T.C.
ÇUKUROVA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

SAĞLIK BİLİMLERİNDE YAPILAN GÖZLEMSEL ÇALIŞMALARDA EĞİLİM SKORU KULLANIM ALANLARI

Hakan ANIL

BİYOİSTATİSTİK TEZLİ YÜKSEK LİSANS PROGRAMI
YÜKSEK LİSANS TEZİ

DANIŞMANI
Doç. Dr. İlker ÜNAL

ADANA-2025

KABUL VE ONAY

Biyoistatistik Yüksek Lisans Programı Çerçevesinde yürütülmüş olan

"**Sağlık Bilimlerinde Yapılan Gözlemsel Çalışmalarda Eğilim Skoru Kullanım Alanları**"

adlı çalışma, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tarihi: 07 / 04 / 2025

TEZ SINAV JÜRİSİ

Doç. İlker ÜNAL
Çukurova Üniversitesi
Başkan

Prof. Dr. Gülşah SEYDAOĞLU
Çukurova Üniversitesi
Üye

Prof. Dr. Semra ERDOĞAN
Mersin Üniversitesi
Üye

Yukarıdaki Tez, Yönetim Kurulunun / / tarih ve sayılı kararı ile kabul edilmiştir.

Prof. Dr. Kübra AKILLIOĞLU
Sağlık Bilimleri Enstitü Müdürü

T.C. ÇUKUROVA ÜNİVERSİTESİ SAĞLIK BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI

ETİK BEYANI

Çukurova Üniversitesi Bilimsel Araştırma ve Yayın Etiği Yönergesini okuduğumu ve anladığımı ve Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde ve ortaya çıkan sonuçlarda herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu, bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim. 24/03/2025

Hakan ANIL

Kayıtlı olunan Program: Biyoistatistik Yüksek Lisans

Tezin Konusu: Sağlık Alanında Yapılan Gözlemsel Çalışmalarda Eğilim Skoru Kullanım Alanları

Tezin Türü:

X Yüksek Lisans: Doktora:

Danışmanın Adı-Soyadı: Doç. Dr. İlker ÜNAL

Danışmanın İletişim Bilgileri Telefon:

E-Posta: _____

Öğrencinin İletişim Bilgileri

Telefon:

E-Posta :

Adres:

TEŞEKKÜR

Lisansüstü eğitime başlamamla beraber bana her zaman destek olan, her türlü soruma cevap veren, ilgisi ve sabrıyla her zaman yanımda olduğunu hissettiğim ve tez sürecimde gösterdiği destekle sürecimi kolaylaştıran Doç. Dr. İlker ÜNAL'a,

Lisansüstü eğitimime başladığım günden itibaren bilgilerinden ve tecrübelerinden istifade ettiğim Prof. Dr. Gülşah SEYDAOĞLU, Doç. Dr. Yaşar SERTDEMİR ve Dr. Öğr. Üyesi Yusuf Kemal ARSLAN'a

Eğitimimin özellikle ilk yılında bizlere yardımcı olan ve sorularımıza içtenlikle cevap veren Biyoistatistik Anabilim Dalı araştırma görevlileri Hülya BİNOKAY, Nazlı TOTİK DOĞAN ve Sevinç Püren YÜCEL KARAKAYA'ya,

Eğitimin süresince akademik ve idari konularda her türlü desteği aldığım Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü yönetimi ve çalışanlarına,

Her zaman her konuda yanımda olan sevgili anneme, babama,

Tez sürecimde en büyük desteği sağlayan sevgili eşime,

En büyük motivasyonum olan canım kızıma en içten teşekkürlerimi sunarım.

HAKAN ANIL

İÇİNDEKİLER

KABUL VE ONAY II
TEŞEKKÜR IV
İÇİNDEKİLER V
ÇİZELGELER DİZİNİ VII
ŞEKİLLER DİZİNİ IX
SİMGELER ve KISALTMALAR DİZİNi X
ÖZET XI
ABSTRACT XII
1 Giriş 1
2 Genel Bilgiler 3
2.1 Eğilim Skoru ve Varsayımları 3
2.2 Eğilim Skorunun Hesaplanması 5
2.3 Eğilim Skoru Yöntemleri 6
2.3.1 Eğilim Skoru ile Eşleştirme 6
2.3.2 Tabakalara Ayırma 10
2.3.3 Ters Olasılıklı Ağırlıklandırma 11
2.3.4 Eğilim Skoru Kullanarak Regresyon Düzeltmesi 12
3 Gereç ve Yöntem 14
3.1 Veri Setlerinin Tanıtılması 14
3.1.1 Parsiyel Nefrektomi Veri Seti 14
3.1.2 Osteoarthritis Veri Setinin Tanıtılması 15
3.2 Veri Setlerinden Türetilen Çeşitli Senaryolar 15
3.3 Eğilim Skoru Yöntemlerinin Performanslarının Karşılaştırılması 16
3.4 İstatistiksel analiz 17

4 Bulgular 19
4.1 İkili Yapıdaki Sonuç Değişkenin Yöntem Karşılaştırmalarına Etkisi 19
4.1.1 Trifekta Varlığına Tedavi Gruplarının Etkisi 19
4.1.2 Osteoartrit Veri Seti ile Eğilim Skoru Modellerinin Karşılaştırılması 26
4.2 Sürekli Yapıdaki Sonuç Değişkenin Yöntem Karşılaştırmalarına Etkisi 33
4.3 Örneklem Büyüklüğünün Yöntem Karşılaştırmalarına Etkisi 34
4.3.1 Alt Veri Setlerinin Elde Edilmesi ve Tanıtılması 34
4.4 İkili Yapıdaki Sonuç Değişkenin Görülme Oranının Yöntem Karşılaştırmalarına Etkisi 38
4.4.1 Veri Setinin Elde Edilmesi 38
5 Tartışma 44
6 Sonuç ve Öneriler 49
7 Kaynaklar 51
Ekler 54
Özgeçmiş 62

ÇİZELGELER DİZİNİ

Çizelge 2.1 Eğilim skoru ile eşleştirme yöntemleri 6
Çizelge 3.1 Parsiyel nefrektomi veri setinde bulunan değişkenler ve bu değişkenlerin özellikleri 15
Çizelge 4.1 Bağımsız değişkenlerin gruplar arasında karşılaştırılması 19
Çizelge 4.2 Potansiyel değişkenlerle lojistik regresyon modeli oluşturulması 20
Çizelge 4.3 Eşleştirme sonrası değişkenlerin dengesi 22
Çizelge 4.4 Lojistik regresyon analizi ile yöntemlerin performanslarının karşılaştırılması 25
Çizelge 4.5 Parsiyel nefrektomi veri setinde yöntemlerin ROC analizi ile karşılaştırılması 26
Çizelge 4.6 Osteoporoz tedavi durumuna göre değişkenlerin karşılaştırılması 27
Çizelge 4.7 Eşleştirme yöntemi sonrası değişkenlerin gruplar arasında karşılaştırılması 29
Çizelge 4.8 Osteoartrit veri setine uygulanan eğilim skoru yöntemlerinin performans karşılaştırılması 32
Çizelge 4.9 Osteoartrit veri setinde modellerin ROC analizi ile karşılaştırılması 33
Çizelge 4.10 Tedavi gruplarının operasyon süresi üzerine olan etkisi 33
Çizelge 4.11 Örneklem sayısı 100 olan alt veri setinde eğilim skoru yöntemlerinin sonuçları 34
Çizelge 4.12 Örneklem sayısı 100 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları 35
Çizelge 4.13 Örneklem sayısı 500 olan alt veri setinde eğilim skoru yöntemlerinin sonuçları 35
Çizelge 4.14 Örneklem sayısı 500 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi 36
Çizelge 4.15 Örneklem sayısı 1000 olan alt veri setinde eğilim skoru yöntemlerinin sonuçları 37
Çizelge 4.16 Örneklem sayısı 1000 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları 38
Çizelge 4.17 Örneklem sayısının 100 olduğu senaryoda modellerin karşılaştırmaları 39

Çizelge 4.18 Örneklem sayısı 100 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları 39
Çizelge 4.19 Örneklem sayısının 500 olduğu senaryoda modellerin karşılaştırmaları 40
Çizelge 4.20 Örneklem sayısı 500 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları 40
Çizelge 4.21 Örneklem sayısının 1000 olduğu senaryoda modellerin karşılaştırmaları 42
Çizelge 4.22 Örneklem sayısı 1000 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları 43

ŞEKİLLER DİZİNİ

Şekil 2. 1 Eğilim skoru ile eşleştirme yönteminin görselleştirilmesi 7
Şekil 2. 2 Eğilim skoru ile tabakalara ayırma 11
Şekil 2. 3 Eğilim skoru kullanarak ters olasılıklı ağırlıklandırma 12
Şekil 2. 4 Eğilim skoru ile regresyon düzeltmesi 13

Şekil 4. 1 Eğilim skoru eşleştirme öncesi ve sonrası denge durumu 21
Şekil 4. 2 Ters olasılıklı ağırlıklandırma puanlarının histogram ile gösterilmesi 23
Şekil 4. 3 Değişkenlerin ağırlıklandırma sonrası dengesi 23
Şekil 4. 4 Tabakalandırma sonrası gruplar arasında oluşan eğilim puanı dengesi 24
Şekil 4. 5 Eşleştirme sonrası ortalama farkların gösterilmesi 29
Şekil 4. 6 Ters olasılıklı ağırlıklandırma öncesi ile sonrasında oluşan gruplar arasındaki denge 30
Şekil 4. 7 Tabakalarda oluşan eğilim skoru dengesi 31
Şekil 4. 8 1000 örneklemli alt veri setinde eğilim skoru yöntemleri uygulanması sonrası oluşan denge 37
Şekil 4. 9 Eşleştirme, ağırlıklandırma ve tabakalandırma yöntemlerinin denge performanslarının farklı oranlı sonuç değişkeni ve örneklem sayısı 1000 olan veri setinde karşılaştırılması 42

SİMGELER VE KISALTMALAR DİZİNİ

RKÇ	Randomize kontrollü çalışma
SH	Standart hata
VKİ	Vücut kitle indeksi,
IQR	Çeyrekler arası uzaklık
KBH	Kronik böbrek hasarı
GFR	Glomerüler filtrasyon hızı
OR	Olasılık oranı
EAA	Eğri altındaki alan
SS	Standart sapma
CM	Santimetre

ÖZET

Sağlık Bilimlerinde Yapılan Gözlemsel Çalışmalarda Eğilim Skoru Kullanım Alanları

Eğilim skoru yöntemleri son zamanlarda gözlemsel çalışmalarda artan sıklıkla kullanılmaktadır. Ancak araştırmacıların hangi veri setinde hangi yöntemi kullanacağına dair net bir bilgi yoktur. Bu tez çalışmasında, farklı eğilim skoru yöntemlerinin kullanım alanları, üstün ve eksik yönleri incelenerek yöntem seçiminde etkili faktörler araştırılmıştır.

Çalışmada eğilim skoru yöntemlerini karşılaştırmak amacıyla Üroloji alanından elde edilen, parsiyel nefrektomi yapılan 239 hastalık bir veri seti ve R programında erişime açık olarak sunulan osteoporoz hastalarına ait verileri içeren "Osteoarthritis" veri seti kullanılmıştır. Verilerde tedavi yönteminin sonuç üzerindeki etkisini tahmin etmek için farklı senaryolar oluşturulmuştur. Eğilim skoru yöntemlerinin tedavi etkisini tahmin etme başarısına; sonuç değişkenin ikili yapıda veya sürekli yapıda olmasının, değişen örneklem büyüklüğünün ve ikili yapıdaki sonuç değişkenindeki değişen yanıt oranın etkisi araştırılmıştır.

Yapılan analizler sonucunda, eşleştirme yöntemi; karıştırıcı değişkenlerin neden olduğu dengesizliği gidermede en etkili yöntem olsa da, küçük örneklem büyüklüğünde tedavi etkisi tahmininde başarısının düştüğü (yüksek standart hata ve düşük sınıflama oranı) gözlenmiştir. Ters olasılıklı ağırlıklandırma yöntemi genel olarak tüm senaryolarda kabul edilebilir bir performans göstermiştir. Tabakalandırma yöntemi, tüm popülasyonu kullanabilmesi ve diğer yöntemlere göre daha yüksek eğri altındaki alan değerine sahip olmasıyla öne çıkarken, geniş güven aralıkları dezavantaj olarak saptanmıştır. Regresyon düzeltmesi, özellikle sürekli sonuç değişkenlerinde iyi performans sergilemiş ve düşük örneklem sayılarında yüksek sınıflama başarısı göstermiştir.

Sonuç olarak, eğilim skoru yöntemi olarak en çok bilinen ve sıklıkla tercih edilen eşleştirme yöntemi yerine verinin yapısal özelliklerine ve çalışmanın amacına yönelik olarak diğer eğilim skoru yöntemlerinin tercih edilmesi önerilmektedir. Bu yöntemler arasında sonuç değişkenin ikili yapıda olması halinde ters olasılıklı ağırlıklandırma veya tabakalandırma yöntemi, sonuç değişkenin sürekli yapıda olması halinde ise regresyon düzeltmesi veya eşleştirme yönteminin kullanılması tahmin başarısını artırarak çalışmanın bilimsel değerini yükseltebilecektir.

Anahtar sözcükler: eğilim skoru, ağırlıklandırma, eşleştirme, tabakalandırma, regresyon düzeltmesi.

ABSTRACT

Applications of Propensity Score in Observational Studies in Health Sciences

The utilization of propensity score methods has seen a marked increase in observational studies in recent times. However, there is a lack of clear information about which method researchers should use in which data set. This thesis study aims to address this knowledge gap by examining the areas of use, advantages, and disadvantages of different propensity score methods. Additionally, it investigates the factors affecting method selection.

To this end, a data set comprising 239 patients who underwent partial nephrectomy and obtained from the field of Urology, as well as the "Osteoarthritis" data set containing data from osteoporosis patients, which is openly available in the R program, were utilized for the study. To evaluate the effect of the treatment method on the outcome in the data, different scenarios were created. The impact of the binary or continuous nature of the outcome variable, the variation in sample size, and the alteration in response rate within the binary outcome variable on the efficacy of propensity score methods in estimating treatment effects was examined. The analyses performed revealed that, while the matching method is the most effective method in eliminating the imbalance caused by confounding variables, its success in estimating the treatment effect decreases in small sample sizes (high standard error and low classification rate). Conversely, the inverse probability weighting method demonstrated consistent and satisfactory performance across all scenarios. The stratification method merits particular attention for its capacity to utilize the entire population and attain a high area under the curve value. However, it is important to note that wide confidence intervals were identified as a notable drawback. Regression adjustment demonstrated notable efficacy, particularly in continuous outcome variables, and exhibited high classification success in low sample sizes.

Consequently, as opposed to the most prominent and frequently favored the propensity matching method, it is advised to prefer alternative propensity score methods depending on the structural characteristics of the data and the objective of the study. Among these methods, if the outcome variable is binary, the inverse probability weighting or stratification methods should be employed. If the outcome variable is continuous, the regression adjustment or matching methods should be used to enhance the scientific value of the study by increasing the success of prediction.

Keywords: propensity score, weighting, matching, stratification, regression adjustment.

1 Giriş

Epidemiyolojik çalışmalar deneysel ve gözlemsel olarak ikiye ayrılmaktadır. Randomize kontrollü çalışmalar (RKÇ) deneysel çalışmaların bir alt türü olup, bireylerin gruplara rastgele atanmaları nedeniyle müdahale etkilerinin değerlendirilmesinde altın standart olarak kabul edilirler (1). Rastgele tedavi atanması, tedavi durumunun ölçülen veya ölçülmeyen bazal özellikleriyle daha objektif değerlendirilmesini sağlar. Ancak günümüzde birçok araştırmacı için etik, maliyet gibi engellerden dolayı RKÇ yapmak mümkün olmamaktadır (2). Bu gibi durumlarda birçok araştırmacı gözlemsel bir çalışma dizayn etmek durumunda kalmaktadır.

Tedavilerin sonuçlar üzerindeki etkilerini tahmin etmek için gözlemsel (veya randomize olmayan) çalışmalara yönelik artan bir ilgi vardır (3). Gözlemsel çalışmalarda, tedavi seçimi genellikle bireylerin özelliklerinden etkilenir. Tedavi edilen bireylerin temel özellikleri genellikle tedavi edilmeyen bireylerin özelliklerinden sistematik olarak farklıdır. Bu nedenle, tedavi edilen ve tedavi edilmeyen bireyler arasındaki temel özelliklerdeki farklılıklar yanlılığa sebebiyet vererek çalışmanın gücünü azaltır (4). Tarihsel olarak araştırmacılar uyguladıkları tedavi veya cerrahi yöntemin, tedavi ve kontrol grubu arasındaki demografik, klinik özellik farklılıklarını elimine etmek için regresyon düzeltmesi yöntemini kullanmışlardır. Ancak regresyon düzeltmesi yöntemi birtakım sınırlamalara sahiptir (5,6). Küçük örneklemlerle çok fazla değişken modele dahil edildiğinde, aşırı uyum gerçekleşebilir (7). Yine bu yöntem aykırı değerlere veya etkili verilere karşı duyarlı olabilir. Birkaç uç gözlem, sonuçları orantısız bir şekilde etkileyebilir ve yanıltıcı sonuçlara yol açabilir (8).

