



ARAŞTIRMA / RESEARCH

Bayesian estimation of diagnostic accuracy measures when there is no gold standard: a *Helicobacter Pylori* data application

Altın standart test yokluğunda tanısal doğruluk ölçütlerinin Bayesci yaklaşım ile tahmini: *Helicobacter Pylori* verisi uygulaması

Yusuf Kemal Arslan¹

¹Erzincan Binali Yıldırım University, Faculty of Medicine, Department of Biostatistics, Erzincan, Turkey

Cukurova Medical Journal 2021;46(4):1548-1557

Abstract

Purpose: Diagnostic tests are used for defining the disease status of person in a heterogeneous population which consists of healthy and diseased people. In this study, it was aimed to obtain the accuracy measures of the diagnostic test with the Bayesian approach when there was no gold standard test, and to compare it with other methods.

Materials and Methods: In the diagnosis of *Helicobacter Pylori*, different tests are used such as HpSA, Culture, CLO, Histology and PCR. In this study, models containing a priori information (prevalence, sensitivity, etc.) were created to obtain parameter estimations with the Bayesian approach. Bayesian methods were used to correct the imperfect gold standard bias that may occur in determining the accuracy of these diagnostic tests. For this purpose, the data of a medical thesis on *H. Pylori* were analyzed with the help of WinBUGS and R package programs.

Results: When Bayesian inference is made, the results of the positive and negative predictive values of the diagnostic test were found to be more reliable, since the prevalence and a priori information about the disease were taken into account. Similar to the predictive values, reliable results were obtained for sensitivity and specificity values. The credible interval obtained by using Bayesian approach is narrowed when a prior information is used.

Conclusion: In the absence of a gold standard, Bayesian approaches using a prior information such as prevalence and diagnostic test information could facilitate diagnosis for clinicians.

Keywords: Bayesian approach, prior information, gold standard, diagnostics tests, *Helicobacter Pylori*, prior information

Öz

Amaç: Tanı testleri, hasta ve sağlıklı bireylerin oluşturduğu heterojen bir kitlede bireyin gerçek durumunu ortaya çıkarmak amacıyla kullanılır. Bu çalışmada, altın standart test yokluğunda, Bayesci yaklaşım ile ilgilenilen tanı testine ait doğruluk ölçütlerini elde etmek ve diğer yöntemlerle karşılaştırması amaçlanmıştır.

Gereç ve Yöntem: *Helicobacter Pylori* tanısında; HpSA, Kültür, CLO, Histoloji ve PCR gibi farklı testler kullanılmaktadır. Bu çalışmada, Bayesci yaklaşımla parametre tahminlerinin elde edilmesinde önsel bilgileri (prevalans, duyarlılık vs.) içeren modeller oluşturulmuştur. Çalışmada bu tanı testlerinin doğruluklarının saptanmasında ortaya çıkabilen kesin olmayan altın standart yanlışlığının düzeltilmesi amacıyla Bayesci yöntemler kullanılmıştır. Bu amaçla *H. Pylori* ile ilgili bir uzmanlık tezine ait veriler WinBUGS ve R paket programları yardımıyla analiz edilmiştir.

Bulgular: Bayesci çıkarsama yapıldığında hastalık ile ilgili prevalans bilgisi ve önsel bilgi dikkate alındığından tanı testine ait pozitif ve negatif kestirim değerlerine ait sonuçların daha güvenilir olduğu saptanmıştır. Kestirim değerlerine benzer şekilde duyarlılık ve seçicilik değerleri içinde güvenilir sonuçlar elde edilmiştir. Bayesci yaklaşım ile elde edilen güvenilir aralık önsel bilgi kullanıldığında daralmıştır.

Sonuç: Altın standardın olmadığı durumlarda prevalans gibi önsel bilgileri kullanan Bayesci yaklaşımlar klinisyenler için tanıya kolaylık sağlayacaktır.

Anahtar kelimeler: Bayes yaklaşımı, altın standart test, tanı testleri, *Helicobacter Pylori*, önsel bilgi

Yazışma Adresi/Address for Correspondence: Dr. Yusuf Kemal Arslan, Erzincan Binali Yıldırım University, Faculty of Medicine, Department of Biostatistics, Erzincan, Turkey. E-mail: ykarslan@gmail.com

Geliş tarihi/Received: 02.10.2021 Kabul tarihi/Accepted: 05.11.2021 Çevrimçi yayın/Published online: 17.11.2021

GİRİŞ

Tanı testleri, hasta ve sağlıklı bireylerin oluşturduğu bir kitlede bireyin gerçek durumunu saptamak amacıyla kullanılan testlerdir. Doğruluğu kesin olarak kanıtlanmış altın standart testler ile bireylere hasta ya da sağlıklı tanısı konulmaktadır. Bazı altın standart testlerdeki uygulama zorlukları, test maliyetlerinin yüksek oluşu ve testin bazı hastalıklarda girişimsel (invaziv) oluşu nedeniyle her durumda kullanımı uygun ve imkânlı değildir^{1,2}. Bu nedenle sağlık alanında altın standart testlere alternatif olabilecek tanı testleri geliştirilmeye çalışılır.

Klinikte uzmanlar hastaları değerlendirirken, olasılık oranlarını sayısal olarak hesaplayıp kaydını tutmazlar. Bunun yerine, her pozitif sonucu hastalık varlığını düşündürücü olarak ve her negatif testi hastalıktan uzaklaşma olarak yorumlarlar. Ön test - son test oranlarını nicel terimlerden ziyade nitel olarak kavramsallaştırırlar. Belirli tanı testi sonucundaki bulgular bazen çok güçlü ve kesin olabilir, fakat bu testler agresif girişim gerektiren altın standart testlerdir. Örneğin omurilik sıvısında gram boyama ile bakteri tespiti için sıvı alınması işlemi gibi girişimsel bir işlem yerine, öykü ve fizik muayene bulgularını kullanmak bilgi verici olabilir. öykü ve fizik muayene bulguları klinikte son derece güçlü araçlardır ve klinisyenlerin büyük güven duyduğu birçok tanı testinden potansiyel olarak daha güçlüdür. Klinik karar verme aslında temelde bayescidir. Duyarlılıkları ve seçicilikleri nadiren tam olarak bilinmesine rağmen, tüm klinik öykü soruları ve fizik muayeneler aslında tanı testidir. Klinisyenler tanının doğası gereği, ayırıcı tanıları sınırlandırmak için bayesci akıl yürütme kullanır. Yeni bir tanı testi geliştirirken literatürde bulunan hastalık prevalansı veya test ile ilgili önsel bilgiyi kullanmak, nicel olarak bayesci karar mekanizmasını desteklemektedir. Klinikte, örneğin prostat kanserinde hastalığın nasıl seyredeceğini önceden görmek amacıyla prognostik faktörleri içeren nomogramlar mevcuttur ve bu nomogramlar sıklıkla kullanılan naive bayes sınıflayıcılarından biridir.

