğruluk
 - H. Alfa hatası
 - I. Beta hatası
 - J. Tarafılık
- K. Kesinlik
- L. Rastgele hata
- M. Duyarlılık
- N. Özgüllük