Regresyon düzeltmesinin belirli alandaki bu kısıtlamalarından yola çıkarak, Rosenbaum ve Rubin 1983 yılında eğilim skorunu bulmuşlar ve literatüre sunmuşlardır (9). Eğilim skoru, belirli değişkenler göz önüne alındığında bir birimin (birey, hasta vb) belirli bir tedaviye atanma olasılığını temsil eder. Bu yöntem, özellikle tıpta, epidemiyolojide ve sosyal bilimlerde, randomize çalışmaların yapılamadığı durumlarda karıştırıcı değişkenlerin etkilerini azaltmak için geniş ölçüde kullanılmaktadır (9). Çoğu araştırmacı retrospektif dizayndaki çalışmasının limitasyonlarını elimine etmek için eğilim skorunun çeşitli yöntemlerini kullanmaktadır (10).

Günümüzde yaygın olarak kullanılan eğilim skoru yöntemleri; eşleştirme (matching), tabakalandırma (stratification), ters olasılıklı ağırlıklandırma (inverse probability weighting) ve skor değerinin regresyon modeline ek değişken olarak eklenmesi (regresyon düzeltmesi) amaçlarıyla kullanılmaktadır (11-13). Genellikle bir veri seti için bu yöntemlerden sadece biri tercih edilerek eğilim skoru hesaplanırken, aynı veri seti üzerinde birden fazla yöntemin uygulanabilmesi de söz konusudur. Pubmed veri tabanında yapılan incelemeye göre eğilim skorunu kullanan çalışmalarda; eşleştirme yönteminin diğer yöntemlere göre daha fazla kullanıldığı görülmektedir. Ancak yöntemlerin birbirleri yerine kullanılabilme durumu veya eşleştirme dışında daha uygun bir yöntemin veride kullanılabilme seçeneğinin çalışmalarda tercih edilmediği görülmektedir. Eğilim skoru yöntemlerinin hangi veri setinde nasıl kullanılacağına dair net bir kılavuz yoktur.

Bu tez çalışmasında amaç, farklı eğilim skoru yöntemlerinin kullanım alanlarını sağlık alanlarından alınan veriler üzerinde incelemek, yöntemlerin üstün ve eksik yönleri sunma, veri setine, ve çalışmanın amacına uygun eğilim skoru yönteminin nasıl belirleneceği konusunda önerilerde bulunmaktır. Bu amaçla, çalışmada farklı veri setlerinde örneklem büyüklüğü, değişken sayısı, sonuç değişkeni türü gibi parametrelerin yöntem tercihindeki etkisi araştırılmıştır.

2 Genel Bilgiler

Bu bölümde eğilim skoru hakkında detaylı bilgi verilecek, eğilim skoru elde edilmesi sonrası kullanılan 4 yöntem detaylı olarak irdelenecektir.

2.1 Eğilim Skoru ve Varsayımları

Rosenbaum ve Rubin tarafından 1983 yılında önerilen eğilim skoru, gözlenen bazal değişkenler verildiğinde bir kişinin tedavi koluna atanmasının koşullu olasılığı olarak tanımlanmıştır (9).

$$e_i = P(T_i = 1|X_i) \qquad (1)$$

Burada X_i bireye ait bazal değişkenlerin değerlerini, T_i bireyin tedaviyi alıp almadığını ve e_i ise eğilim skorunu göstermektedir.

Eğilim skoru, tedavi alan ve almayan grupların tedavi dışında kalan özelliklerini dengelemek amacıyla kullanılan bir skordur. Eğilim skorunu kullanarak, ölçülen bazal değişkenlerinin dağılımı tedavi alan ve almayan bireyler arasında benzer olması amaçlanır. Dolayısıyla, aynı eğilim skoruna sahip bireylerin olduğu bir kümeden tedavi kollarına atama işlemi sonucunda, gözlenen bazal değişkenlerinin tedavi edilen ve edilmeyen bireylerde aynı olması sağlanır ve bu da tedavinin sonuç üzerindeki etkisinin yansız olarak ölçülmesini sağlar.

Eğilim skoru hem deneysel hem de gözlemsel çalışmalarda kullanılabilir. Deneysel çalışmalarda, gerçek eğilim skoru bilinir ve çalışma tasarımı ile tanımlanır. Gözlemsel çalışmalarda ise, gerçek eğilim skoru bilinmez. Ancak, çalışma verileri kullanılarak tahminde bulunabilinir. Pratikte eğilim skoru çoğu zaman tedavi durumunun gözlenen başlangıç özelliklerine göre analiz edildiği lojistik regresyon modeli kullanılarak tahmin edilir. Eğilim skorunu tahmin etmek için sıklıkla lojistik regresyon yöntemi kullanılsa da bunun dışında başka yöntemler de kullanılmaktadır.

- Ağaç tabanlı yöntemler (recursive partitioning or tree-based methods)
- Rastgele orman metodu (Random forests)

- Yapay sinir ağları (neural networks)
- Torbalama veya artırma (Bagging, boosting)

gibi yöntemler kullanılarak da eğilim skoru hesaplanabilir (14-16).

Eğilim skorunun iki temel varsayımı mevcuttur. Bunlar koşullu bağımsızlık ve pozitiflik varsayımıdır.

Koşullu bağımsızlık varsayımı, 1983 yılında Rosenbaum potansiyel mevcut ortak değişkenlerin modele alındığını ve sonucu etkileyecek karıştırıcı faktörlerin kalmaması gerektiğini ifade eder. Diğer bir ifadeyle tedavi ve gözlem grupları arasında farklılık yaratacak ve en önemlisi tedaviye yanıtı etkileyecek bir değişkenin kalmaması gerekliliğidir. Bu koşul formülize edilecek olduğunda; tedaviye atanma olasılığı (T), olası sonuçlar (Y), değişkenler ise (X) ile gösterilecek olursa koşullu bağımsızlık şöyle formülize edilebilir;

$$Y_i(0), Y_i(1) \perp T_i | X_i \qquad (2)$$

İkinci koşul ise **pozitiflik varsayımıdır**. Pozitiflik varsayımı bireylerin tedavi veya kontrol gruplarından birine atanması ihtimalinin sıfırdan farklı ve pozitif olması durumuna dayanır. Her bir birey iki farklı gruba atanabilir ancak bu değer 0'dan büyük 1'den küçük olmalıdır. Bu durum şu şekilde formülize edilebilir;

$$0 \leq e_i \leq 1 \;\; veya \; 0 < P(T = 1 \,|\, X) < 1 \qquad (3, 4)$$

Bu iki durum sonrası dikkat edilmesi gereken hususlardan birisi de eğilim skoru hesabı dışında kalan değişkenlerin gözlenen değişkenlerden bağımsız olması durumudur. Koşullu bağımsızlık varsayımı sonrası eğilim skoru modeli oluşturan değişkenlerin tedavi sonucunu veya sonuç değişkenlerini etkileyebilir olması gerekmektedir. Sonuç değişkenini etkilemeyen değişkenlerle oluşturulan bir model varlığında sistematik hata ortaya çıkacaktır (17).

2.2 Eğilim Skorunun Hesaplanması

Bu bölümde eğilim skorunun nasıl hesaplandığı adım adım anlatılıp, daha sonra eğilim skorunu kullanan dört farklı yöntemin ayrıntıları incelenecektir.

İlk olarak tedavi değişkenin türüne göre eğilim skoru hesaplanmasında kullanılacak yöntem seçilir. Kategorik yapıda olan tedavi değişkeni için lojistik regresyon başta olmak üzerine ilk bölümde bahsedilen yöntemlerden (diskriminant analizi, random forest, neural networks vb.) birisi tercih edilir. Tedavinin sürekli yapıda olması halinde ise literatürde çalışmalar önerilmiştir (51, 52).

İkinci basamakta modele eklenecek değişkenlerin seçimi yapılır. Bu basamakla ilgili net bir fikir birliği olmamakla beraber çoğu zaman araştırmacı kendi gözlemi ve literatür verilerine dayanarak bu değişkenleri belirler (18).

Üçüncü basamakta eğilim skoru hesaplanır. Bu değer elde edildikten sonra bu skorun kullanılabileceği dört yöntem olan eşleştirme, tabakalara ayırma, regresyon düzeltmesi, ters olasılıklı ağırlıklandırma yöntemi arasında seçim yapılmalıdır. Bu seçimle yapılacak olan sistematik hatanın önüne geçmek amaçlanmaktadır.

Son olarak gruplar arasında bu skorun dengesi sağlanmalıdır. Bu dengeyi değerlendirmek için kullanılan yöntemlerde aşağıda özetlenmiştir.

- **Tanımlayıcı istatistikler:** ortalama, ortanca değerlerin karşılaştırılması üzerinden p değeri ile denge sağlanır. Küçük örneklem sayılarında bu yöntemin gücü düşer.
- **Standardize ortalama fark:** tedavi ve kontrol grubuna ait skor ortalamaları arasındaki mutlak fark ve bu değerlerin ortak standart sapması ile hesaplanır. Dengeyi kontrol etmede en sık kullanılan yöntemlerden biridir (19).
- **Grafikle inceleme:** histogram ve box-plot gibi grafiklerle denge değerlendirmesi yapılabilir.
- **Eğilim skoru aralıkları (Overlap):** tedavi ve kontrol gruplarının eğilim skoru dağılımlarındaki çakışma (overlap) miktarını ölçer.

2.3 Eğilim Skoru Yöntemleri

2.3.1 Eğilim Skoru ile Eşleştirme

Bu yöntem tedavi alan ve kontrol grubunun eğilim puanlarını eşleştirerek karşılaştırmayı amaçlar. Vaka-kontrol çalışma tasarımlarında genellikle kontrol grubu örneklem sayısı daha fazla, tedavi grubu daha az sayıda örnekleme sahip olmaktadır. Böyle bir senaryoda eğilim skoru ile eşleştirme tedavi grubuna en yakın özellikteki kontrol grubundaki vakalarla eşleyecektir. Böylece karşılaştırılabilir iki grup elde edilecektir. Diğer bir deyişle eğilim skoru eşleştirme tedavi uygulanan ve uygulanmayan bireylerin benzer eğilim skorlarına göre eşleştirilmesini sağlamaktadır. Bu yöntem, tedavi edilenler üzerindeki ortalama tedavi etkisini tahmin etmek için kullanılmaktadır (20-21). En sık kullanılan yöntem ise, **bire bir eşleştirme (1:1 matching)** olup, tedavi gören ve kontrol grubunda bireylerin birbirine benzer eğilim skorlarına göre çiftler halinde eşleştirilmesini amaçlar (10). Eğilim skoru eşleştirmesinde kullanılan birçok yöntem bulunmaktadır (22) (Çizelge 2.1).

Çizelge 2.1 Eğilim skoru ile eşleştirme yöntemleri

1- En yakın komşuluk eşleştirmesi (Nearest neighbor matching)
2- Kaliper Eşleştirme (Caliper Matching)
3- Mahalanobis Metrik Eşleştirmesi (Mahalanobis Metric Matching)
4-Mahalanobis Metrik Eşleştirmesi ile Eğilim Skorlarının Birleşimi
5- Tabakalı Eşleştirme (Stratified Matching)
6- Optimal eşleştirme
7- Tam eşleştirme
8- Radius eşleştirmesi
9- Kesin eşleştirme

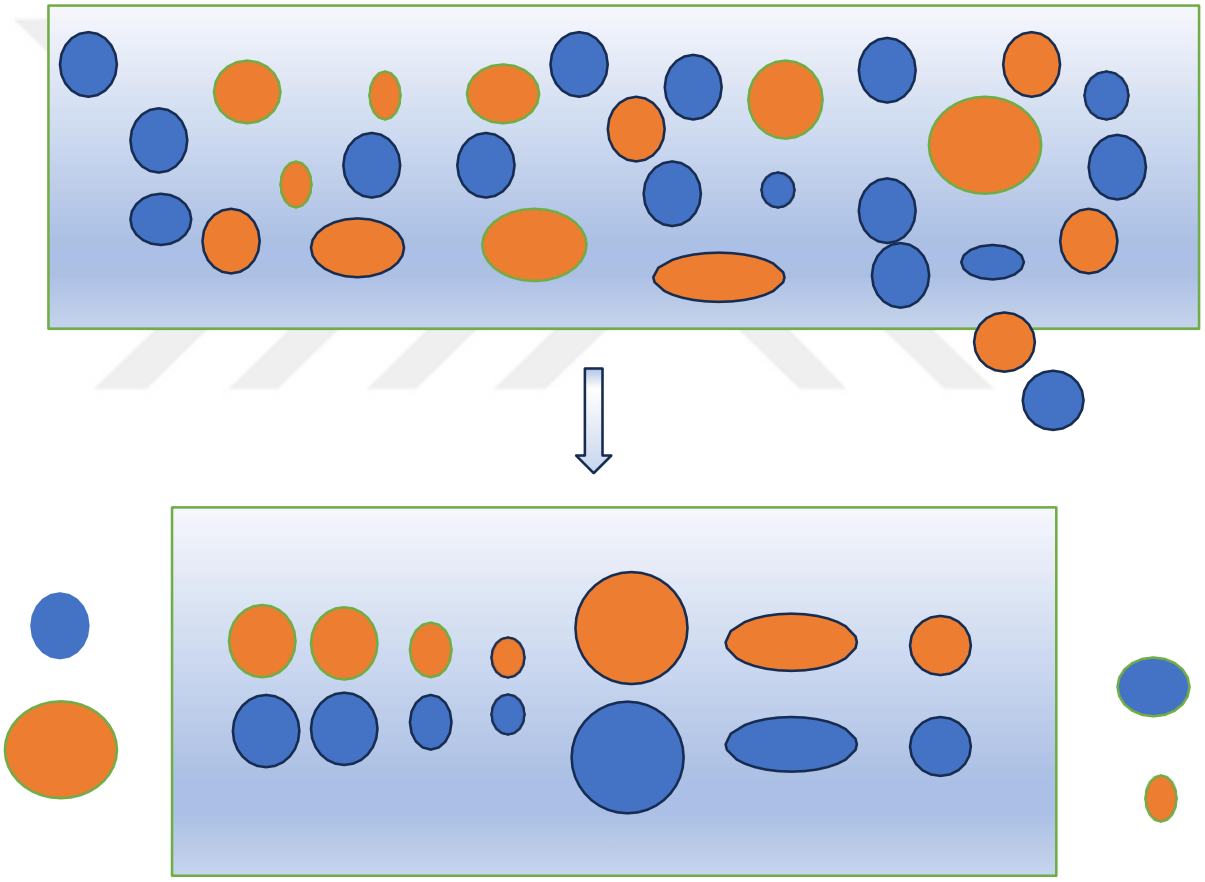
2.3.1.1 En yakın komşuluk eşleştirmesi

En yakın komşu eşleştirme, eşleştirme yöntemleri içerisinde en sık olarak uygulanan yöntemdir. Bu yöntem, mesafe ölçütlerini temel alarak kontrol ve tedavi gruplarındaki bireyleri eşleştirir. Diğer bir deyişle, her bir tedavi grubundaki bireye, mesafe ölçütlerine göre en yakın olan kontrol grubundaki birey atanır (23). Bu teknik,

benzerlikleri baz alarak gruplar arasında denge kurmak ve karşılaştırılabilir bir grup karşılaştırması için kullanılır (24).

$$C(P_i) = min_j |P_i - P_j| \qquad (4)$$

Formülde; $C(P_i)$ j. kontrol grubu birimi ile eşleşen i. tedavi grubu birimi arasındaki en küçük farkı, P_i; i. tedavi grubu birimi için hesaplanan eğilim skorunu, P_j; j. kontrol grubu birimi için hesaplanan eğilim skorunu gösterir



Şekil 2. 1 Eğilim skoru ile eşleştirme yönteminin görselleştirilmesi

2.3.1.2 Kaliper Eşleştirme

Kaliper mesafesi içinde en yakın komşu eşleştirme, en yakın komşu eşleştirmenin bir türüdür; ancak, eşleşen bireylerin eğilim puanları arasındaki mutlak farkın belirli bir

eşik değerinin (kaliper mesafesi) altında olması koşulu eklenmiştir. Bu değer genellikle 0.2 olarak kabul edilir. Bu yöntemde, bir tedavi grubundaki birey için, eğilim puanı belirlenen mesafe içinde olan tüm tedavi almamış bireyler belirlenir**.** Bu sınırlı tedavi almamış bireyler grubundan, eğilim puanı tedavi grubundaki bireyin eğilim puanına en yakın olan birey eşleştirme için seçilir. Eğer tedavi grubundaki bireyin eğilim puanına, belirlenen kaliper mesafesi içinde kalan hiçbir tedavi almamış birey bulunamazsa, bu tedavi grubundaki birey herhangi bir tedavi almamış birey ile eşleştirilmez. Bu durumda, eşleştirilemeyen tedavi grubundaki birey, ortaya çıkan eşleştirilmiş örneklemden hariç tutulur (13,26).

$$|P_i - P_j| < e \qquad (5)$$

e: burada belirlenen standart hatadır ve birinci çeyreklik temsil etmesi için 0.20 değeri genellikle kabul gören değerdir.