Altın standart test ile gerçek durumu belirlenen bireylere alternatif test uygulanarak, yeni geliştirilen alternatif tanı testlerinin sınıflama gücünü gösteren doğruluk ölçütleri elde edilir. Altın standart varlığında ve yokluğunda bu doğruluk ölçütleri farklı yaklaşımlar kullanılarak elde edilmektedir. Altın standart var olsa bile bazen kesin tanıyı belirlemek yerine, maliyeti daha düşük olan ve girişimsel olmayan (non-invasiv)

aynı zamanda altın standart teste alternatif olabilecek belirli bir hata oranına sahip alternatif tanı testleri tercih edilmektedir. Altın standart test var olmadığında kesin olmayan altın standart (KOAS) yanlılığı ortaya çıkabilmekte ve yeni geliştirilen testin doğruluk ölçütleriyle ile hatalı sonuçlar elde edilebilmektedir.

Tanı testlerinin değerlendirilmesinde birey ile ilgili gerçek durum bilgisinin bulunmaması (yani altın standart testin yokluğu) sorun yaratabilmektedir. Bu durumda KOAS kullanılmaktadır. KOAS referans olarak kullanıldığında yeni bir test geliştirilmek istendiğinde, yeni geliştirilen testin doğruluk ölçütleriyle ilgili yanlış bir sonuç elde edilebilmektedir. Bu alternatif testten kaynaklanan yanlış durumu düzeltmek için bazı yaklaşımlar önerilmiştir. Bu yanlılığı giderebilmek için değerlendirme yöntemleri; tutarsızlık çözücü, bileşke referans standardı, gizli sınıf analizi (LCA) ve Bayes yaklaşımıdır. KOAS ile yeni geliştirilmeye çalışılan tanı testi arasında tutarsızlık olduğu durumlarda aradaki tutarsızlığı çözmek için çözücü test kullanılabileceği belirtilmiş ve "Tutarsızlık Çözücü-Discrepant Resolution" yaklaşımı önerilmiştir³. Zhou ve ark. literatürde yayınlanmış ve alternatif test kullanıldığı durumlarda yeni tanı testi olarak önerilen bir veya birden fazla testin doğruluk ölçütleri için düzeltmeleri incelemiş ve birçok düzeltmenin, önerilen yeni tanı testinin ikili testler olması durumunda geçerli olduğunu bildirmiştir⁴. Aynı çalışmada alternatif test varlığında kategorik tanı testlerinin doğruluk ölçütleri için düzeltmeler verilirken, tanı testi sayısı ve popülasyon sayısını da göz önünde bulundurmışlardır. Bu soruna çözüm olarak önerilen yaklaşımlardan birisi de araştırmacılar tarafından değişik alanlarda yaygın olarak uygulanmaya başlanan Bayesci yaklaşımdır. Joseph ve ark. 1995 yılında Bayes yaklaşımı ile altın standart test yokluğunda doğruluk ölçütlerinin elde edilmesi üzerine çalışmıştır⁵ ve sonraki birçok çalışmaya kaynak olmuştur. Daha sonraki yıllarda klinik araştırmalarda; Bayesci yaklaşım ile ilgili önsel belirlenmesi, sonsal dağılıştan örnekleme, altın standart yokluğunda değerlendirme gibi durumlar, Dendukuri⁶⁻⁸, Joseph⁹, Lesaffre¹⁰ ve Broemeling¹¹ gibi birçok bilim adamı tarafından çalışılmıştır. Bayesci yaklaşım Thomas Bayes (1702-1761) tarafından sunulmuş ve Bayesci tahmin ilk kez Laplace tarafından kullanılmıştır¹³.

Bugünlerde Bayesci istatistik, araştırmacılar tarafından değişik alanlarda yaygın olarak uygulanmakta ve uygulamada BUGS, WinBUGS ve

R gibi programlar kullanılmaktadır. Gerçek problemlerden kaynaklanan karmaşa ile başa çıkma kapasitesi yüzünden farklı alanlardaki araştırmacılar Bayesci yaklaşımı tercih etmişlerdir. Bayesci yaklaşım araştırmacıya ilgilendiği parametre tahminlerinin elde edilmesinde önsel bilgiyi (prior) kullanma olanağı sağlamaktadır. Bayesci yaklaşım ile tanı testlerinin değerini ölçen duyarlılık ve seçicilik gibi ölçütlerin güvenilirliklerinin artırılmasına yönelik katkı tartışılmaktadır. Bayesci yaklaşım son yıllarda çok popüler hale gelmiş ve tanıda kullanıldığı gibi birçok alanda da Bayesci çıkarsama yapılmaya başlanmıştır. Bu alanlara örnek olarak; yaşam analizi, tekrarlı ölçümler analizi, meta analiz, bioassay çalışmaları, uyum çalışmaları, ölçüm hataları ve hastalık haritalama verilebilir¹³⁻¹⁵. Daha pahalı ve girişimsel yöntemlerin altın standart olarak kullanmak yerine tercih edilen alternatif test ve yeni geliştirilen testlerin kombinasyonları, hastalığa ve teste dair önsel olasılık bilgilerinin varlığı ve yokluğunu bir arada değerlendirebilecek modeller geliştirilmesi son derece önemli olacaktır.