2.3.1.3 Mahalanobis Metrik Eşleştirmesi

Mahalanobis mesafesi, çok boyutlu bir uzaydaki mesafeleri belirleyen bir ölçüttür ve Mahalanobis metrik mesafe eşleştirmenin temelini oluşturur. Tedavi ve kontrol gruplarındaki rastgele sıralanmış birimler arasında, tedavi grubundaki ilk birim, kontrol grubundaki en yakın birimle eşleştirilir ve aralarındaki mesafe Mahalanobis mesafesi olarak tanımlanır.

$$D_{ij} = (x_i - y_j)^T S^{-1} (x_i - y_j) \qquad (6)$$

şeklinde formülize edilir.

2.3.1.4 Mahalanobis Metrik Eşleştirmesi ile Eğilim Skorlarının Birleşimi

Bu yöntemde Mahalanobis uzaklığı hesaplanır. Ek olarak eğilim skoru hesaplanır ve bu değere eklenir. Bu iki skorun birleşimi ile tedavi ve kontrol grubu eşleştirilir.

2.3.1.5 Tabakalı Eşleştirme

Tabakalı eşleştirmede veriler eğilim puanına göre tabakaları ayrılır. Tabaka sayısı konusunda bir fikir birliği bulunmamakla birlikte genellikle 5 katman kullanılır. 5 tabakalı eşleştirmede %90-95 oranında objektif karşılaştırılabilir gruplar elde edildiği ifade edilmiştir (26,27). Örneğin 0-1 arasında değerleri olan bir eğilim skorunda; 1.tabaka eğilim skoru 0 – 0.2 arasında olanlar, 2. tabaka 0.2 – 0.4 arası, 3. tabaka 0.4 – 0.6 arası, 4. tabaka 0.6 – 0.8, 5. tabaka ise 0.8 – 1.0 arası puan alan birimlerden oluşacaktır. Ve her tabakada tedavi ve kontrol grubu eşleşmesi ile bu yöntem uygulanabilir.

2.3.1.6 Optimal Eşleştirme

Bu eşleştirmede baz alınan nokta tedavi grubu ile kontrol grubu arasında eğilim skoru farklılığıdır. Tedavi grubundaki bir bireyin eğilim puanı ile kontrol grubundaki bireyin eğilim puanı arasındaki farkın en az olduğu birimler birbiriyle eşleştirilir. Eşleşmeyen bireyler çalışmadan çıkarılır. Bu yöntem gruplar arasında daha az farklı eşleşmeler sağlayacağı için, dengesiz dağılan birimleri dengeleyebilir (28). Optimal eşleştirme eşleşen birimler arasındaki farkın toplamının en az olmasını amaç edinirken, haris (greedy) eşleştirme ilk birimden başlayarak en yakın kontrol eşleşmesini amaç edinir. Bu eşleştirme şu şekilde hesaplanabilir. Tedavi grubu X, kontrol grubu Y olsun. Tedavi grubu birimleri i ile, kontrol grubu birimle j ile temsil edelim. δ (X_i, Y_J), tedavi ve kontrol birimleri arası mesafedir. Burada ağırlıklandırma ile elde edilen uzaklık formülü

$$D = \Sigma w(|X_i|, |Y_j|)\, \delta\,(X_i, Y_j) \qquad (7)$$

şeklinde olacaktır.

2.3.1.7 Tam Eşleştirme

Tam eşleştirme optimal eşleştirmenin bir varyantıdır. Optimal eşleştirmeden farklı olarak tabakalar oluşturulur ve tedavi birimi birden fazla kontrol birimi ile eşleşebilir. Örneklem sayısının özellikle kontrol grubunda azaltılması istenmeyen durumlarda kullanılabilir (29).

2.3.1.8 Radius Eşleştirmesi

Dizaynı kaliper eşleşmesine benzerdir. Fakat kaliper eşleşmesinde temel farkı belirlenen sınıra giren tüm tedavi ve kontrol gruplarının eşleşmesidir (30).

2.3.1.9 Kesin Eşleştirme

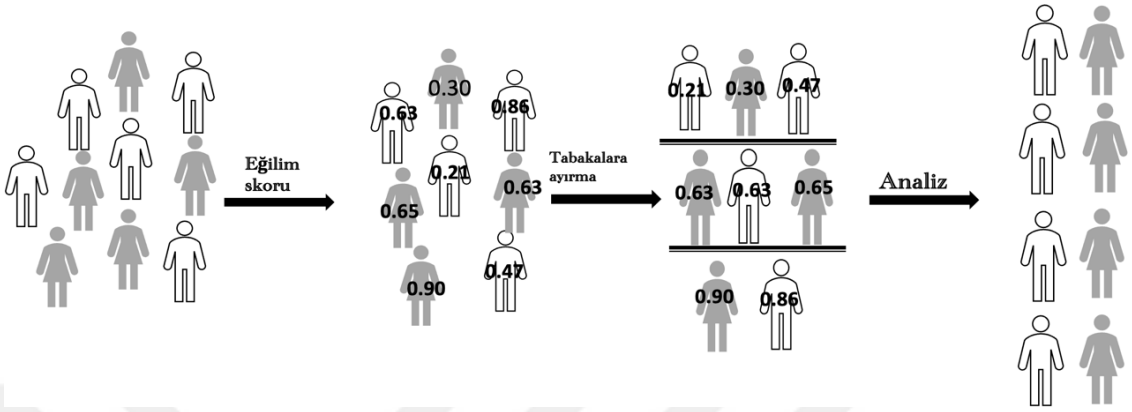
Her bir değişken için aynı eğilim puanına sahip olanları eşleştirmeyi amaçlar. Ancak bu durumun sağlanması için yüksek örnek sayılı veri setlerine ihtiyaç vardır. Bu da kullanımını sınırlamaktadır. Değişken sayısının fazla olması da uygulanabilirliğini kısıtlamaktadır.

2.3.2 Tabakalara Ayırma

Tabakalara ayırma yöntemi eğilim skoru hesaplanması sonrası her bir bireyin önceden belirlenen eşik değere göre tabakalara ayrılmasını içerir. Genellikle 5 tabaka kullanılır. Bu 5 tabaka kullanıldığında yanlılığın %90 oranında azaltıldığı bildirilmiştir (31). Daha fazla tabaka kullanıldığında, tedavi edilen ve edilmeyen gruplar içinde daha homojen alt gruplar oluşturulabilir. Bu durum, her tabakada tedavi grubu ve kontrol grubu arasında daha iyi bir denge kurulmasını sağlar ve bu da yanlılığı azaltır. Yani, tedavi ve kontrol gruplarını daha fazla alt gruplara ayırmak, bu grupların birbirine daha benzer hale gelmesini sağlayabilir. Ancak tabaka sayısı arttıkça, her yeni tabakanın yanlılığı azaltma etkisi giderek küçülür. Başlangıçta tabaka sayısını artırmak önemli bir fark yaratırken, tabaka sayısı çok fazla artınca ekstra tabakaların katkısı giderek azalır (31,32). Bu yüzden, genellikle optimum sayıda tabaka seçilmesi önerilir.

Her bir tabaka içinde, tedavinin sonuçlar üzerindeki etkisi, tedavi edilen ve edilmeyen bireylerin sonuçlarının doğrudan karşılaştırılmasıyla tahmin edilebilir. Tedavi etkisinin tabakaya özgü tahminleri, daha sonra tüm tabakalar arasında birleştirilerek genel bir tedavi etkisi tahmini elde edilir (11). Bu bağlamda, tabakaya özgü ortalama farklılıkları veya risk farkları tahmin edilebilir. Bu tahminler, genel bir ortalama farkı veya risk farkı oluşturmak için birleştirilebilir. Genel olarak, tabaka özgü etki tahminleri, her tabaka içinde yer alan bireylerin oranı ile ağırlıklandırılır. Örneğin, örneklem β eşit büyüklükte tabakaya ayrıldığında, her tabakanın özgü ağırlığı $1/\beta$ olarak belirlenir ve bu, genel tedavi etkisini tahmin etmeye olanak tanır (33). Eşleştirme yönteminde olduğu gibi,

tabaka içinde regresyon ayarlaması yapılarak, tedavi edilen ve edilmeyen bireyler arasındaki kalan farklılıklar hesaba katılabilir (33-34).



Şekil 2. 2 Eğilim skoru ile tabakalara ayırma

2.3.3 Ters Olasılıklı Ağırlıklandırma

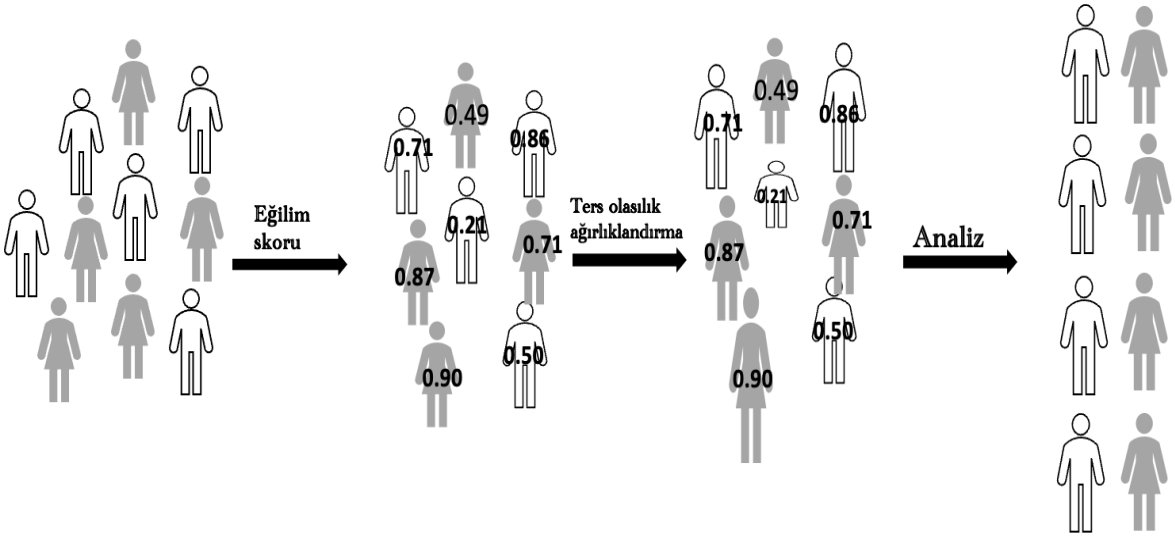
Eğilim skorunu kullanarak ters olasılıklı ağırlıklandırma yöntemi tedavi grubundaki bireylerin tedavi alma olasılıklarının tersini (1/olasılık), kontrol grubundaki bireylerin ise tedavi almama olasılıklarının tersini kullanılır.

Bu yöntemde eğilim skoru hesaplandıktan tedavi gruplarının sonuç üzerindeki etkisini ölçmek için bir ağırlık hesabı yapılır. Buradaki ağırlık değeri şöyledir:

$$w = \frac{T}{e} + \frac{1-T}{1-e} \qquad (9)$$

Burada T tedaviyi, e ise eğilim skorunu göstermektedir.

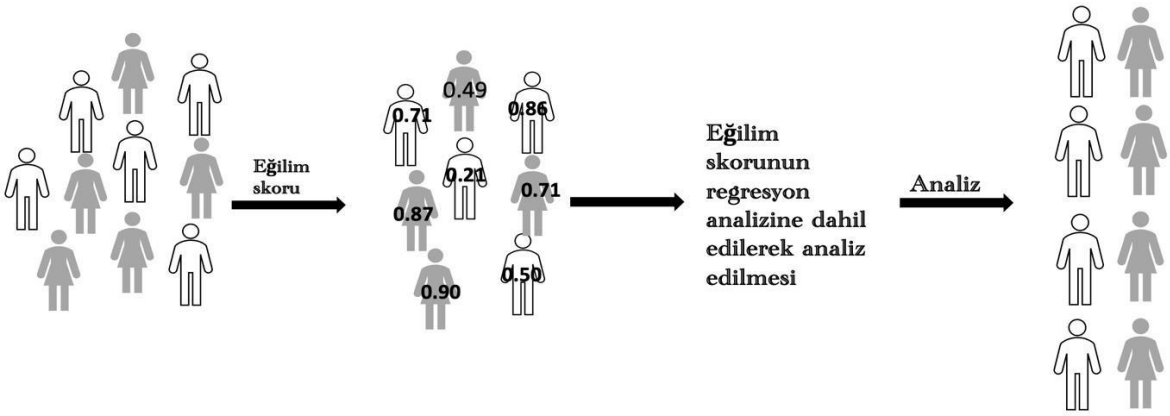
Eğilim skoru 0-1 arasında olduğunda, ağırlıklandırma yönteminde eğilim skorunun çarpmaya göre tersi alınmasından dolayı uç değerler ortaya çıkabilir. Bu uç değerler sistematik hataya neden olacağı için bu bireyler çalışmadan çıkartılması sistematik hatayı ortadan kaldıracak ve daha objektif sonuçlar elde edilecektir (35).



Şekil 2. 3 Eğilim skoru kullanarak ters olasılıklı ağırlıklandırma

2.3.4 Eğilim Skoru Kullanarak Regresyon Düzeltmesi

Eğilim skoru kullanarak Regresyon düzeltmesi, gözlemsel çalışmalarda eğilim skorunun dört temel kullanım yönteminden biridir. Bu yöntem, bağımlı değişkenin (sonuç değişkeni), tedavi durumu göstergesi ve tahmin edilen eğilim skoru üzerinde regresyon analizi yapılmasını içerir. Kullanılacak regresyon modeli, sonuç değişkeninin türüne bağlıdır. Sürekli değişkenler için lineer model, kategorik değişkenler içinse lojistik regresyon analizi kullanılır. Tedavinin etkisi, uyarlanmış regresyon modelinden tahmin edilen regresyon katsayısı kullanılarak belirlenir. Burada hesaplanan eğilim skoru değeri regresyon analize dahil edilerek karıştırıcı faktörler için ayarlama yapılır (10).



Şekil 2. 4 Eğilim skoru ile regresyon düzeltmesi

3 Gereç ve Yöntem

Bu tez çalışmasında eğilim skorunun yaygın olarak kullanıldığı eşleştirme, ters olasılıklı ağırlıklandırma, tabakalandırma ve regresyon düzeltmesi yöntemleri sağlık alanından elde edilen veri setlerinde uygulanmıştır. İlk veri seti Anıl ve ark. tarafından daha önce yayınlanmış bir çalışmanın verileridir (36). Bu veri seti böbreğin anterior ve posterior lokalizasyonlarına bulunan kitlelere göre iki gruba ayrılan 239 hastadan oluşmaktadır. İkinci veri seti ise R programında MatchThem kütüphanesinde bulunan eksik verilerin dışlanması sonrası 2363 örneklem sayısına sahip, erişimi açık olan "Osteoartritis" veri setidir.

3.1 Veri Setlerinin Tanıtılması

3.1.1 Parsiyel Nefrektomi Veri Seti

Böbrek tümörü nedeniyle parsiyel nefrektomi uygulanan 239 hastadan oluşan bir veri setidir. Hastalar, kitlelerinin bulunduğu anatomik lokalizasyona göre iki gruba ayrılmıştır. Operasyon yöntemi laparoskopik retroperitoneal parsiyel nefrektomidir. Bu yöntem böbrek posteriorunda bulunan kitlelerde daha çok tercih edilmektedir. Böbreğin anteriorunda bulunan kitleler içinse bu yaklaşım cerrahi zorluklar barındırmaktadır. Bu nedenle posterior kitleler kontrol, anterior yerleşimli kitleler tedavi grubu olarak belirlenmiştir. Veri seti, grupların pre-op, per-op ve post-op ölçümlerini içermektedir. Cerrahinin başarısını ölçen trifekta varlığı (cerrahi sınırların negatif olması, komplikasyon yaşanmaması ve böbrek fonksiyonlarının korunması) ikili yapıda sonuç değişkenidir. Operasyon süresi ise sürekli sonuç değişkeni olarak belirlenmiştir. Çalışmada tedavi başarı ölçütü olarak bu ikili yapıdaki değişken ele alınmış ve buna göre analizler yapılmıştır. Çalışmada karıştırıcı faktörler, yani tedavinin sonuç değişkeni üzerine etki etmesini etkileyebilecek preoperatif sürekli ve kategorik değişkenler de bulunmaktadır. Bu değişkenlerden sürekli yapıda olanlar yaş, vücut kitle indeksi, preoperatif glomerüler filtrasyon düzeyi, kitle boyutu, nefrometri skorudur. Kategorik yapıda olan değişkenler ise cinsiyet, kronik böbrek hasar derecesi, klinik evre ve ek komorbidite varlığıdır.

Çizelge 3.1 Parsiyel nefrektomi veri setinde bulunan değişkenler ve bu değişkenlerin özellikleri

Değişken adı	*Değişken türü*	*Verilen kod*
Yaş	Sürekli, bağımsız	
Cinsiyet	Kategorik, bağımsız	1=erkek, 2=kadın
Ek komorbidite varlığı	Kategorik, bağımsız	0=hayır, 1=evet
eGFR	Sürekli, bağımsız	
Tümör boyutu	Sürekli, bağımsız	
Vücut kitle indeksi	Sürekli, bağımsız	
Nefrometri skoru	Sürekli, bağımsız	
Tedavi	Kategorik	0=posterior yerleşimli böbrek kitleleri, 1=anterior yerleşimli böbrek kitleleri
Operasyon süresi	Sürekli, bağımlı	
Kanama miktarı	Sürekli, bağımlı	
Kan tranfüzyonu ihtiyacı	Kategorik, bağımlı	0=hayır, 1=evet
Hastanede yatış süresi	Sürekli, bağımlı	
Trifekta	Kategorik, bağımlı	0=hayır, 1=evet

3.1.2 Osteoarthritis Veri Setinin Tanıtılması

Çalışmada kullanılan ikinci veri seti R programı MatchThem kütüphanesinde bulunan "osteoarthritis" veri setidir. Bu veri seti kronik osteoartrit hastalığını etkileyen faktörlerin araştırıldığı çalışmanın verileridir. Veri setinde 2363 hastanın verileri bulunmaktadır. Hastalar osteoporoz tedavi alma durumuna göre iki gruba ayrılmıştır. Veri setinde yaş, vücut kitle indeksi, cinsiyet, ırk, sigara kullanımı değişkenleri bulunmaktadır. Çalışmanın amacı osteoporoz tedavisi almanın kronik osteoartrit gelişimi üzerine etkisini ortaya çıkarmaktır. Bu veri setinin yeterli örneklem büyüklüğüne sahip olması nedeniyle alt veri analizleri bu veri setinden türetilmiştir.