Bu çalışmanın amacı; tanı testlerini altın standart varlığında ve yokluğunda Bayesci yaklaşım ile değerlendirmek ve standart yaklaşımlar yerine Bayesci yaklaşımın (Bayesci çıkarsama, Gibbs örnekleme, Markov Zinciri Monte Carlo yöntemleri kullanılarak) tanı testlerinin değerlendirmesine olan katkısını belirlemektir. Bu amaçla *H. Pylori* tanısında altın standart testin olmadığı durumlarda farklı testleri ile ilgili önsel bilgiyi ve hastalık prevalansına ait bilgiyi kullanarak bayesci yaklaşım ile tanısal doğruluk ölçütleri elde edilmiştir.

GEREÇ VE YÖNTEM

Bu çalışmada bayes yaklaşımıyla ilgili modellemeler kurulup literatürde önerilen diğer yaklaşımlarla ilgili sonuçlarla karşılaştırılmıştır. Gerçek veri seti olarak Çukurova Üniversitesi Tıp Fakültesi Gastroenteroloji Anabilim dalında yapılmış olan ve erişime açık olan bir uzmanlık tezinin verileri kullanılmıştır¹⁶. Çalışma için etik kurul onayı gerekmemektedir. Bu veri setinde yetişkinlerde ve çocuklarda gaitada *Helicobacter Pylori* (HP) antijeni araştıran HpSA testine ait bilgiler bulunmaktadır. Bu çalışmada sadece yetişkinler için elde edilen veriler değerlendirilmiş, tanı amaçlı kullanılan 5 farklı test ve bu testlerin kombinasyonları kullanılarak karşılaştırmalar yapılmıştır. Verilerin analizinde WinBUGS¹⁷, IBM SPSS 20 paket programı¹⁸ ve R programlama dili¹⁹ kullanılmıştır. R'da beta dağılımı için dönüştürme kodu yazılmış ve

bayesci analiz için hazırlanan WinBUGS kodları da kullanılmıştır. WinBUGS'da modeller kurulmuştur ve temel olarak daha önce yapılan çalışmaların algoritmaları ve önsel bilgileri dikkate alınmıştır. Bilindiği gibi *Helicobacter Pylori* tanısında; HpSA, Kültür, CLO, Histoloji ve PCR gibi testler kullanılmaktadır. Buna dayanarak bu çalışmada HpSA, Kültür, CLO, Histoloji ve PCR testlerinin isimleri sırasıyla Test0, Test1, Test2, Test3, Test4 olarak kodlanmıştır. Çalışmada tanı testlerinin doğruluklarının saptanmasında ortaya çıkabilen kesin olmayan altın standart yanlışlığının düzeltilmesi amacıyla bazı yöntemler kullanılmıştır.

Çözücü test: Koçak'ın tezinde Kültür testinin Kesin Olmayan Altın Standart olduğu ve diğer 3 testten herhangi birinin ise Çözücü test olarak kabul edildiği görülmüştür.

Bileşke Referans Standardı yaklaşımı: Kesin olmayan altın standart yanlışlığını düzeltmede kullanılan ve literatürde önerilen Bileşke Referans Standardı yaklaşımı 2'den fazla tanı testi sonucu içeren değerlendirmelerde kullanılabilir. Bileşke Referans Standardı oluşturabilmek için tek sayıda test olması kolaylık sağlayacaktır. Ayrıca Kültür testi dışındaki diğer 3 test kullanılarak da Bileşke Referans Standardı oluşturulabilmektedir.

Bayes yaklaşımı: Önsel bilgi ve veriden elde edilen olabilirlik kullanılarak sonsal ortalama medyan ve güven aralıkları elde edilir. Tanı testlerinde uygulama yapıldığından ilgililenen veya korele testlerin duyarlılık seçicilik değerleri ile ilgili önsel bilgi kullanılır. Prevalans ile ilgili bilginin varlığı ve yokluğu da modellemelere dahil edilmektedir.

Önsel bilgilerin belirlenmesi

Literatürde önsel dağılımın belirlenmesiyle ilgili birçok çalışma yapılmıştır. Önsel bilgi içerme yönünden tıp olarak 2 başlık altında toplanabilir, Bilgi içeren önsel (INF) ve Bilgi içermeyen önsel (NI). Bilgi içeren önseller, bilginin tipi yönünden de 2'ye ayrılır, Sübjektif ve objektif önsel. Bu çalışmada kullanılan önseller aşağıda özetlenmiştir.

Noninformatif (NI) bilgi içermeyen önseller

İlgilenilen testle veya korele herhangi bir testle ilgili daha önce yapılan bir çalışma yoksa veya güvenilir bir bilgi yoksa bu durumda bilgi içermeyen bir önsel kullanımı tercih edilebilir. Bu çalışmada tanı testleriyle ilgili NI önsel dağılımı olarak $B(1,1)$ dağılımı kullanılmıştır.

Ampirik önseller (AI)

Bu çalışmada bileşke referans testler kullanılarak her test için ayrı ayrı ampirik bilgi içeren değerler elde edilmiş ve bu değerlerden beta dağılımına uygun parametreler gerekli kodlarla hesaplanmıştır. Elde edilen testlere özgü değerler değerlendirilmiş ve en iyi 3 testten bir bileşke referans oluşturulup HpSA için doğruluk ölçütleri elde edilmiştir. Beta dağılımının kullanılması sebebi test sonucunun dikotom olması ve binom dağılımının beta dağılımının eşlenik önseli olmasıdır. Ayrıca beta dağılım parametrelerindeki ufak değişikliklerle ortalama ve standart sapma değişikliği sağlanabilmektedir.

Literatür önseller (LI)

Literatür önselleri belirlenirken testler için (T0,T1,T2,T3,T4) örnek verinin toplandığı yıla ait duyarlılık, seçicilik ve hastalıkla prevalansı ile ilgili önsel bilgiler dünya gastroenteroloji örgütünün ilgili yılda yayınladığı rapordan elde edilmiştir²⁰.

Modellemeler

Modellemeler yapılırken ilgilenilen test ile ilgili bilgi içeren ve içermeyen önsel olma durumu göz önüne alınmıştır. İkili kombinasyonlar kullanıldığında tüm testlerle ilgili önsel bilgiler göz önüne alınmıştır. Ampirik bilgi hesaplanan ve ampirik değerlendirmede gerek genel doğruluğu gerekse doğruluk ölçütleri daha iyi olan 3 test seçilmiştir. Seçilen 3 testten ilk olarak en iyi 2 test, daha sonra diğer ikili modele konulmuştur ve 3'lü kombinasyonlar oluşturulmuştur. Modellemeler yapılırken oluşturulan test kombinasyonları Tablo 1'de verilmiştir.

Kullanılan önsel tipine, test kombinasyonlarına ve prevalans bilgisi varlığı/yokluğu durumuna göre modeller kurulmuş ve bu modellerden elde edilen sonuçlar sunulmuştur. İlgilenilen testle ilgili (Test 0) sınıflama yapılmış ve her durumda modelin prevalans bilgisi içeren ve içermeyen test sonuçları bulgular kısmında grafiklerle sunulmuştur.

Tablo 1. Toplam 6 ana model için 84 alt modelde kullanılan test ve önsel bilgi kombinasyonları

| Ana model | Alt model sayısı | Test 0 (HpSA) | Test 1 (Kültür) | Test 2 (CLO) | Test 3 (Histoloji) |
|-----------|------------------|--------------------|--------------------|--------------------|--------------------|
| Model 1 | 3 | NI veya AI veya LI | - | - | - |
| Model 2 | 9 | NI veya AI veya LI | NI veya AI veya LI | - | - |
| Model 3 | 9 | NI veya AI veya LI | - | NI veya AI veya LI | - |
| Model 4 | 9 | NI veya AI veya LI | - | - | NI veya AI veya LI |
| Model 5 | 27 | NI veya AI veya LI | - | NI veya AI veya LI | NI veya AI veya LI |
| Model 6 | 27 | NI veya AI veya LI | NI veya AI veya LI | NI veya AI veya LI | - |

NI: Bilgi içermeyen önsel, AI: Ampirik önsel bilgi, LI: Literatür önsel bilgisi

BULGULAR

Bayesci çıkarsama gerçekleştirilirken ana modeller ve alt modelleri kurulmuş ve modeller ile ilgili açıklamalar listelenmiştir. Bu alt modeller ve test kombinasyonlarına göre kişi sayıları Tablo 2'de verilmiştir. Modellerde kullanılan tanı testlerinin duyarlılık ve seçiciliği ile ilgili kullanılan önsel bilgiler Tablo 3'te sunulmuştur.

İlgilenilen testle ilgili bilgi içermeyen önsel (NI) içeren modelde prevalans varlığında duyarlılık için güven aralığının prevalans içermeyen modele göre daha dar olduğu görüldü. Testle ilgili Ampirik bilgi (AI) varken

duyarlılık değerinin içermeyen NI modele göre daha yüksek olduğu bulundu. Benzer şekilde literatür bilgisi (LI) içeren modelde de duyarlılık değerlerinin prevalans varlığı ve yokluğu durumlarının her ikisinde de diğer modellere göre daha yüksek olduğu gözlemlendi. Seçicilik değerlerinin de duyarlılık değerlerine benzer şekilde değiştiği görüldü. Pozitif kestirim değeri (PKD) ve negatif kestirim değeri (NKD) ile ilgili değerlendirme yapıldığında NI önsel içeren model için prevalans varlığı pozitif kestirim değerini ciddi oranda artırıp güvenilir aralığını daraltmıştır. AI önsel içeren modelde ve LI içeren modelde hem prevalans varlığında hem de prevalans bilgisi olmadığı

durumlarda değerler diğer modellere göre daha yüksek ve güvenilir aralıklar daha dar elde edilmiştir.

İlgilenilen test ile ilgili bilgi yokken Test 1'in (Kültür) katkısı incelendiğinde, AI önsel içeren modelde tüm doğruluk ölçütleri için değerler yükselmiş ve güvenilir

aralıklar daralmıştır. Eğer test1 ile ilgili LI önsel modele dâhil edilirse aynı şekilde tüm doğruluk ölçütleri için değerlerin yükseldiği görüldü ve güvenilir aralıklar daraldı. Bu modellerde prevalans etkisine bakıldığında prevalanstan en fazla etkilenen ölçütün PKD olduğu belirlendi.

Tablo 2. Test kombinasyonlarından olabirlik belirlenmesi için kişi sayıları

| Tekli durum | - | + | | | | | | |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Test 0 (HpSA) | 31 | 48 | | | | | | |
| İkili kombinasyonlar | - / - | - / + | + / - | + / + | | | | |
| Test 0/Test 1 (HpSA/Kültür) | 31 | 0 | 16 | 32 | | | | |
| Test 0/Test 2 (HpSA/CLO) | 18 | 13 | 8 | 40 | | | | |
| Test 0/Test 3 (HpSA/Histoloji) | 19 | 12 | 10 | 38 | | | | |
| Test 0/Test 4 (HpSA/PCR) | 25 | 6 | 6 | 42 | | | | |
| Üçlü kombinasyonlar | + / + / + | + / + / - | + / - / + | + / - / - | - / + / + | - / + / - | - / - / + | - / - / - |
| Test 0/Test 1/Test 2 (HpSA/Kültür/CLO) | 30 | 2 | 10 | 6 | 0 | 0 | 13 | 18 |
| Test 0/Test 2/Test 3 (HpSA/CLO/Histoloji) | 37 | 3 | 1 | 7 | 12 | 1 | 0 | 18 |

Tablo 3. Duyarlılık ve seçicilik için önsel bilgiler

| | Duyarlılık(%) | | | | Seçicilik(%) | | | |
|-------------------|---------------|-------|----------|-------|--------------|-------|----------|-------|
| | AI | | LI | | AI | | LI | |
| | Ortalama | Sapma | Ortalama | Sapma | Ortalama | Sapma | Ortalama | Sapma |
| Test1 (Kültür) | 57,7 | 5,6 | 86 | 3,9 | 92,6 | 2,95 | 99 | 1,2 |
| Test2 (CLO) | 95,3 | 2,4 | 94 | 2,7 | 66,7 | 5,3 | 94 | 2,7 |
| Test3 (Histoloji) | 90,7 | 3,3 | 97,5 | 1,76 | 69,4 | 5,18 | 97,5 | 1,76 |
| Test0 (HpSA) | 76,9 | 4,7 | 95 | 2,45 | 70,4 | 5,13 | 94 | 2,7 |