3.2 Veri Setlerinden Türetilen Çeşitli Senaryolar

Çalışmada, sonuç değişkeninin kategorik veya sürekli yapıda olmasının analiz sonuçları üzerindeki etkisi, parsiyel nefrektomi veri seti kullanılarak incelenmiştir. Örneklem büyüklüğünün ve sonuç değişkeninin görülme sıklığının etkisi ise Osteoartrit veri seti üzerinde oluşturulan senaryolar aracılığıyla değerlendirilmiştir.

Parsiyel nefrektomi veri setinin ilk senaryosunda, sonuç değişkeninin kategorik yapıda olmasının yöntemler üzerindeki etkisi incelenmiştir. Bu kapsamda, ikili yapıdaki sonuç değişkeni (trifekta varlığı) üzerinde tedavinin etkisi değerlendirilmiştir. İlk adımda, tedavi grupları arasında demografik ve klinik özellikler açısından istatistiksel farklılıklar araştırılmıştır. Tedavi grupları arasında farklılık gösteren ve klinik olarak anlamlı olduğu düşünülen değişkenlerin tedavi seçimindeki etkisi, lojistik regresyon analizi ile analiz

edilmiş ve bu analiz sonucunda eğilim skoru elde edilmiştir. Bu veri setinde trifekta varlığının oranı %82'dir. Osteoartrit veri setinde ise osteoporoz tedavisi almanın, kronik osteoartrit gelişimi üzerine etkisi araştırılmıştır. Kronik osteoartrit görülme oranı %50.7'dir.

Eğilim skorunun hesaplanmasının ardından, eşleştirme, ters olasılıklı ağırlıklandırma, tabakalandırma ve regresyon düzeltmesi yöntemleri veri setine uygulanmıştır. Eşleştirme yöntemleri arasından, literatürde en sık kullanılan en yakın komşuluk eşleştirmesi ve kaliper (0.20) metodu tercih edilmiştir. Tabakalandırma yönteminde ise, veri seti eğilim puanına göre 5 tabakaya ayrılarak analiz edilmiştir.

İkinci senaryoda, sonuç değişkeninin sürekli değişken olmasının etkisi araştırılmıştır. Çalışmada kullanılan parsiyel nefrektomi veri setinde bu değişken, operasyon süresi olarak belirlenmiştir. Eğilim puanı hesaplanmasının ardından, eğilim skoru yöntemleri veri setine uygulanmıştır. Sonuç değişkenin sürekli yapıda olması nedeniyle, operasyon süresi üzerindeki tedavinin etkisi, lineer regresyon analizi ile incelenmiştir. Bu analiz, tedavinin operasyon süresi üzerindeki etkisini daha net ve istatistiksel olarak anlamlı bir şekilde değerlendirmeyi amaçlamıştır.

Üçüncü senaryoda, örneklem büyüklüğünün etkisi incelenmiştir. Bu senaryoda Osteoartrit veri seti kullanılmıştır. İkili yapıdaki sonuç değişkeninin tüm veri setindeki yanıt oranları korunarak (yanıt oranı %50), toplamda 100, 500 ve 1000 hastalık alt gruplar rastgele seçilmiştir. Bu senaryonun amacı, örneklem büyüklüğünün eğilim skoru yöntemlerinin performansı üzerindeki etkisini ortaya koymaktır. Üçüncü senaryonun ikinci kısmında ise, sonuç değişkeni olan osteoartrit oranı 1/3 (%33) oranında sabit tutularak, 100, 500 ve 1000 örneklem büyüklüğüne sahip veri setleri oluşturulmuştur. Bu veri setlerinde eğilim skoru yöntemleri karşılaştırılarak, örneklem büyüklüğünün ve sonuç değişkeninin görülme sıklığının yöntemlerin performansına olan etkisi detaylı bir şekilde analiz edilmiştir.

Bu senaryolarla, yöntemlerin farklı örneklem koşulları altındaki tutarlılığını ve güvenilirliğini değerlendirmeyi amaçlamaktadır.

3.3 Eğilim Skoru Yöntemlerinin Performanslarının Karşılaştırılması

Bu çalışmada, dört farklı eğilim skoru yönteminin performanslarının karşılaştırılmasında temel kriterler, tedavinin sonuç değişkeni üzerindeki tahmini etkisi

ve bu etkinin standart hata değeridir. Sürekli sonuç değişkeninde R^2 değeri diğer bir sonuç karşılaştırma parametresidir. Bunun yanı sıra, ROC analizi ile elde edilen eğri altında kalan alan (EAA), yöntemlerin performansını değerlendirmek için ek bir parametre olarak değerlendirilmeye alınmıştır. EAA'nın kullanımının temel amacı çalışmaya ek bir gösterge olabilme potansiyelidir. Standart hatadaki yakın farklılıklarda yol gösterici bir parametre olabileceği düşünülmüştür.

Yöntemlerin eğilim skoru dağılımları ve dengeleme başarıları, grafikler ve ilgili değerlendirmelerle detaylı bir şekilde sunulmuştur. Bu görsel ve istatistiksel analizler, yöntemlerin tedavi grupları arasındaki dengelenme başarısını ve sonuç değişkeni üzerindeki etkilerini karşılaştırmak için kullanılmıştır. Bu sayede, farklı eğilim skoru yöntemlerinin güçlü ve zayıf yönleri ortaya konularak, hangi yöntemin hangi durumlarda daha uygun olduğuna dair bilimsel bir temel sağlanmıştır.

3.4 İstatistiksel analiz

Çalışmada tedavi seçiminde etkili olan değişkenleri belirlemek amacıyla önce tek değişkenli analizler, sonra ise bu analizlerin sonuçlarından hareketle çoklu analizler kullanılmıştır. Tek değişkenli analizlerde sürekli verilerin normal dağılıma uygunluğu Kolmogrov-Smirnov testi ile yapılmıştır. Normal dağılıma uygun sürekli değişkenlerin tedavi grupları arasında karşılaştırılmasında Student t testi, normallik varsayımının sağlanamaması durumunda ise Mann Whitney U testi kullanılmıştır. Bu analizler sonucunda istatistiksel farklılığa sahip değişkenlerle, tedavi gruplarına atanmada potansiyel yanlılık oluşturacak değişkenler kullanılarak lojistik regresyon analizi yapılmış ve eğilim puanı hesaplanmıştır. Elde edilen eğilim puanına göre 4 yöntem için veriler elde edilmiştir.

1- Eşleştirme yöntemi için "Matchit" fonksiyonu kullanılarak en yakın eşleştirme yöntemi ile tedavi grupları eşleştirilmiş yeni bir veri seti elde edilmiştir. Bu veri setinde tedavi gruplarının sonuç değişkeni üzerindeki etkisi sonuç değişkeninin türüne göre lojistik regresyon veya lineer regresyon analizi ile hesaplanmıştır.

2- Ağırlıklandırma yöntemi için elde edilen eğilim puanı kullanılarak ters olasılıklı Ağırlıklandırma puanı elde edilmiştir.

$$Ters\ olasılıklı\ ağırlıklandırma\ puanı = \frac{Tedavi\ grubu}{Eğilim\ puanı} + \frac{1-Tedavi\ grubu}{1-Eğilim\ puanı} \quad (10)$$

3- Tabakalama yöntemi için elde edilen eğilim puanına göre veri seti 5 tabakaya ayrılmıştır. Her tabaka için ayrı ayrı analizler yapılarak ortalama tedavi etkisine ait regresyon katsayısı ve standart hata hesaplanmıştır.
4- Regresyon düzeltmesi yönteminde ise regresyon analizinde oluşturulan modele grup değişkenine ilave olarak eğilim puanı da eklenerek analizler yapılmıştır.

Çalışmanın istatistiksel analizi R 4.2.2 programında yapılmıştır. Analizler sırasında R da bulunan "haven", "broom", "MatchIt", "dplyr", "survey", "tableone", "ggplot2" isimli paketler kullanılmıştır. $P<0.05$ olması istatistiksel anlamlı olarak kabul edilmiştir.

4 Bulgular

4.1 İkili Yapıdaki Sonuç Değişkenin Yöntem Karşılaştırmalarına Etkisi

İkili yapıdaki sonuç değişkeni için parsiyel nefrektomi veri setinde tümör lokalizasyonu grup (0=posterior, 1=anterior), trifekta varlığı (0=yok, 1=var) ise ikili yapıdaki sonuç değişkenidir. Trifekta varlığı iyi bir cerrahi göstergedir. Trifektanın varlığı (1) olması cerrahi sonuçların daha iyi olduğuna işaret eden bir parametredir. İkinci veri seti olan osteoartrit veri setinde ise grup değişkeni osteoporoz tedavisi (0=tedavi almayan, 1=tedavi alan), ikili yapıdaki sonuç değişkeni ise kronik osteoartrit (0=yok, 1=var) gelişimidir.

4.1.1 Trifekta Varlığına Tedavi Gruplarının Etkisi

4.1.1.1 Preoperatif Değişkenlerin Tedavi Grupları Arasında Değerlendirilmesi

Çalışmanın ilk aşamasında değişkenlerin tedavi grupları arasındaki farklılıkları değerlendirilmiştir. İki grup arasında istatistiksel farklılık oluşturan değişkenler saptanarak, regresyon modeli oluşturulması amaçlanmıştır. Çizelge 4.1'de görüldüğü üzere preoperatif değişkenler gruplar arasında farklılıklar bulunmaktadır.

Çizelge 4.1 Bağımsız değişkenlerin gruplar arasında karşılaştırılması

Değişkenler	*Posterior grup N=168*	*Anterior grup N=71*	*p*
Yaş, medyan (IQR)	54.93 (10.57)	59.82 (8.10)	**0.001** †
Cinsiyet, n (%)			
Erkek	122 (72.6)	51(71.8)	0.901‡
Kadın	46 (27.4)	20(28.2)	
VKİ, medyan (IQR)	29.24 (3.68)	28.09 (3.18)	**0.022** †
ASA skoru, medyan (IQR)	2 (1.8-2.05)	2 (2-2)	0.684 †
Tümör çapı, medyan (IQR)	4.29 (1.57)	3.69 (1.56)	**0.007** †
Nefrometri skoru, medyan (IQR)	5.65 (1.25)	5.31 (1.48)	0.073 †
Klinik evre n (%)			
T1a	11 (6.5)	10 (14.1)	**0.001**‡
T1b	112 (66.7)	29 (40.8)	
T1c	45 (26.8)	32 (45.1)	
Preoperatif KBH evresi, n (%)			
Evre 1	74 (44.0)	56 (78.9)	**<0.001**‡
Evre 2	86 (51.2)	13 (18.3)	
Evre 3	8 (4.8)	2 (2.8)	

VKİ= vücut kitle indeksi, IQR= çeyrekler arası uzaklık, KBH= kronik böbrek hasarı, †Mann-Whitney-U testi, ‡Ki-Kare testi.

Çizelge 4.1 incelendiğinde gruplar arasında farklılık oluşturan ve tedavinin sonuçlarını etkileyebilecek değişkenler bulunmaktadır. Bu değişkenler yaş, vücut kitle indeksi, klinik evre, KBH evresi ve tümör çapıdır.

4.1.1.2 Potansiyel Değişkenlerle Regresyon Modelinin Oluşturulması ve Eğilim Puanın Hesaplanması

Tek değişkenli analizlerde anlamlı bulunan değişkenlere ilave olarak anlamlı olmadığı halde klinik olarak önemli kabul edilen nefrometri skoru değişkeni ile lojistik regresyon modeli oluşturulmuştur. Oluşturulan modeldeki değişkenlerin olasılık oranları, bu oranların %95 güven aralıkları ve ilgili p değerleri Çizelge 4.2'de sunulmuştur.

Çizelge 4.2 Potansiyel değişkenlerle lojistik regresyon modeli oluşturulması

Prediktörler	*Odds Oranı*	*%95 Güven aralığı*		*P*
Yaş	1.046	1.014	1.081	**0.005**
Vücut kitle indeksi	0.941	0.860	1.027	0.178
Tümör boyutu	1.488	1.047	2.182	**0.033**
Nefrometri skoru	0.955	0.746	1.217	0.708
Preoperatif,KBH evresi				0.080
Evre 2(1)	0.415	0.148	1.168	0.096
Evre 3(2)	0.817	0.288	2.137	0.705
Klinik evre				**<0.001**
T1b (1)	0.099	0.030	0.323	**0.0001**
T2a (2)	0.044	0.004	0.447	**0.008**

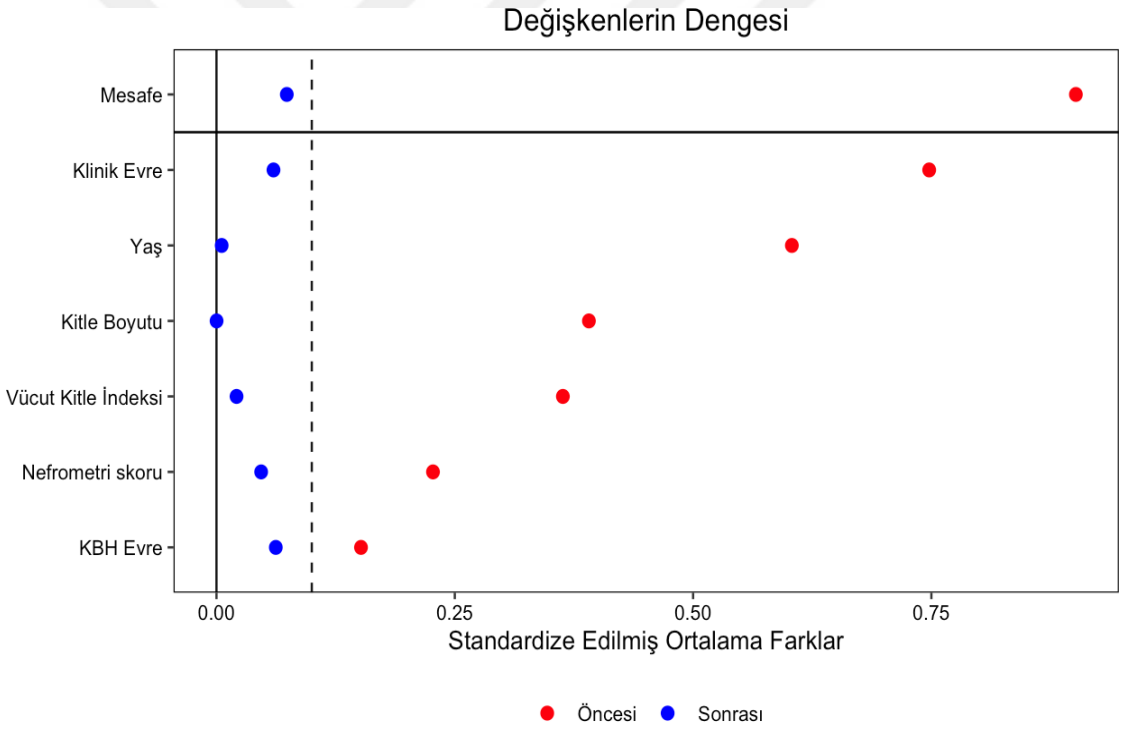
KBH= kronik böbrek hasarı.

Bu kurulan modelle hastaların gruplara atanma olasılığını temsil eden eğilim skoru değerleri elde edilmiştir. Elde edilen puanla eğilim skorunun 4 yöntemi veri setinde uygulanmıştır. Ancak, eğilim skoru yöntemleri uygulanmadan önce, sonuç değişkeni

üzerindeki tedavi etkisini ölçmek amacıyla bir ham model oluşturulmuş ve bu model, diğer yöntemlerin karşılaştırılmasında referans olarak kullanılmıştır. Ham model ile ilk değerlendirilen sonuç değişkeni tedavi sonrası trifekta durumudur. Yapılan lojistik regresyon analizinde; alınan tedavide kitlenin anterior yerleşimli olmasının trifekta oranına etkisinin olmadığı görülmüştür (β=-0.542, standart hata= 0.350, odds oranı: 0.582, P=0.122).

4.1.1.3 Eşleştirme Yönteminin Veri Setinde Uygulanması

Eşleştirme yöntemi için en yakın komşuluk modeli kaliper değeri 0.20 alınarak uygulanmıştır. Eşleştirme öncesi ve sonra değişkenlerin dengesi Şekil 4.1'de gösterilmiştir.