AI: Ampirik önsel bilgi, LI: Literatür önsel bilgisi

Test 0 (HpSA) ile ilgili AI önsel varken test 1'in (Kültür) katkısı incelendiğinde duyarlılık için elde sonuçlarda test 1 ile ilgili önsel bilginin sonuçları etkilediği fakat değişken önsel tipinin (NI, AI, LI) benzer sonuçlar verdiği gözlenmektedir. Duyarlılık için güvenilir aralık sırasıyla 0,76-0,90; 0,75-0,90; 0,76-0,90 olarak bulundu. Seçicilik için elde sonuçlar duyarlılık ile benzerdi. Prevalans varlığı ve yokluğu kullanılarak değerlendirme yapıldığında ise önceki modeller gibi PKD değerleri prevalanstan

etkilenmiştir. NKD için durum PKD'ye benzerdir.

Test 0 (HpSA) ile ilgili LI varlığı test 1(Kültür) ile ilgili NI ve AI önsel varlığına baskındır. Prevalans varlığı ve yokluğu durumu PKD için güvenilir aralığı daraltmakta ve en çok etkiyi PKD üzerinde yapmaktadır. Test 0 (HpSA) ile ilgili bilgi yokken AI önsel içeren test 2 (CLO) modele eklendiğinde tüm doğruluk ölçütleri için değerler yükselmiş ve güvenilir aralıklar daralmıştır. Eğer test 2 ile ilgili LI önsel dâhil edilirse aynı şekilde tüm doğruluk ölçütleri için

değerler artmıştır. Test0 ile ilgili AI önsel AI veya LI önsel içeren test2 modele eklendiğinde duyarlılık ve seçicilik değerlerinde benzer sonuçlar elde edilmiştir.

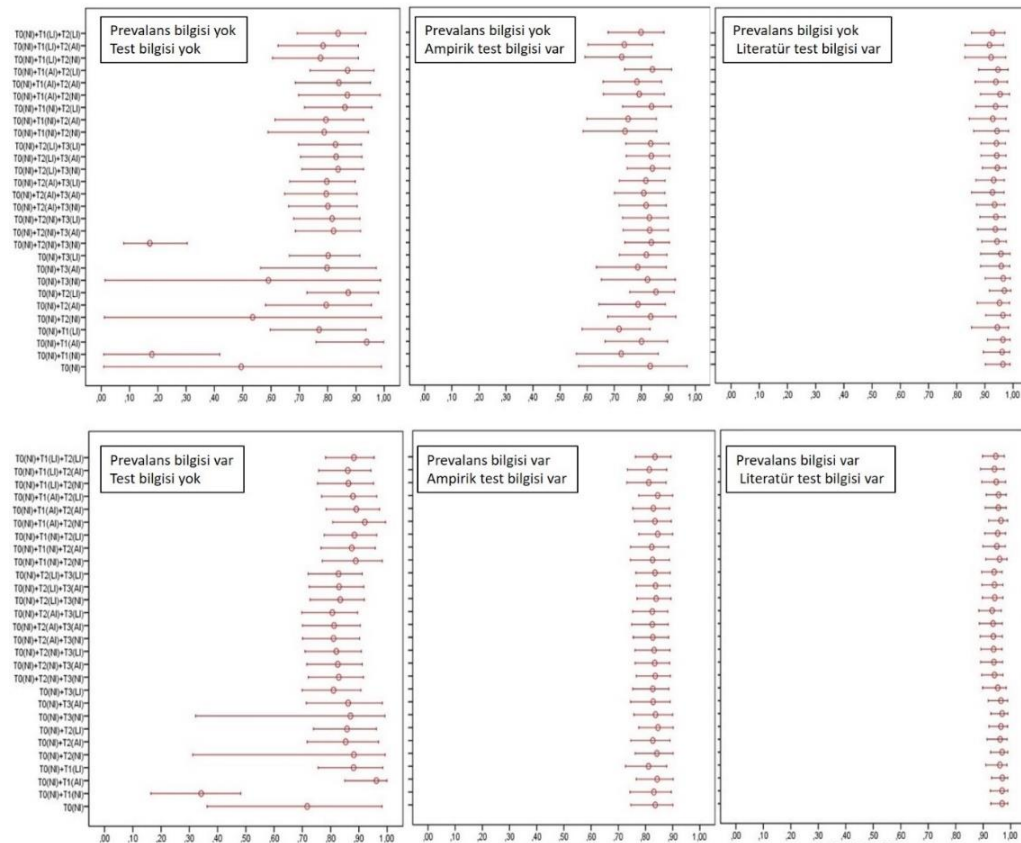
Test 0 (HpSA) ile ilgili bilgi yokken AI önsel içeren test 3 (Histoloji) modele eklendiğinde tüm doğruluk ölçütleri için değerler yükselmiştir. Test 0 ile ilgili AI önsel varken test 3 ile ilgili NI önsel içeren modelde NKD'de belirgin bir artış olmuştur. Test 3 için olabilirliğin NKD'yi arttıracak etki yaptığı gözlenmiştir.

Test 0 ile ilgili LI önsel AI önsel içeren test 3 modele eklendiğinde tüm doğruluk ölçütleri için değerler yükselmiş ve güvenilir aralıklar daralmıştır. LI önsel eklendiğinde NKD ve duyarlılıkta düşüş olmuştur.

Kombinasyon içeren alt modeller incelendiğinde 17 modelde test 0 (ilgilenilen test- HpSA) ile ilgili önsel bilgi NI veya AI tipindedir. Durum böyle olduğunda test 2 veya test 3 hangi önsel tipine sahip olursa olsun

(NI, AI, LI) sonuçlar üzerinde etkisi yoktur. Kalan 9 modelde test 0 ile ilgili LI önseli vardır ve sonuçlar belirgin düzeyde artmıştır. Prevalans varlığı duyarlılık ve seçicilik değerlerini etkilememiştir.