Şekil 4. 1 Eğilim skoru eşleştirme öncesi ve sonrası denge durumu

Şekil 4.1'de görüldüğü üzere, eşleştirme öncesinde değişkenlerin gruplar arasındaki dağılımı önemli farklılıklar göstermekte ve bu durum yanlılığa sebebiyet

verebilecek düzeydedir. Eşleştirme sonrası ise değişkenlerin gruplar arasındaki standardize edilmiş ortalama farklılıklarının 0.10'un altına düştüğü görülmüştür.

Çizelge 4.3 Eşleştirme sonrası değişkenlerin dengesi

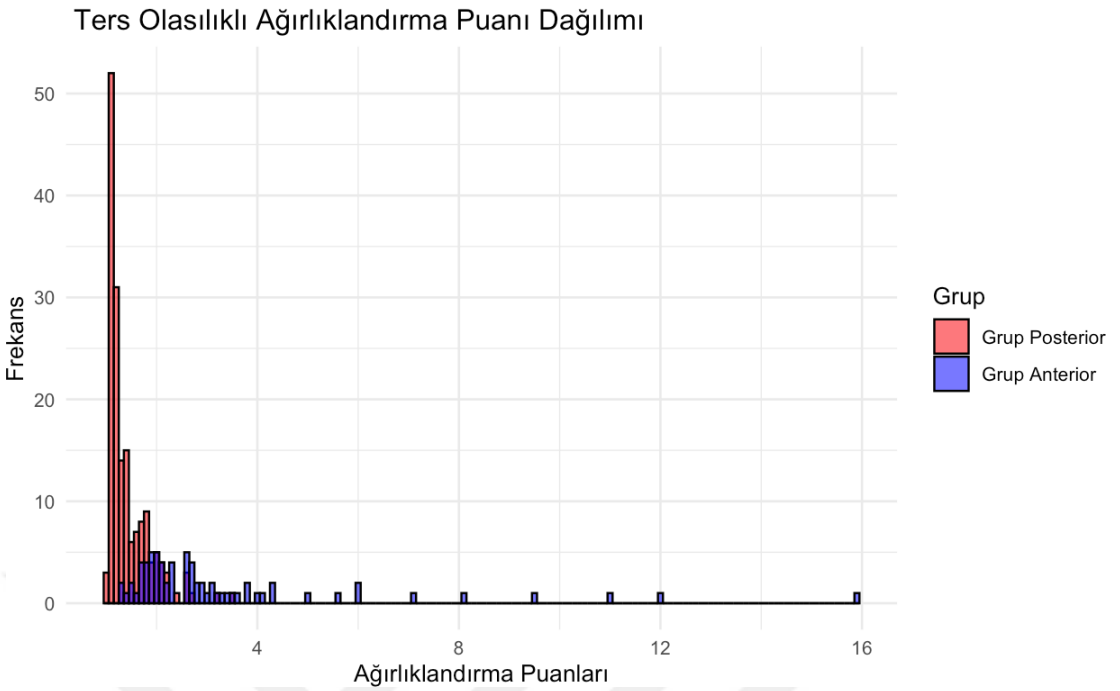
Değişkenler	*Posterior grup, n=68*	*Anterior grup, n=68*	P
Yaş, medyan(IQR)	59.82 (10.62)	59.87 (8.24)	0.978†
Cinsiyet, n (%)			
Erkek	50 (73.5)	48 (70.6)	0.702‡
Kadın	18 (26.5)	20 (29.4)	
VKİ, medyan (IQR)	28.14 (3.12)	28.20 (3.11)	0.901†
ASA skoru, medyan (IQR)	2 (0)	2 (0)	0.684
Tümör çapı, medyan (IQR)	3.61 (1.51)	3.61 (1.50)	1.000†
Nefrometri skoru, medyan (IQR)	5.38 (1.17)	5.31 (1.51)	0.765†
Klinik evre n (%)			
T1a	4 (5.9)	10 (14.7)	0.137‡
T1b	37 (54.4)	28 (41.2)	
T2a	27 (39.7)	30 (44.1)	
Preoperatif KBH evresi, n (%)			
Evre 1	50 (73.5)	53 (77.9)	0.621‡
Evre 2	17 (25.0)	13 (19.1)	
Evre 3	1 (1.5)	2 (2.9)	

VKİ= vücut kitle indeksi, IQR= çeyrekler arası uzaklık, KBH= kronik böbrek hasarı, †Mann-Whitney-U testi, ‡Ki-Kare testi.

Çizelge 4.3'de eşleştirme sonrası örneklem sayısının 239'dan 136'ya düştüğü görülmektedir. Eşleştirme sonrasında, olası karıştırıcı faktörlerin gruplar arasında istatistiksel olarak anlamlı bir farklılık yaratmadığı gözlemlenmiştir. Eşleştirme yönteminde tedavinin sonuç değişkenine olan etkisi incelendiğinde; β=-0.702, standart hata= 0.458, odds oranı: 0.496, P=0.126 verileri elde edilmiştir.

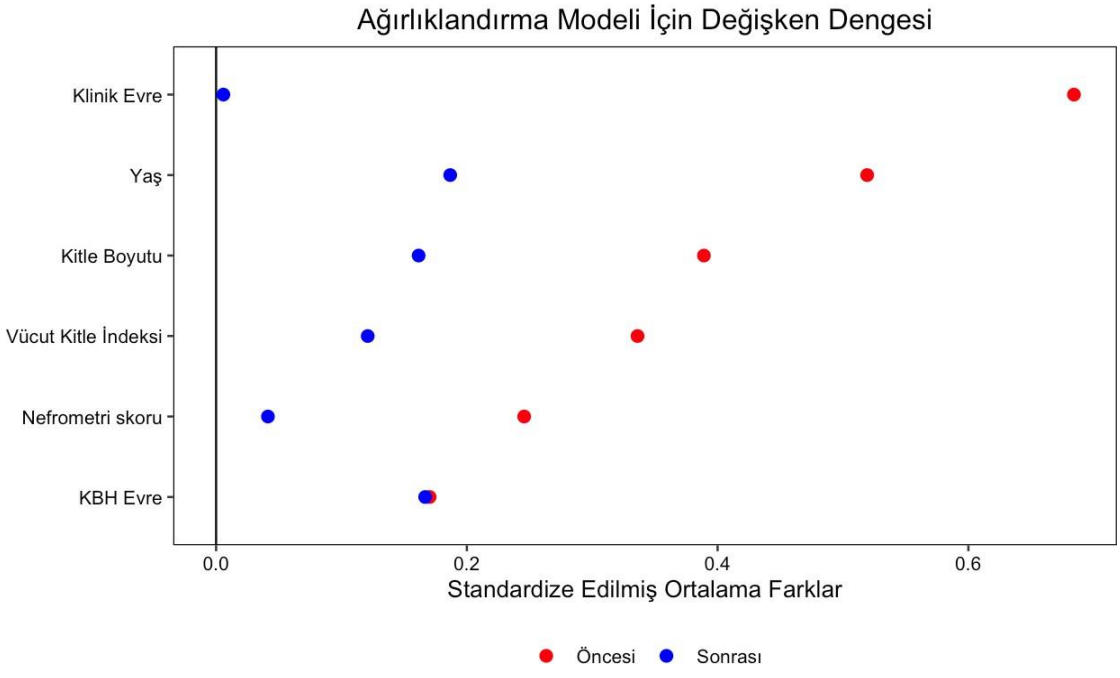
4.1.1.4 Ters Olasılıklı Ağırlıklandırma Yönteminin Veri Setinde Uygulanması

Elde edilen eğilim puanı sonrası ters olasılıklı ağırlıklandırma puanı elde edilmiştir. Şekil 4.2'de ters olasılıklı ağırlıklandırmaların frekansları histogram grafiği ile gösterilmiştir.



Şekil 4. 2 Ters olasılıklı ağırlıklandırma puanlarının histogram ile gösterilmesi

Ağırlıklandırma işlemi sonrası gruplar arasındaki dengenin sağlandığı gözlenmiştir. Bu denge Şekil 4.3'de görselleştirilmiştir.

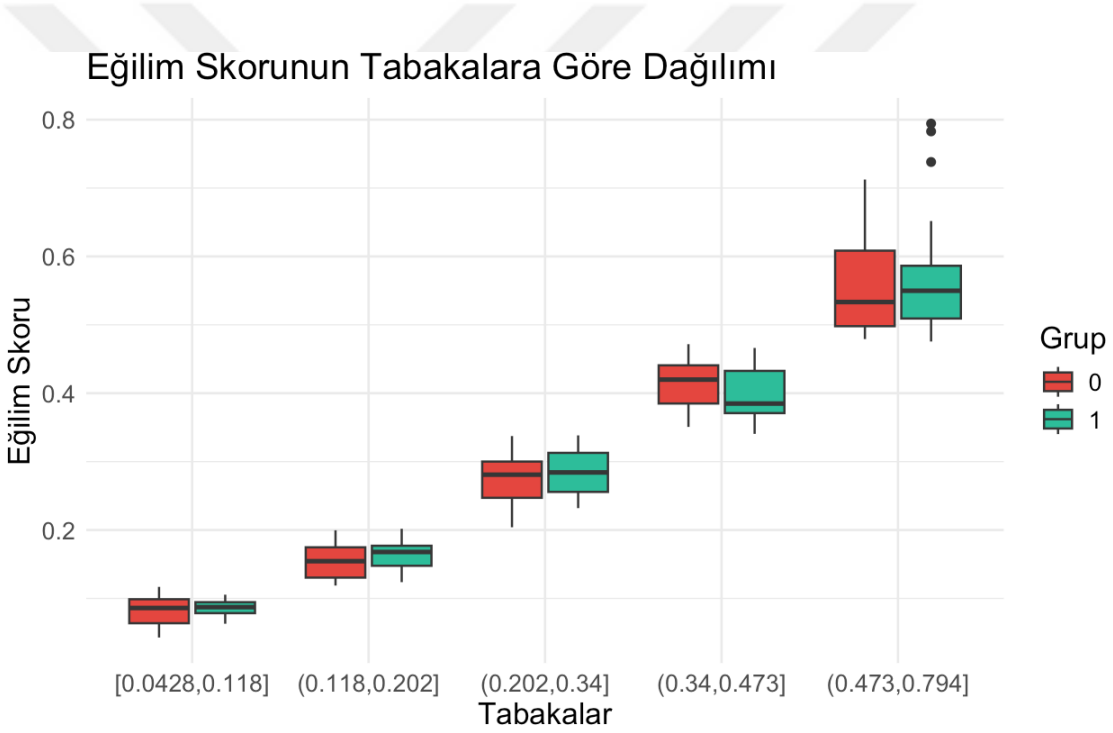


Şekil 4. 3 Değişkenlerin ağırlıklandırma sonrası dengesi

Kurulan denge sonrası yapılan regresyon analizinde ters olasılıklı ağırlıklandırma modelinin çıktıları; β=-0.980, standart hata= 0.426, P=0.022 olarak bulunmuştur. Odds oranı 0.375 olarak hesaplanmıştır. Anterior yerleşimli kitlelerin posterior yerleşimli kitlelere kıyasla trifekta oranını istatistiksel olarak anlamlı düzeyde azalttığı saptanmıştır.

4.1.1.5 Tabakalama Yönteminin Veri Setinde Uygulanması

Tabakalama metodunda eğilim skoruna göre veri seti 5 tabakaya ayrılmıştır. Her tabaka kendi içinde analiz edilerek, ortalama tedavi etkisi ortaya çıkarılmıştır. Tabakalara ayırma sonrası gruplar arasında oluşan denge Şekil 4.4'de gösterilmiştir.



Şekil 4. 4 Tabakalandırma sonrası gruplar arasında oluşan eğilim puanı dengesi

Tabakalandırma işlemi sonrası uygulanan lojistik regresyon analizi sonucunda; β=-0.882, standart hata= 0.398, odds oranı: 0.414, P=0.027 olarak bulunmuştur. Tabakalandırma yöntemi, ağırlıklandırma yöntemiyle paralel olarak anterior yerleşimli kitlelere yapılan cerrahinin trifekta oranını azalttığını ortaya koymuştur.

4.1.1.6 Regresyon Düzeltmesi Yönteminin Veri Setinde Uygulanması

Bu yöntemde elde edilen eğilim skoru regresyon modeline dahil edilerek tedavi etkisi hesaplanmıştır. Bu analizde elde edilen sonuçlar; β=-0.755, standart hata= 0.390, odds oranı: 0.470, P=0.053 bulunmuştur. İstatistiksel anlamlılık sınırda olmakla beraber, kitle lokalizasyonun trifekta oranına etkisi gruplar arasında farklılık saptanmamıştır.

4.1.1.7 Modellerin Performanslarının Karşılaştırılması

Eğilim skoru kullanılarak oluşturulan dört model ile ham modelin performansları, elde edilen lojistik regresyon analizi sonuçları üzerinden karşılaştırılmıştır. Ayrıca, yöntemlerin performansını değerlendirmek için ROC analizi kullanılmış ve bu bölümde sunulmuştur. Bu karşılaştırmalarla, modellerin tahmin gücünü ve dengelenme başarısını detaylı bir şekilde değerlendirmek amaçlanmıştır.

4.1.1.7.1 Lojistik Regresyon Sonuçları

Elde edilen lojistik regresyon analiz sonuçları, farklı eğilim skoru yöntemlerinin kitlenin bulunduğu lokalizasyona göre sonuç değişkeni olan trifekta üzerinde etkisinin nasıl değiştiğini göstermektedir (Çizelge 4.4).

Çizelge 4.4 Lojistik regresyon analizi ile yöntemlerin performanslarının karşılaştırılması

Modeller	*Beta*	*SH*	*P*	*OR*	*%95 güven aralığı*
Ham	-0.542	0.350	0.122	0.582	0.294-1.170
Eşleştirme	-0.702	0.458	0.126	0.496	0.195-1.195
Ağırlıklandırma	-0.980	0.426	**0.022**	0.375	0.163-0.868
Tabakalandırma	-0.882	0.398	**0.027**	0.414	0.189-0.903
Regresyon düzeltmesi	-0.755	0.390	0.053	0.470	0.219-1.014

SH=standart hata, OR= odds oranı.

Ham analiz sonuçlarına göre, kitle yerleşimi lokalizasyonunun trifekta sonuçlarına etkisi istatistiksel olarak anlamlı değildir (OR = 0.582, %95 GA: 0.294-1.170, p = 0.122). Eşleştirme yöntemi uygulandığında, olasılık oranı azalmış olsa da istatistiksel anlamlı farklılık bulunamamıştır (OR = 0.496, %95 GA: 0.195-1.195, p = 0.126). Tabakalandırma yöntemi veri setine uygulandığında, kitle lokalizasyonunun etkisi istatistiksel olarak anlamlı hale gelmiştir (OR = 0.414, %95 GA: 0.189-0.903, p = 0.027).

Ağırlıklandırma yöntemi kullanıldığında, en düşük odds oranı elde edilmiştir ve bu istatistiksel açıdan anlamlıdır (OR = 0.375, %95 GA: 0.163-0.868, p = 0.022). Son olarak, regresyon düzeltmesi yöntemi ile yapılan analizde istatistiksel anlamlılık sınırında bir sonuç elde edilmiştir (OR = 0.470, %95 GA: 0.219-1.014, p = 0.053). Bu sonuçlar ağırlıklandırma ve tabakalandırma yöntemlerinde kitle lokalizasyonun anteriorda olmasının trifekta oranını azalttığını göstermiştir.

4.1.1.7.2 ROC Analizi Sonuçları

Modellerin performans ölçütlerinin karşılaştırılmasında diğer bir yöntem ROC analizi ile EAA hesaplanması olmuştur. Bu analizlerde tabakalandırma modeli 0.611 ile en yüksek eğri altında kalan alana sahiptir. En yüksek duyarlılık regresyon düzeltmesi, en yüksek özgüllük eşleştirme modeline ait olduğu görülmüştür. Yöntemlerin sonuçları Çizelge 4.5'de sunulmuştur.

Çizelge 4.5 Parsiyel nefrektomi veri setinde yöntemlerin ROC analizi ile karşılaştırılması

Modeller	*EAA*	*%95 güven aralığı*
Ham	0.587	0.524- 0.650
Eşleştirme	0.585	0.503- 0.666
Ağırlıklandırma	0.587	0.524- 0.650
Tabakalandırma	0.611	0.540- 0.682
Regresyon düzeltmesi	0.583	0.509- 0.658

EAA= eğri altındaki alan

4.1.2 Osteoartrit Veri Seti ile Eğilim Skoru Modellerinin Karşılaştırılması

4.1.2.1 Değişkenlerin Gruplar Arasındaki Dağılımını Belirleme ve Potansiyel Karıştırıcıların Saptanması

Osteoartrit veri setinde, yaş ve vücut kitle indeksi sürekli değişkenler olarak, cinsiyet, sigara kullanımı ve ırk ise kategorik değişkenler olarak tanımlanmıştır. Gruplar, osteoporoz tedavisi almayanlar (0) ve osteoporoz tedavisi alanlar (1) şeklinde

kodlanmıştır. Sonuç değişkeni ise, hastalarda kronik osteoartritin gelişmesi durumu olarak belirlenmiştir.

Analizin ilk aşamasında, gruplar arasında potansiyel karıştırıcı faktörlerin ikili karşılaştırması yapılmış ve bu karşılaştırma sonuçları Çizelge 4.6'da sunulmuştur. Bu adım, gruplar arasındaki başlangıç farklılıklarını anlamak ve eğilim skoru yöntemlerinin uygulanması öncesinde dengelenmesi gereken değişkenleri belirlemek amacıyla gerçekleştirilmiştir.

Çizelge 4.6 Osteoporoz tedavi durumuna göre değişkenlerin karşılaştırılması

Değişken	**Osteoporoz (-) N=1921**	**Osteoporoz (+) N=442**	**p**
Yaş	68.24 (5.32)	69.02 (5.62)	**0.006**†
Cinsiyet, n (%)			
Erkek	886 (46.1)	31 (7.0)	**<0.001**‡
Kadın	1035 (53.9)	411 (93.0)	
VKİ, medyan (IQR)	28.84 (4.47)	26.15 (4.28)	**<0.001**†
Irk, n (%)			
Beyaz	30 (1.6)	3 (0.7)	
Afro-Amerikan	1581 (82.3)	398 (90.0)	**0.001**‡
Asyalı	295 (15.4)	38 (8.6)	
Karışık	15 (0.8)	3 (0.7)	
Sigara kullanımı, n (%)			
İçmeyen	926 (48.2)	241 (54.5)	**0.019**‡
İçen	995 (51.8)	201 (45.5)	

VKİ=Vücut kitle indeksi, IQR=çeyrekler arası uzaklık, †Mann-Whitney-U testi, ‡Ki-Kare testi.

Çizelge 4.6'da görüldüğü üzere osteoporoz (+) grubunda bireylerin yaş ortalaması, osteoporoz (-) grubuna kıyasla anlamlı düzeyde daha yüksek bulunmuştur (p=0.006). Cinsiyet dağılımı açısından, osteoporoz (+) grubunda kadın oranı belirgin şekilde daha yüksek olup, bu fark istatistiksel olarak anlamlıdır (p<0.001). VKİ osteoporoz (+) grubunda daha düşük olup, bu farklılık istatistiksel olarak anlamlıdır (p<0.001). Irk dağılımı incelendiğinde, Afro-Amerikan bireylerin osteoporoz tedavisi alan grupta daha yüksek oranda temsil edildiği görülmüştür (p=0.001). Sigara kullanımı açısından, osteoporoz (+) grubunda sigara kullanmayanların oranı daha yüksek bulunmuş

olup, bu fark istatistiksel olarak anlamlıdır (p=0.019). Bulgular, osteoporoz tedavi grupları arasında yaş, cinsiyet, VKİ, ırk ve sigara kullanımının istatistiksel olarak farklı olduğunu göstermiştir.

Bu sonuçlar temel alınarak, bir lojistik regresyon modeli kurulmuş ve bu modelden eğilim skoru elde edilmiştir. Eğilim skoru modelleri oluşturulmadan önce, osteoporoz tedavisi almanın kronik osteoartrit gelişimi üzerindeki etkisi hesaplanmış ve bu hesaplama sonucunda ham model çıktıları elde edilmiştir. Ham model, herhangi bir dengelenme işlemi yapılmadan önce tedavinin veya risk faktörünün sonuç değişkeni üzerindeki etkisini değerlendirmek için kullanılmıştır.

4.1.2.2 Ham Model ile Elde Edilen Sonuçlar

Ham model sonuçlarına göre, osteoporoz tedavisinin kronik osteoartrit gelişimi üzerindeki etkisi istatistiksel olarak anlamlı bulunmuştur (P<0.001). Katsayı değeri (β=-0.398) negatif yönlü bir ilişki göstermiştir, bu da osteoporoz tedavisi almanın osteoartrit gelişme olasılığını azalttığını göstermiştir. Standart hata değeri 0.107 olarak hesaplanmıştır. Negatif yönlü ilişki, osteoporoz tedavisinin osteoartrit riskini azaltıcı bir faktör olabileceğine işaret etmiştir.

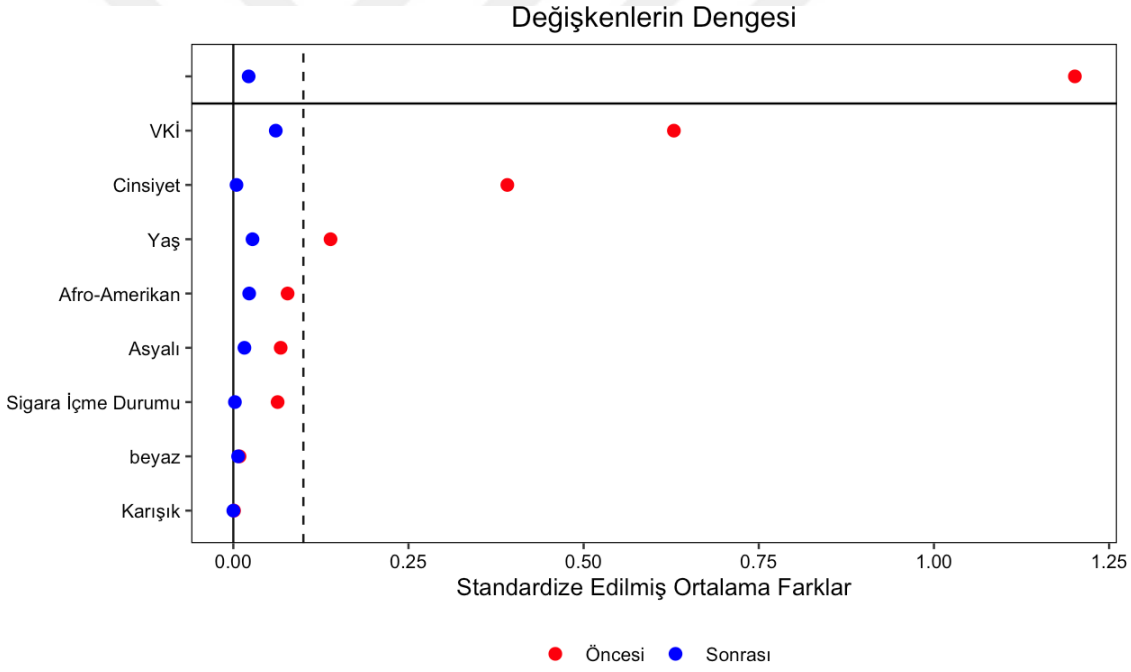
4.1.2.3 Eşleştirme Yönteminin Uygulanması ve Sonuçların Elde Edilmesi

Eşleştirme yönteminde en yakın komşuluk eşleştirme ve kaliper (0.2) yöntemi kullanılmıştır. Eşleştirme sonrası her grupta 442 hasta olmak üzere 884 hastalık yeni bir eşleştirilmiş veri seti elde edilmiştir. Eşleştirme sonrası değişkenlerin gruplar arasında dağılımın dengelendiği gözlenmiştir. Değişkenlerin ortalama farklılıklarının ortadan kaldırıldığı Çizelge 4.7 ve Şekil 4.5'de gösterilmiştir.

Çizelge 4.7 Eşleştirme yöntemi sonrası değişkenlerin gruplar arasında karşılaştırılması

Değişken	*Osteoporoz (-) N=442*	*Osteoporoz (+) N=442*	*p*
Yaş, medyan (IQR)	68.87 (5.34)	69.02 (5.62)	0.677†
Cinsiyet, n (%)			
Erkek	29 (6.6)	32 (7)	0.894‡
Kadın	413 (93.4)	411 (93.0)	
VKİ, medyan (IQR)	26.41 (4.32)	26.15 (4.28)	0.370†
Irk, n (%)			
Beyaz	0 (0.0)	3 (0.7)	
Afro-Amerikan	408 (92.3)	398 (90.0)	0.280‡
Asyalı	31 (7.0)	38 (8.6)	
Karışık	3 (0.7)	3 (0.7)	
Sigara kullanımı, n (%)			
İçmeyen	3 (0.7)	3 (0.7)	0.999‡
İçen	200 (45.2)	201 (45.5)	

VKİ=vücut kitle indeksi, †Mann-Whitney-U testi, ‡Ki-Kare testi, IQR=çeyrekler arası uzaklık

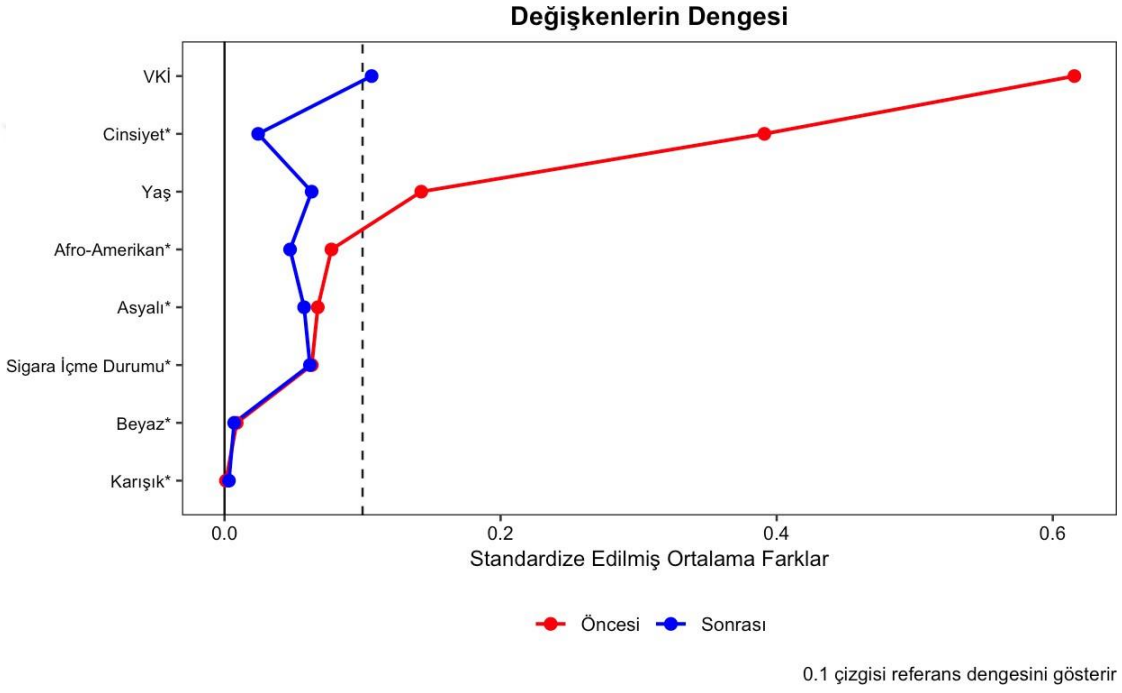


Şekil 4. 5 Eşleştirme sonrası ortalama farkların gösterilmesi

Eşleştirme yöntemi sonrasında yapılan lojistik regresyon analizinde, osteoporoz tedavisi almanın kronik osteoartrit gelişimi üzerindeki etkisinin istatistiksel anlamlılığını kaybettiği görülmüştür (p=0.278). Beta katsayısı -0.147, odds oranı 0.863 ve standart hatası 0.136 olarak bulunmuştur. Eşleştirme sonrası elde edilen verilere göre, osteoporoz tedavisinin osteoartrit riskini istatistiksel olarak artırmadığı sonucuna varılmıştır.

4.1.2.4 Ters Olasılıklı Ağırlıklandırma Yöntemi

Elde edilen eğilim skorları daha önce bahsedildiği üzere hesaplanarak veri setine eklenmiştir. Yapılan analiz öncesi ve sonrası ters olasılıklı ağırlıklandırma puanlarının dağılımı Şekil 4.6' da gösterilmiştir.

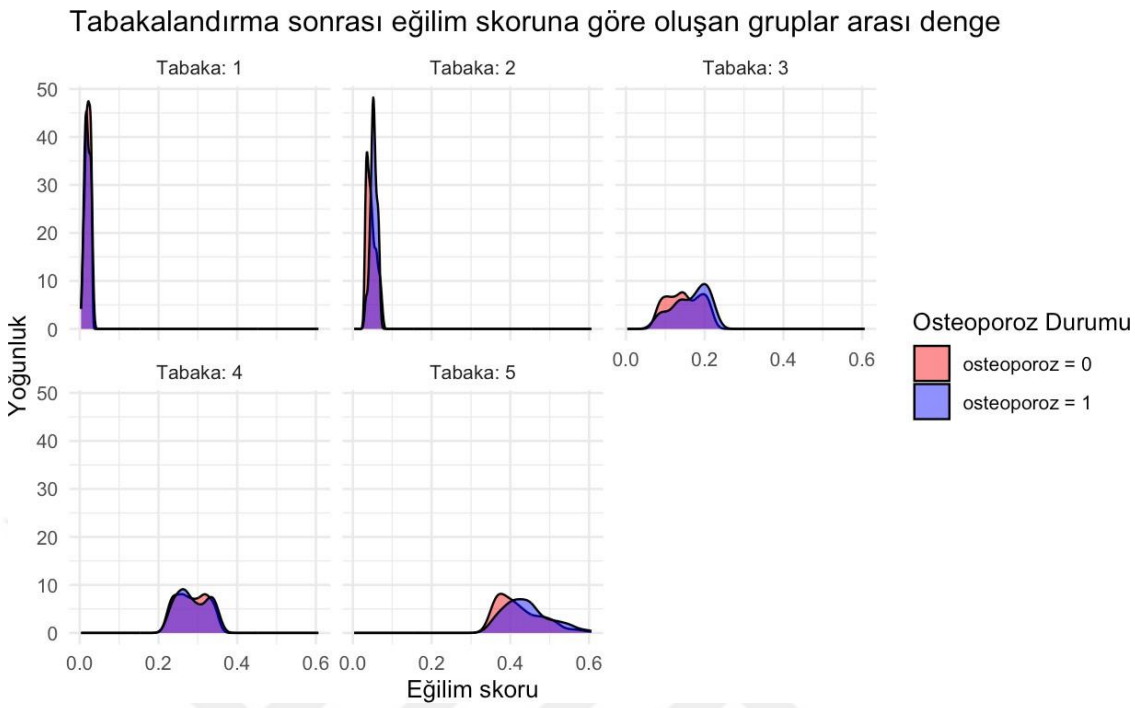


Şekil 4. 6 Ters olasılıklı ağırlıklandırma öncesi ile sonrasında oluşan gruplar arasındaki denge

Ters olasılıklı ağırlıklandırma puanının elde edilmesi sonrası kurulan lojistik regresyon modelinde osteoporoz tedavisinin kronik osteoartrit üzerine olan etkisi istatistiksel olarak anlamlı bulunmamıştır (P=0.264). Beta değeri -0.299, odds oranı 0.796, standart hata değeri 0.204 olarak hesaplanmıştır.

4.1.2.5 Tabakalandırma Sonrası Osteoartrit Veri Setinde Elde Edilen Bulgular

Tabakalandırma yönteminde ilk veri setinde olduğu gibi 5 tabaka kullanılmıştır. Eğilim puanına göre elde edilen tabakaların kendi içlerinde dengesi Şekil 4.7'de gösterilmiştir



Şekil 4. 7 Tabakalarda oluşan eğilim skoru dengesi

Kurulan lojistik regresyon modeli sonrasında, tabakaların birleştirilmiş etkisine ilişkin sonuçlar beta değeri: -0.196, odds oranı: 0.822 ve standart hata: 0.117 olarak bulunmuştur. Osteoporoz tedavisinin kronik osteoartrit gelişimi üzerindeki etkisi, istatistiksel açıdan anlamlı bulunmamıştır (p=0.094). Bu sonuç, osteoporoz tedavisinin osteoartrit riskini istatistiksel anlamlı olarak azaltmadığını göstermiştir.

4.1.2.6 Regresyon Düzeltmesi

Elde edilen eğilim skorunun lojistik regresyon analizine dahil edilmesi sonrası elde edilen çıktılarda; Beta değeri -0.160, odds oranı 0.852, standart hata 0.117 olarak hesaplanmıştır. Diğer eğilim skoru yöntemlerinde olduğu gibi osteoporoz tedavisi kronik osteoartrit gelişimi arasında istatistiksel anlamlı bir ilişki saptanmamıştır (p=0.172).

4.1.2.7 Osteoartrit Veri Setinde Uygulanan Eğilim Skoru Yöntemlerinin Karşılaştırılması

Osteoporoz tedavisi almanın kronik osteroartrit veri setinde uygulanan eğilim skoru yöntemlerinin sonuçları Çizelge 4.8' de özetlenmiştir. Bu sonuçlara göre;

- Osteoporoz tedavisi almak ham modelde istatistiksel anlamlı olarak osteoartrit gelişimini azaltmıştır.
- Eğilim skoru modelleri uygulanması sonrası elde edilen ß değerlerinin istatistiksel açıdan anlamlı olmaması, literatürde bu tedavinin osteoartrit görülmesine etki etmediği görüşünü desteklemektedir.
- Eğilim skoru yöntemleri arasında en düşük standart hata değerleri tabakalandırma ve regresyon düzeltmesi modellerinde elde edilmiştir.