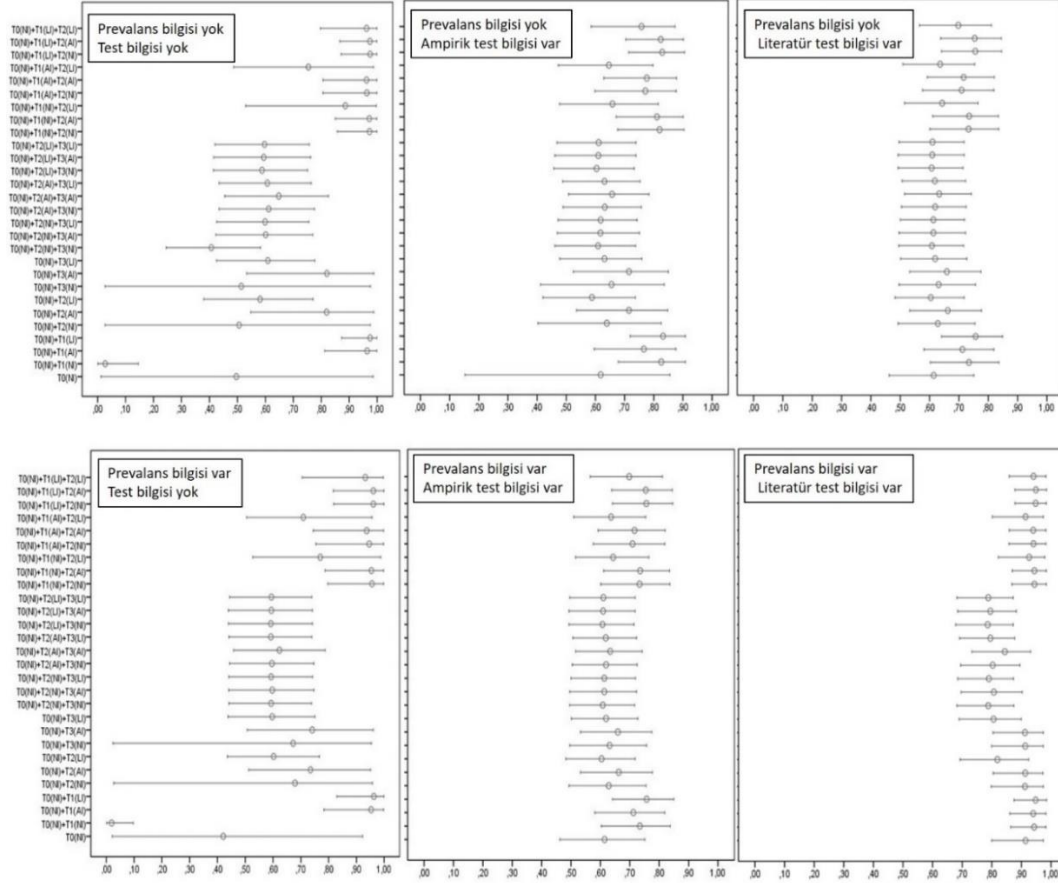
Üç teste ait kombinasyonları içeren modellerde test 0 ile ilgili NI, AI, LI önsel içeren modellerde Test 1 ilgili AI önsel modele eklendiğinde seçicilik ve PKD'nin arttığı, LI önsel modele eklendiğinde seçicilik ve PKD'nin azaldığı gözlenmiştir. Genel Test 0 ile ilgili eklenen önsel bilgi (LI) 9 modelde sonuçları belirgin düzeyde artmıştır. Modeller üzerinde diğer test bilgisi katkıları ve olabilirlik etkisi daha azdır. Prevalans bilgisinin varlığı test 0 ile ilgili LI önsel olmayan durumlarda NKD ve PKD'yi etkilememiştir. Test 0 ile ilgili LI önsel varken NKD ve PKD de değişim azalmıştır. Test0 önsel bilgi tipine (NI, AI, LI) ve prevalans bilgi tipine (NI, INF) bağlı olarak elde edilen alt modellerdeki PKD değerleri Şekil 1'de sunulmuştur.



Şekil 1. Tüm senaryolarda Bayesci yaklaşım ile elde edilen Pozitif Kestirim Değerleri

Benzer şekilde NKD'ne ait deęişimler Şekil 2'de gösterilmiştir. Tüm test ve modellere (6 ana, 84 alt

model) ait sayısal bulgular makale eklerinde sunulmuştur.



Şekil 2. Tüm senaryolarda Bayesci yaklaşım ile elde edilen Negatif Kestirim Deęerleri

TARTIŞMA

Altın standart testlerin birçoęu girişimsel olduğundan onlara alternatif olan KOAS testler saęlığın her alanında geliştirilmektedir. Bu testlere özgü doğruluk ölçütleri elde edilirken çeşitli yanlışlıkları gidermek adına Tutarlılık Çözücü, Bileşke Referans Standardı, Gizli Sınıf Analizi (LCA) yaklaşımları kullanılmaktadır. Bu çalışmada, *Helicobacter Pylori* veri setine Bayesci yaklaşım uygulanmış ve daha önce elde edilen sonuçlarla²¹⁻²³, Bayesci yaklaşım sonuçları karşılaştırılarak yaklaşımın katkısını ortaya konmuştur. Literatürde, tanı testlerinde geleneksel referans testin düşük duyarlılığa sahip olduğunu ve

prevalansı eksik deęerlendirebileceęi söylenmektedir^{24,25}. Dorny ve arkadaşlarının (2004) çalışmalarında popülasyon tabanlı araştırmalarda parazit veya farklı enfeksiyon prevalansının tahmininde altın standart olmayan 2 tanı testi, gaita incelemesi ve serolojik testler kullanılmaktadır²⁶. Ancak bilindięi üzere gaita incelemesi genellikle hastalık prevalans deęerini göz ardı ettiğinden doğruluk ölçütleri elde edilirken yanlış kestirimler yapılabilir. Bayesci yaklaşım kullanılarak popülasyon prevalansının, duyarlılığın, seçiciliğinin, pozitif ve negatif kestirim deęerlerinin her tanı testi için eş zamanlı çıkarımları yapılabilir. Gilks ve arkadaşları sürekli ölçümlerden elde edilen tanı testleri üzerinde

durarak, laboratuvar tanı testinin doğruluğunun değerlendirilmesinde ve kişisel test sonuçlarının yorumlanmasında kullanılan istatistiksel yöntemleri incelemiştir²⁷. Bu çalışmada benzer şekilde belirli bir hastalık için uygulanan test kombinasyonları kullanılmıştır.