Çizelge 4.8 Osteoartrit veri setine uygulanan eğilim skoru yöntemlerinin performans karşılaştırılması

Model	*Beta*	*SH*	*P*	*Odds oranı*	*%95 güven aralığı*	
Ham	-0.398	0.107	**<0.001**	0.672	0.545	0.828
Eşleştirme	-0.147	0.136	0.278	0.863	0.661	1.126
Ağırlıklandırma	-0.229	0.204	0.264	0.796	0.533	1.188
Tabakalandırma	-0.196	0.117	0.094	0.822	0.653	1.034
Regresyon düzeltmesi	-0.160	0.117	0.172	0.852	0.678	1.071

SH=standart hata

Çizelge 4.9'da görüldüğü üzere modelleme yöntemlerinin (Ham model, eşleştirme, tabakalandırma, ağırlıklandırma ve regresyon düzeltmesi) EAAve güven aralıkları karşılaştırılmıştır. Çizelge 4.9'da sunulan sonuçlara göre;

- Tabakalandırma yöntemi en yüksek eğri altında kalan alana sahiptir.
- En yüksek duyarlılık ağırlıklandırma modelinde gözlenmiştir.
- Öte yandan, en yüksek özgüllük tabakalandırma modelinde elde edilmiştir.

Çizelge 4.9 Osteoartrit veri setinde modellerin ROC analizi ile karşılaştırılması

Model	*EAA*	*%95 güven aralığı*
Ham	0.530	0.514-0.546
Eşleştirme	0.518	0.485-0.552
Ağırlıklandırma	0.530	0.514-0.546
Tabakalandırma	0.594	0.569-0.614
Regresyon düzeltmesi	0.566	0.544-0.590

EAA= eğri altındaki alan

4.2 Sürekli Yapıdaki Sonuç Değişkenin Yöntem Karşılaştırmalarına Etkisi

Sonuç değişkeninin sürekli olduğu senaryoda parsiyel nefrektomi veri seti kullanılmıştır. Kitle lokalizasyonun operasyon süresine olan etkisi bu bölümde incelenmiştir. Sonuç değişkeni sürekli bir değişken olduğu için lineer regresyon modeli ile veriler analiz edilmiştir. Analizlerin beta katsayıları, standart hataları, P değerleri ve %95 güven aralıkları elde edilmiştir. Bölüm 4.1'de kullanılan veri seti olduğu için modellerin ayrıntılı bilgisine yer verilmemiştir. Tek değişkenli analizde posterior grupta ortalama operasyon süresi 80.9 ± 11.4, anterior grup için 89.7 ± 11.3 dakika olup, aradaki farklılık istatistiksel olarak anlamlıdır (P<0.001). Posterior cerrahi grubunda operasyon süresinin daha kısa olduğu görülmektedir. Bu farklılığın eğilim skoru modelleri uygulanması sonrası nasıl bir değişiklik gösterdiği irdelenmiştir. Modellerin uygulanması sonrası alınan lineer regresyon analizi sonuçları Çizelge 4.10'da özetlenmiştir.

Çizelge 4.10 Tedavi gruplarının operasyon süresi üzerine olan etkisi

Model	R^2	*Beta*	*SH*	*P*	*%95 güven aralığı*	
Ham	0.110	8.750	1.614	**<0.001**	5.570	11.931
Eşleştirme	0.194	10.426	1.835	**<0.001**	6.797	14.056
Ağırlıklandırma	0.146	10.291	2.442	**<0.001**	5.480	15.102
Tabakalandırma	0.133	10.373	1.747	**<0.001**	6.931	13.815
Regresyon düzeltmesi	0.134	10.454	1.730	**<0.001**	7.046	13.862

SH=standart hata

Çizelge 4.10 incelendiğinde;

- Tüm modellerin istatistiksel olarak anlamlı bir şekilde sonuç değişkenini tahmin edebildiği görülmüştür.

- Regresyon düzeltmesi modelinin, eğilim skoru yöntemleri arasında en düşük standart hataya ve en dar güven aralığına sahip olduğu belirlenmiştir. R^2 değeri en yüksek yöntemin ise eşleştirme modelinde olduğu görülmüştür.

4.3 Örneklem Büyüklüğünün Yöntem Karşılaştırmalarına Etkisi

4.3.1 Alt Veri Setlerinin Elde Edilmesi ve Tanıtılması

Bu senaryoda osteoartrit veri seti kullanılmıştır. Sonuç değişkeni olan kronik osteoartrit oranı sabit tutularak (%50) sırasıyla 100, 500, 1000 örneklem veri setinden çekilerek alt veri setleri oluşturulmuştur. Bölüm 4.1.2'de yapılan analizler her bir alt veri setinde uygulanmıştır. İlk olarak görece düşük örneklem sayılı (100) veri setinin analizleri yapılmıştır.

4.3.1.1 Örneklem Sayısı 100 Olan Veri Setinde Model Performansları

Çizelge 4.11'de görüldüğü üzere modeller istatistiksel olarak anlamlı sonuçlara sahip değildir. Ham ve eşleştirme modelinde negatif yönlü bir ilişki saptanır iken, ağırlıklandırma, tabakalandırma ve regresyon düzeltmesi modelleri pozitif bir orana sahip olduğu görülmüştür. Modeller içerisinde ağırlıklandırma yönteminin en düşük standart hata ve dar güven aralığına sahip olduğu görülmüştür.

Çizelge 4.11 Örneklem sayısı 100 olan alt veri setinde eğilim skoru yöntemlerinin sonuçları

Model	*Beta*	*SH*	*P*	*Odds oranı*	*%95 güven aralığı*	
Ham	-0.393	0.515	0.446	0.675	0.239	1.842
Eşleştirme	-0.213	0.653	0.744	0.808	0.220	2.920
Ağırlıklandırma	0.205	0.322	0.524	1.227	0.654	2.315
Tabakalandırma			0.438	1.630	0.478	5.794
Regresyon düzeltmesi	0.265	0.590	0.654	1.303	0.406	4.208

SH=standart hata

ROC analiz sonuçlarına göre, yöntemlerin hesaplanan performans ölçütleri arasında belirgin farklılıklar gözlemlenmiştir. Ham model, yüksek duyarlılık göstermesine rağmen özgüllük açısından zayıf bir performans sergilemiştir. Eşleştirme

yöntemi, duyarlılık ve özgüllük açısından dengeli bir dağılım gösterse de ayırt ediciliği sınırlı kalmıştır. Ağırlıklandırma yöntemi, maksimum duyarlılık elde etmesine karşın özgüllük açısından yetersiz bulunmuştur. Tabakalandırma yöntemi hem duyarlılık hem de özgüllük açısından daha dengeli bir performans sergileyerek özgüllüğü en yüksek yöntemlerden biri olmuştur. Regresyon düzeltmesi yöntemi de yüksek duyarlılık ve orta derece de özgüllükle diğer yöntemlere göre tabakalandırma yöntemi ile beraber daha iyi bir performans göstermiştir. Yöntemlerin EAA değerleri ve güven aralıkları çizelge 4.12'de sunulmuştur.

Çizelge 4.12 Örneklem sayısı 100 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları

Model	*EAA*	*%95 güven aralığı*
Ham	0.530	0.452-0.607
Eşleştirme	0.526	0.362-0.690
Ağırlıklandırma	0.470	0.392-0.547
Tabakalandırma	0.708	0.611-0.813
Regresyon düzeltmesi	0.636	0.524-0.754

EAA=eğri altındaki alan

4.3.1.2 Örneklem Sayısı 500 Olan Veri Setinde Model Performansları

Çizelge 4.13'te incelendiğinde yöntemlerin hiçbiri istatistiksel olarak anlamlı sonuçlara sahip değildir. Ağırlıklandırma yöntemi, düşük standart hatası ve dar güven aralığı ile daha güçlü bir ilişki gösterdiği görülmüştür.

Çizelge 4.13 Örneklem sayısı 500 olan alt veri setinde eğilim skoru yöntemlerinin sonuçları

Model	*Beta*	*SH*	*P*	*Odds oranı*	*%95 güven aralığı*	
Ham	-0.319	0.223	0.151	0.727	0.468	1.122
Eşleştirme	-0.039	0.281	0.888	0.961	0.554	1.667
Ağırlıklandırma	0.154	0.126	0.220	1.167	0.912	1.494
Tabakalandırma	-0.086	0.243	0.723	0.917	0.568	1.478
Regresyon düzeltmesi	-0.161	0.245	0.511	0.852	0.526	1.375

SH=standart hata

Çizelge 4.14 incelendiğinde en yüksek eğri altında kalan alana tabakalandırma yönteminin sahip olduğu görülmüştür. Duyarlılığı en yüksek model ağırlıklandırma iken,

özgüllük konusunda zayıf bir performans gösterdiği saptanmıştır. Bu bulgulara göre tabakalandırma yöntemi özgüllüğünün ve eğri altında kalan alanın üstünlüğü ile diğer yöntemlere göre daha iyi bir performans göstermiştir.

Çizelge 4.14 Örneklem sayısı 500 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları

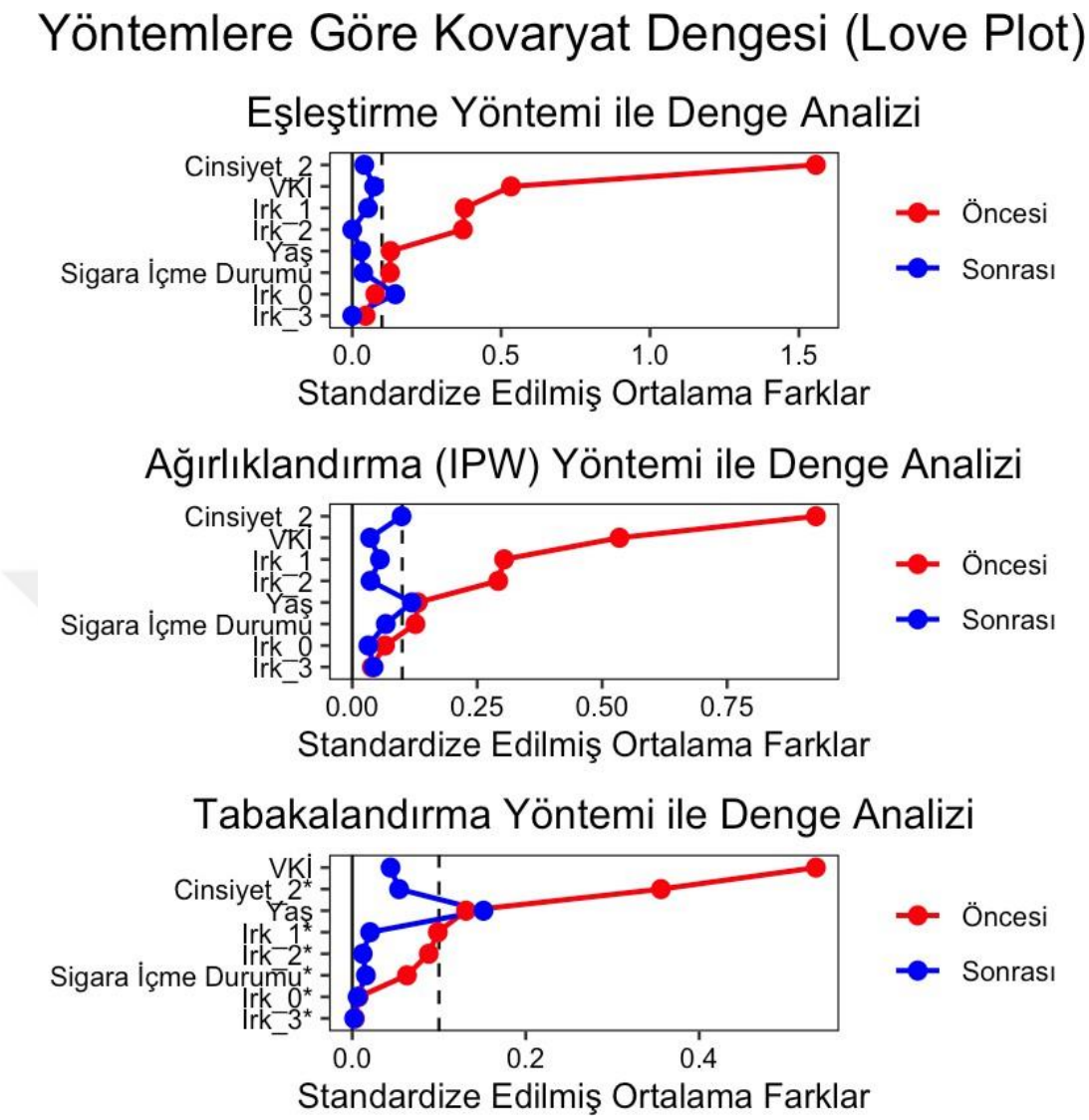
Model	*EAA*	*%95 güven aralığı*
Ham	0.538	0.503-0.561
Eşleştirme	0.550	0.481-0.620
Ağırlıklandırma	0.538	0.503-0.561
Tabakalandırma	0.600	0.553-0.651
Regresyon düzeltmesi	0.547	0.517-0.588

EAA=eğri altındaki alan

4.3.1.3 Örneklem Sayısı 1000 Olan Veri Setinde Model Performansları

Yöntemlerin standardize edilmiş ortalama farkları dengelemedeki performansı Şekil 4.8'de gösterilmiştir. Eşleştirme ortalamalar arası farkları en düşük seviyeye indirgeyen yöntem olduğu görülmüştür.

Çizelge 4.15 incelendiğinde osteoporoz tedavisinin kronik osteoartrit gelişimi üzerine etkisi ham modelde istatistiksel anlamlı olarak saptanmıştır. Örneklem sayısı arttıkça gerçek veri setine yakın sonuçlar elde edilmiştir. Diğer yöntemler istatistiksel açıdan anlamsız olup, osteoporoz tedavisinin kronik osteoartrit gelişimi üzerine etkisinin olmadığını göstermiştir.



Şekil 4. 8 1000 örneklemli alt veri setinde eğilim skoru yöntemleri uygulanması sonrası oluşan denge

Çizelge 4.15 Örneklem sayısı 1000 olan alt veri setinde eğilim skoru yöntemlerinin sonuçları

Model	*Beta*	*SH*	*P*	*Odds oranı*	*%95 güven aralığı*	
Ham	-0.391	0.155	**0.012**	0.676	0.498	0.915
Eşleştirme	-0.150	0.193	0.440	0.861	0.589	1.258
Ağırlıklandırma	-0.141	**0.091**	0.120	0.868	**0.726**	**1.038**
Tabakalandırma	-0.132	0.172	0.443	0.877	0.626	1.228
Regresyon düzeltmesi	-0.077	0.171	0.653	0.926	0.662	1.295

SH=standart hata

Çizelge 4.16 Örneklem sayısı 1000 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları

Model	*EAA*	*%95 güven aralığı*
	0.533	0.507-0.558
Eşleştirme	0.519	0.362-0.690
Ağırlıklandırma	0.533	0.507-0.558
Tabakalandırma	0.602	0.567-0.636
Regresyon düzeltmesi	0.581	0.545-0.616

EAA=eğri altındaki alan

Örneklem sayısının 1000 olduğu veri setinin ROC analizi sonuçlarında diğer örneklemlerde olduğu gibi tabakalandırma en yüksek eğri altında kalan alana sahip olduğu görülmüştür. Ağırlıklandırma ve ham model yüksek duyarlılıkla beraber düşük oranda özgüllük performansı göstermiştir.

4.4 İkili Yapıdaki Sonuç Değişkenin Görülme Oranının Yöntem Karşılaştırmalarına Etkisi

4.4.1 Veri Setinin Elde Edilmesi

Osteoartrit veri setinde ikili yapıdaki sonuç değişkeni olan kronik osteoartritin görülme oranı 1/3 olacak şekilde veri seti üretilmiştir. 1/3 oranı sabit tutularak sırasıyla 100, 500, 1000 örneklem sayılı alt veri setleri oluşturulmuştur.

4.4.1.1 Sonuç Değişkenin 1/3 Oranında Görüldüğü, Örneklem Sayısı 100 Olan Veri Seti

Farklı yöntemlerle elde edilen katsayılar ve olasılık oranları incelendiğinde, ham model ve eşleştirme yöntemleri tedavi değişkenin yüksek etkisini gösterirken, ağırlıklandırma ve regresyon düzeltmesi yöntemleri daha düşük etkilere işaret etmiştir. İstatistiksel açısından hiçbir model anlamlı bir ilişkiye sahip bulunmamıştır. ROC analizi sonuçlarına göre, tabakalandırma yöntemi en yüksek EAA değerine ulaşarak sınıflandırma performansı açısından en iyi sonucu vermiştir. Regresyon düzeltmesi yöntemi de diğer yöntemlere kıyasla yüksek bir EAA değeri ile dikkat çekerken, ham model ve ağırlıklandırma yöntemleri yüksek duyarlılık ve düşük özgüllüğe sahip olduğu görülmüştür.

Çizelge 4.17 Örneklem sayısının 100 olduğu senaryoda modellerin karşılaştırmaları

Model	*Beta*	*SH*	*P*	*Odds oranı*	*%95 güven aralığı*	
Ham	-0.591	0.609	0.332	0.554	0.173	1.963
Eşleştirme	-0.318	0.801	0.691	0.727	0.143	3.512
Ağırlıklandırma	-0.046	**0.416**	0.913	0.955	**0.429**	**2.218**
Tabakalandırma	-0.383	0.745	0.607	0.682	0.155	3.033
Regresyon düzeltmesi	-0.077	0.735	0.917	0.926	0.223	4.173

SH=standart hata

Çizelge 4.18 Örneklem sayısı 100 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları

Model	*EAA*	*%95 güven aralığı*
Ham	0.542	0.448 - 0.637
Eşleştirme	0.540	0.336 - 0.744
Ağırlıklandırma	0.542	0.448 - 0.637
Tabakalandırma	0.592	0.461 - 0.723
Regresyon düzeltmesi	0.574	0.434 - 0.715

EAA=eğri altındaki alan

4.4.1.2 Sonuç Değişkenin 1/3 Oranında Görüldüğü Örneklem Sayısı 500 Olan Veri Seti

Ham model, negatif yönde bir ilişki göstermesine rağmen, bu ilişki istatistiksel olarak anlamlı değildir. Eğilim skoru yöntemleri arasında en küçük standart hata ve dar güven aralıkları ağırlıklandırma yöntemine ait bulunmuştur. Tabakalandırma ve regresyon düzeltmesi modelleri ise benzer sonuçlar vermiş, ancak her iki model de istatistiksel olarak anlamlı bir ilişki göstermemiştir. Bu bulgular, kullanılan modellerin hiçbirinin istatistiksel olarak anlamlı bir etkiye sahip olmadığını ortaya koymaktadır.