Joseph ve ark. altın standart olmadığında, Bayes teoremi için gerekli hesaplamaların zor yapılmakta olduğunu dile getirmiş ve bu yüzden Gibbs örnekleme gibi Monte Carlo algoritmaları geliştirilmişlerdir²⁸. Geliştirilen tekniklerle *Strongyloides* (Nematoda kolu solucanların bir cinsi; iplik solucanları) enfeksiyonu prevalansının tahmini ve iki tanı testinin özelliklerinin incelenmesi için araştırma verilerine uygulanmışlardır. Bu çalışmada sonsal yoğunlukları aynı şekilde Gibbs örnekleme kullanılarak elde edilmiştir. Modellerde, prevalans bilgisinin dikkate alınması gerektiği, aksi takdirde tanı testinin Bayesci değerlendirme sonucunda doğruluk ölçütlerinin daha yüksek veya daha düşük sonuçlar verebileceği gözlemlendi. Prevalanstan en çok etkilenen ölçütlerin PKD ve NKD olduğu görüldü. Bland ve Altman 1994 yılında yaptıkları çalışmalarında PKD ve NKD değerlerinin sadece teste değil prevalans bilgisine bağlı olduğunu dile getirmişlerdir²⁹.

Çalışmada modellere genel olarak bakıldığında uyumsuz ve genele uymayan birkaç modele rastlanmış ve modeller irdelendiğinde bu 5 modelde prevalans ve testlerle ilgili önsellerin NI önsel olduğu görülmüştür. Altın standart test olmayan durumlarda Joseph ve arkadaşları Bayesci yaklaşımla ilgili sundukları yöntemlerin en iyi olası yöntem olduğunu savunmuştur²⁸. Fakat yayınlarında kullanılan önsel bilgi için güvenilir aralığın çok geniş veya çok dar olması durumunda yanlış sonuçlar elde edebileceğini de dile getirmişlerdir. Bu modellerin sıkıntılı olmasındaki sebep tüm önsel bilgilerin en geniş aralıkta (düz önsel- flat prior) olmasıdır. Bu modellerde prevalans bilgisi içeren durumlarda Ünal'ın doktora tezinde farklı yöntemlerle elde ettiği sonuçlara yakın sonuçlar verdiği gözlenmiştir²¹.

Bayesci yaklaşımda önsel kullanılırken fazla test bilgisi içeren modellerin bazı durumlarda hatalı sonuçlar verebileceği göz önünde bulundurulmalıdır. Literatürde yapılan bazı çalışmalarda fazla test kombinasyonu kullanmanın sonuç elde edebilecek örnek büyüklüğünü arttırdığı dile getirilmiştir³⁰. Eğer yeterli örnek büyüklüğü ile çalışılmazsa kullanılan önsel bilgi, veriden elde edilen olabirliğe baskın gelip sonsal dağılışı kendine benzetme eğilimindedir.

Prevalans bilgisi NKD ve PKD için etkili olduğundan muhakkak modele eklenmeli fakat çalışmada gereğinden fazla tanı testi kombinasyonu kullanılmaya dikkat edilmelidir. Dendukuri çalışmasında bu durumun aksi yönünde bulgu elde etmiştir. Test sayısının artmasıyla gerekli örnek büyüklüğünün küçüldüğü durumlar olmuştur fakat bu durum çalışmada açıkladığı, U (uniform prior-NI), L (low), M (moderate), H (high) seviyelerindeki önsel bilgi aralıklarından kaynaklanmaktadır³⁰.

Örnek büyüklüğünü etkileyen diğer bir faktör de testler arası koşullu bağımlılık ve korelasyon durumudur. Bu çalışmada elde edilen her ana modelin ilk alt modelleri NI bilgi içermektedir ve önsel bilgi eksikliğinden kötü sonuçlar elde edilmiştir. Modellerde elde edilen değerler incelendiğinde daha önce yapılan çalışmalarla benzer sonuçlara ulaşıldığı görülmüştür. Bayesci yaklaşımla elde edilen bazı ana modellerin alt modellerinde elde edilen yüksek ölçüt değerleri Burgut ve ark.'nın çözücü test yöntemiyle elde ettiği sonuçlara benzerdir^{22,23}. Burada değerlerin çok yüksek ve LI önsele yakın çıkma sebebi Dendukuri'nin yayınında bahsettiği dar aralıklı yüksek önsel (H) kullanılmasıdır ve daha büyük örnek ölçümüne ihtiyaç duyulmaktadır. İlgilenilen test ile ilgili LI önsel içeren birçok durumda diğer test kombinasyonlarının etkisiz olduğu gözlenmektedir. Literatürde yapılan bazı çalışmalarda fazla test kombinasyonu kullanmanın sonuç elde edebilecek örnek büyüklüğünü (test bağımlılığına göre) arttırdığı dile getirilmiştir³⁰. Eğer yeterli örnek büyüklüğü ile çalışılmazsa kullanılan önsel veriden elde edilen olabirliğe baskın olarak sonsal dağılışı kendine benzetmeye eğilimlidir. İlgilenilen testle ilgili önsel belirlerken önsel aralığını çok dar veya çok geniş tutmak yanlış sonuçlara sebep olabilir. Bu çalışmanın kısıtlılığı, teorik bir yaklaşımın yalnızca *H. Pylori* verisine uygulanmış olmasıdır. Ayrıca basılan bir uzmanlık tezinde bulunan sayılar kullanıldığı için analizde kullanılan örneklem ölçümü kısıtlıdır. Daha büyük veri setleri ve tanı testleriyle ilgili farklı çalışmalar planlanabilir.

Günümüzde tıpta artan maliyet, ilginin bu alana odaklanmasına neden olmuştur. Bu nedenle tanıda doğru karar verme ile ilgili artan sayıda makale görülmektedir. Ancak tıp literatüründe bulunan makaleler incelendiğinde bazen, prediktif değerlerinin önsel olasılık veya yaygınlık değiştikçe değiştiği gerçeğinin göz ardı edildiği, araştırmacıların hem pozitif hem de negatif işlemlerin sonuçlarını yorumlamada önsel olasılığın (veya prevalansın)

oynadığı önemli rolü kavramadıkları, sadece duyarlılık ve özgüllük temalarını kullanan tahmin değerleri kullandıkları izlenmektedir. Bu yöntem, aynı zamanda hangi tanı ya da tarama testine kaynak ayırma noktasında seçim yapması gereken karar vericilere de yardımcı olacaktır. Veri setine uygun önsel bilgileri kullanmak, verilen önselin yoğunluğunu belirlemek en önemli faktördür. Girişimsel testlere alternatif olarak klinikte sıklıkla geliştirilen kesin olmayan altın standart testlerde oluşabilecek yanlışlıkları gidermek adına, hastalık prevalans bilgisini ve alternatif testlerle ilgili tanısal doğruluk ölçütleri bilgisini içeren Bayesci kestirim yöntemleri tercih edilmelidir.

Yazar Katkıları: Çalışma konsepti/Tasarımı: YKA; Veri toplama: YKA; Veri analizi ve yorumlama: YKA; Yazı taslağı: YKA; İçeriğin eleştirel incelenmesi: YKA; Son onay ve sorumluluk: YKA; Teknik ve malzeme desteği: YKA; Süpervizyon: YKA; Fon sağlama (mevcut ise): yok.

Etik Onay: Bu yazıda etik onaya gerek yoktur.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazarlar çıkar çatışması beyan etmemişlerdir.

Finansal Destek: Yazarlar finansal destek beyan etmemişlerdir.

Author Contributions: Concept/Design : YKA; Data acquisition: YKA; Data analysis and interpretation: YKA; Drafting manuscript: YKA; Critical revision of manuscript: YKA; Final approval and accountability: YKA; Technical or material support: YKA; Supervision: YKA; Securing funding (if available): n/a.

Ethical Approval: This article does not any kind of ethical approval.

Peer-review: Externally peer-reviewed.

Conflict of Interest: Authors declared no conflict of interest.

Financial Disclosure: Authors declared no financial support

KAYNAKLAR

- Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: getting better but still not good. *JAMA*. 1995;274:645-51.
- Genç Y. Tanı testi çalışmalarında metodolojik standartların kullanılması. *Ankara Üniversitesi Tıp Fakültesi Mecmuası*. 2003;56:259-64.
- Hadgu A. The discrepancy in discrepant analysis. *Lancet*. 1996;348:592-3.
- Zhou XH, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. New York, Wiley, 2002.
- Joseph L, Gyorkos T W, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*. 1995;141:263-72.
- Dendukuri N, Rahme E, Bélisle P, Joseph L. Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test. *Biometrics*. 2004;60:388-97.
- Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*. 2001;158-67.
- Dendukuri N, Rahme E, Bélisle P, Joseph L. Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test. *Biometrics*. 2004;60:388-97.
- Joseph L, Gyorkos, TW. inferences for likelihood ratios in the absence of a " gold standard". *Med Decis Making*. 1996;16:412-17.
- Lesaffre E, Lawson AB. *Bayesian Biostatistics*. New York, Wiley, 2012
- Broemeling LD. *Bayesian biostatistics and diagnostic medicine*. CRC Press. 2007.
- Bayes T. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philos Trans R Soc Lond*. 1763;53:370-418.
- Ashby D. Bayesian statistics in medicine: a 25 year review. *Stat Med*. 2006;25:3589-631.
- Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 1987;43:671-81.
- Broemeling LD. *Bayesian Methods for Measures of Agreement*. CRC Press. 2009.
- Koçak F. *Helicobacter pylori infeksiyonunda yeni bir tanı yöntemi olan stool antijen testinin değerlendirilmesi (Tıpta Uzmanlık Tezi)*. Adana, Çukurova Üniversitesi, 2002.
- WinBUGS, MRC Biostatistics Unit Cambridge Biomedical Campus. Cambridge Institute of Public Health.
- IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY:IBM Corp. IBM Corp. Released. 2011.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Hunt RH, Xiao SD, Megraud F, Bazzoli F, Hamid S, Vakil N, Malfer-theiner P et al. *Helicobacter pylori* in developing countries. WGO Practice Guideline 2006.
- Ünal İ. Tanı testlerinin değerlendirilmesinde kullanılan standartlar ve analitik yöntemler. *Doktora Tezi*. Adana, Çukurova Üniversitesi, 2010
- Burgut R, Bozdemir N, Erdoğan F, Sertdemir Y, Unal I. Approaches to determining the value of a new test or tests in Diagnostic medicine when there is no gold standard. Invited presentation on the 4rt conference of the EMR of the International Biometric Society, Eilat, Israel, January 23-25, 2007.
- Burgut R, Biostatistical approaches to determining the value of a new test or tests in diagnostic medicine: an example of *Helicobacter Pylori* Infection. WCMCQ, Qatar October 10, 2007.
- Goetghebeur E, Liinev J, Boelaert M, Van der Stuyft P. Diagnostic test analyses in search of their gold standard: latent class analyses with random effects. *Stat Methods Med Res*. 2000;9:231-48.
- Shapiro DE. The interpretation of diagnostic tests. *Stat Methods Med Res*. 1999;82:113-34.
- Dorny P, Phiri IK, Vercruyse J, Gabriel S, Willingham AL 3rd, Brandt J, Victor B et al. A

- Bayesian approach for estimating values for prevalence and diagnostic test characteristics of porcine cysticercosis. *Int J Parasitol.* 2004;34:569-76.
27. Gilks WR, Clayton DG, Spiegelhalter DJ, Best NG, McNeilet AJ et. al. Modelling complexity: applications of Gibbs sampling in medicine. *J R Stat Soc Series B Stat Methodol.* 1993;55:39-52.
 28. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol.* 1995;141:263-72.
 29. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. *BMJ.* 1994;309:102.
 30. Dendukuri N, Bélisle P, Joseph L. Bayesian sample size for diagnostic test studies in the absence of a gold standard: Comparing identifiable with non-identifiable models. *Stat Med.* 2010;29:2688-97.