Çizelge 4.19 Örneklem sayısının 500 olduğu senaryoda modellerin karşılaştırmaları

Model	*Beta*	*SH*	*P*	*Odds oranı*	*%95 güven aralığı*	
Ham	-0.075	0.273	0.782	0.927	0.550	1.611
Eşleştirme	0.120	0.346	0.729	1.127	0.571	2.232
Ağırlıklandırma	0.172	**0.151**	0.255	1.187	**0.884**	**1.597**
Tabakalandırma	0.149	0.289	0.607	1.160	0.666	2.079
Regresyon düzeltmesi	0.133	0.292	0.648	1.142	0.653	2.056

SH=standart hata

Model performansı açısından, ham model ve ağırlıklandırma modeli, %84.5 duyarlılık oranı ile en yüksek duyarlılığa sahipken, özgüllük oranları %26.4 ile nispeten düşük kalmıştır. Tabakalandırma modeli, %72 özgüllük oranı ile en yüksek özgüllüğe sahip olmasına rağmen, duyarlılık oranı %43.4 ile sınırlıdır. Eşleştirme modeli, %54.3 duyarlılık ve %60 özgüllük oranları ile dengeli bir performans sergilemiştir. Regresyon düzeltmesi modeli ise %68.5 duyarlılık ve %46.4 özgüllük oranları ile orta düzeyde bir performans göstermiştir.

Eğri altındaki alan değerleri incelendiğinde, tüm modellerin 0.495 ile 0.572 arasında değişen orta düzeyde bir tahmin gücüne sahip olduğu görülmektedir. Tabakalandırma modeli, 0.572 değeri ile diğer modellere kıyasla nispeten daha yüksek bir tahmin performansı sergilemiştir.

Çizelge 4.20 Örneklem sayısı 500 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları

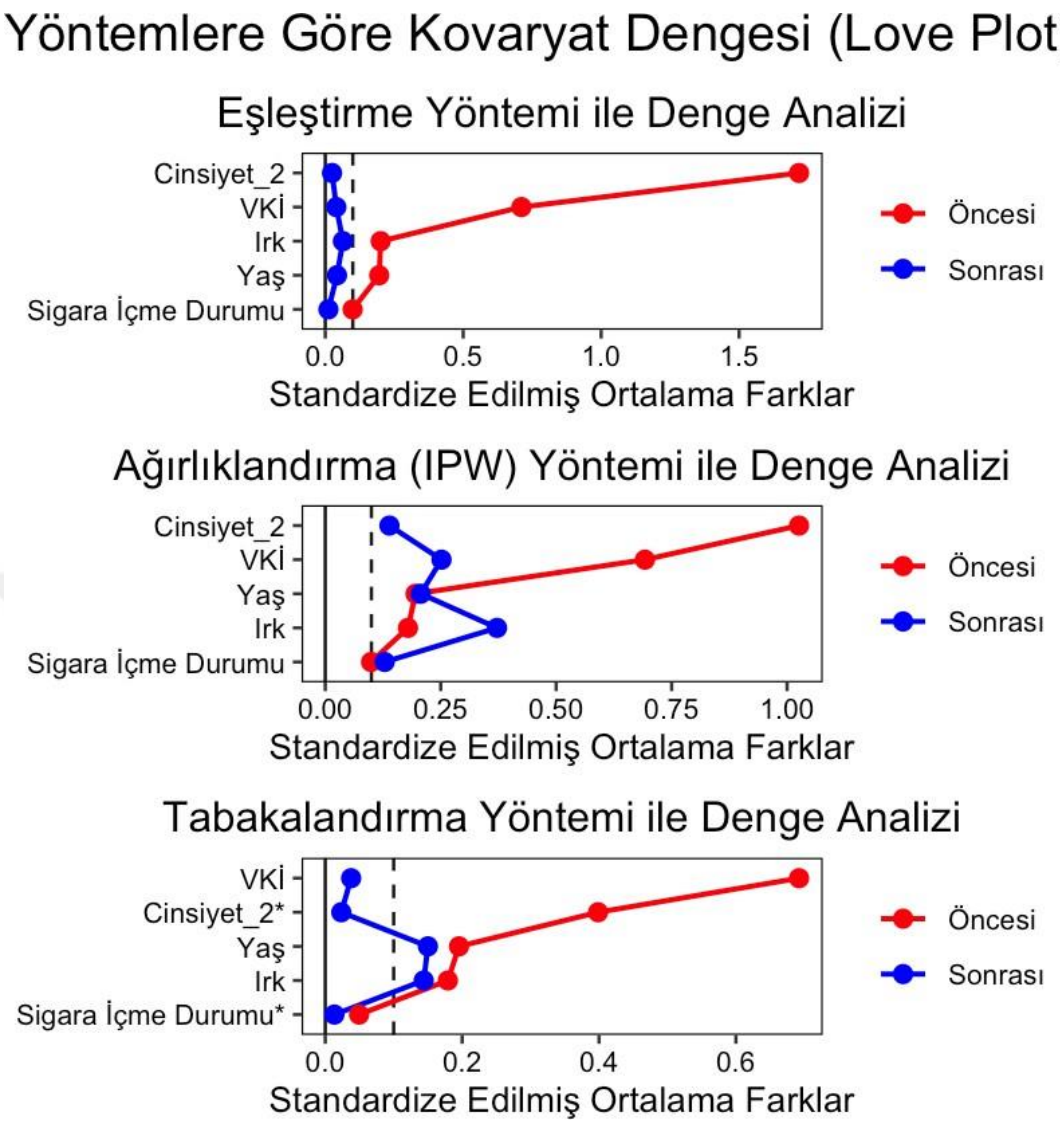
Model	*EAA*	*%95 güven aralığı*
Ham	0.505	0.467 - 0.544
Eşleştirme	0.515	0.429 - 0.600
Ağırlıklandırma	0.495	0.456 - 0.533
Tabakalandırma	0.572	0.515 - 0.629
Regresyon düzeltmesi	0.565	0.507 - 0.623

EAA=eğri altındaki alan

4.4.1.3 Sonuç Değişkenin 1/3 Oranında Görüldüğü Örneklem Sayısı 1000 Olan Veri Seti

Yöntemlerin standardize edilmiş ortalama farkları dengelemedeki performası Şekil 4.9'da gösterilmiştir. Eşleştirme ortalamalar arası farkları en düşük seviyeye indirgeyen yöntem olduğu görülmüştür.

Son senaryonun bulguları incelendiğinde, ham modelde osteoporoz tedavisinin istatistiksel anlamlı olarak kronik osteoartrit gelişimini azalttığı görülmüştür. Bu etki ve istatistiksel anlamlılık, ağırlıklandırma modelinde de görülmemiş, ancak bu modelin standart hatasının daha düşük olduğu ve daha dar bir güven aralığına sahip olduğu saptanmıştır. Bu durum, ağırlıklandırma modelinin güvenilir tahminler ürettiğini göstermektedir. Tabakalandırma, eşleştirme ve regresyon düzeltmesi modellerinde ise osteoporoz tedavisi almanın kronik osteoartrit gelişmini istatistiksel olarak etkilemediği saptanmıştır.



Şekil 4. 9 Eşleştirme, ağırlıklandırma ve tabakalandırma yöntemlerinin denge performanslarının farklı oranlı sonuç değişkeni ve örneklem sayısı 1000 olan veri setinde karşılaştırılması

Çizelge 4.21 Örneklem sayısının 1000 olduğu senaryoda modellerin karşılaştırmaları

Model	*Beta*	*SH*	*P değeri*	*Odds oranı*	*%95 güven aralığı*	
Ham	-0.459	0.181	**0.011**	0.632	0.445	0.906
Eşleştirme	-0.053	0.229	0.819	0.949	0.605	1.487
Ağırlıklandırma	-0.011	**0.101**	0.916	0.989	**0.811**	**1.206**
Tabakalandırma	-0.279	0.202	0.168	0.756	0.510	1.129
Regresyon düzeltmesi	-0.269	0.201	0.181	0.764	0.517	1.138

SH=standart hata

Eğri altındaki alan değerleri incelendiğinde, tüm modellerin tahmin gücünün sınırlı olduğu görülmüştür. Ağırlıklandırma modeli yüksek duyarlılık oranı, tabakalandırma ise en yüksek eğri altındaki alana sahip yöntem olarak saptanmıştır.

Çizelge 4.22 Örneklem sayısı 1000 olan alt veri setinde eğilim skoru yöntemlerinin ROC analizi sonuçları

Model	*EAA*	*%95 güven aralığı*
Ham	0.535	0.506 - 0.564
Eşleştirme	0.513	0.457 - 0.569
Ağırlıklandırma	0.535	0.506 - 0.564
Tabakalandırma	0.610	0.572 - 0.648
Regresyon düzeltmesi	0.544	0.500 - 0.588

EAA=eğri altındaki alan

5 Tartışma

Bu tez çalışmasında, gözlemsel çalışmalarda son yıllarda popülerleşen eğilim skoru yöntemleri farklı senaryolarda karşılaştırmalı olarak incelenmiştir. Literatürde en sık kullanılan yöntemin eşleştirme olduğu, diğer yöntemlerin ise daha az tercih edildiği tespit edilmiştir. Çalışmamızın ana bulgusu, veri setindeki değişkenlerin yapısı, örneklem sayısı, sonuç değişkeninin türü ve görülme sıklığı gibi faktörlere bağlı olarak yöntemlerin performanslarının önemli ölçüde değişmesidir. Sonuç değişkenin ikili yapıda olması durumunda ters olasılıklı ağırlıklandırma yöntemi, çoğu senaryoda düşük standart hata, dar güven aralıkları ve yüksek duyarlılık ile öne çıkarken; tabakalandırma yöntemi, ROC analizinde en yüksek eğri altında kalan alana sahip model olarak üstünlük göstermiştir. Sonuç değişkenin sürekli yapıda olması halinde ise regresyon düzeltmesi ve eşleştirme modellerinin ön plana çıktığı saptanmıştır.

Eğilim skoru eşleştirme, eğilim skorunun en yaygın kullanım yöntemlerinden biridir. Bu yöntem, gözlemsel çalışmalarda tedavi etkisini değerlendirirken karıştırıcı faktörlerin etkisini azaltmayı amaçlar. Kurduğu denge ile karşılaştırılabilir dengeli gruplar oluşturması en önemli avantajıdır (25). Eğilim skoru eşleştirmede iki önemli problem ortaya çıkabilir. Bunlardan ilki örneklem kaybıdır. Eşleştirme yaparken bazı bireyler dışarıda kalır. Eğer başlangıçtaki örneklem genişliği küçükse veya eğilim skoru dağılımlarında çok az örtüşme varsa, kaybedilen bireylerin sayısı önemli olabilir. Bu durum, çalışmanın sonuçlarını tehlikeye atarak, gerçekte var olan farkları tespit edememe riskini artırabilir (tip 2 hata). İkinci problem ise temsiliyet sorunudur. Örtüşen eğilim skoru değerlerine sahip bireyler, bulundukları ana grupları tam olarak temsil etmeyebilir (37-38). Tıp literatüründe sıkça kullanılan bir yöntem olmasına rağmen, uygulanması ve raporlanmasında eksiklikler olduğu tespit edilmiştir (39). Elze ve ark. (40) Kardiyoloji alanında kullanılan geniş örneklemli veri setlerinde yaptıkları karşılaştırma sonucunda eşleştirme modelinin güvenilir bir değişken dengesi sağladığını, uygulamasının basit ve anlaşılır olduğuna vurgu yapmışlardır. Dezavantaj olarak eşleştirme sonrası ortalama %60 oranında veri kaybının kesinlik ve genelleştirilebilirlik kaybına neden olduğu bildirdiler.

Çalışmamızda, parsiyel nefrektomi veri setinde %43.1, osteoartrit veri setinde ise %62.6 oranında örneklem kaybı meydana gelmiştir. Eşleştirme modeli, hem parsiyel

nefrektomi hem de osteoartrit veri setlerinde güçlü bir değişken dengesi sağlamıştır. Parsiyel nefrektomi veri setinde yapılan analizde, eşleştirme yöntemi ile gerçekleştirilen değerlendirme, ham analizle benzer sonuçlar vermiş ve tedavinin trifekta üzerindeki etkisi istatistiksel olarak anlamsız bulunmuştur. Osteroartrit veri setinde ham modelde osteoporoz tedavisi almanın kronik osteoartrit gelişimini azalttığını göstermiştir. Ancak eşleştirme sonrası bu etki ortadan kaybolmuştur. Bulgular eşleştirme yönteminin potansiyel karıştırıcı faktörleri dengelediğini ve osteoporoz tedavisi ile osteoartrit arasındaki ilişkinin başlangıçta düşünüldüğü kadar etkili olmayabileceğini göstermiştir.

Üretilen senaryolar, eşleştirme yönteminin grup karıştırıcıları arasında en iyi dengeyi sağladığını, ancak diğer yöntemlere kıyasla daha yüksek standart hata ile sonuçlandığını göstermektedir. Sürekli sonuç değişkenini araştıran senaryoda en yüksek R^2 değerine sahiptir. EAA açısından zayıf bir. Performans göstermiştir.

Ters olasılıklı ağırlıklandırma yöntemi eğilim skorunu kullanarak her bir bireyi gerçek tedavisini alma olasılığının tersine göre ağırlıklandırır (41). Ağırlıklandırma yöntemi, tedavi grupları arasında etkin kovaryans dengesine sahip bir sahte popülasyon oluşturarak, değişkenlerin etkisini eşitlemeye yardımcı olur. Bu sayede, tedavi etkisi üzerinde karıştırıcı değişkenlerin etkisi en aza indirilir ve daha doğru bir tedavi etkisi ortaya çıkarılabilir (42-43). Literatür incelendiğinde ağırlıklandırma yönteminin basit, tüm çalışma katılımcılarını içeren bir yöntem olması nedeniyle tutarlı sonuçlar verdiği bildirilmiştir (44-45). Ancak değişkenler arasında belirgin bir farklılık olduğunda eğilim skoru değeri 0 ve 1'e çok yaklaşır. Bu da ağırlıklandırma puanında uç değerlere yol açıp, yöntemin etkinliğini düşürebilir.

Austin ve ark. dört farklı eğilim skorunu karşılaştırdıkları çalışmada eşleştirme ve ağırlıklandırma yöntemlerinin, sonuç değişkeni tedaviden olaya kadar geçen süre olduğunda, öne çıktığını bildirmişlerdir (44,46). Diğer bir çalışmada, ağırlıklandırma yönteminin, eşleştirme ve tabakalandırma yöntemlerine kıyasla değişkenler arasındaki dengeyi sağlamada daha üstün olmadığı gösterilmiştir (40). Yine aynı çalışmada ağırlıklandırmanın önemli sayıda karıştırıcıların olduğu az sayılı örneklemlerde kesin olmayan tahminler verdiğini bildirmişlerdir.

Bizim çalışmamızda, ilk senaryoda 239 örneklem içeren veri setinde, ağırlıklandırma yöntemi diğer yöntemlerden (ham model ve eşleştirme yöntemi) farklı olarak tedavi etkisinin trifektayı negatif yönde etkilediği göstermiştir. Ayrıca, bu veri

setinde ağırlıklandırma yöntemi, daha yüksek beta katsayısı, daha düşük standart hata ve daha dar güven aralığına sahip olması nedeniyle daha güçlü bir yöntem olarak değerlendirilmiştir. Bu veri seti üzerine yapılan çalışmanın, eşleştirme yöntemi ile kurgulanmış bir yayına dayanması, eğilim skoru yöntemlerinin doğru kullanımının önemini bir kez daha vurgulamaktadır. Çünkü kullanılan yöntem, çalışmanın yorumunu önemli ölçüde değiştirebilmektedir. Nitekim, eşleştirme modelinde tedavi almanın trifekta üzerine anlamlı bir etkisi bulunmazken, ağırlıklandırma yönteminde bu etkinin negatif yönde ve istatistiksel olarak anlamlı olduğu görülmüştür. Ağırlıklandırma yöntemi, diğer senaryolarda da genel olarak dar bir güven aralığı, yüksek duyarlılık ve düşük özgüllük sergilemiştir. Örneklem sayısı azaldıkça özgüllüğün düştüğü, duyarlılığın ise arttığı gözlemlenmiştir. Örneklem sayısı arttıkça ise ağırlıklandırma yönteminin en düşük standart hataya sahip yöntem olduğu belirlenmiştir.

Tabakalandırma yöntemi, eğilim puanının belirli sayıda tabakaya bölünerek gözlemsel çalışmalarda kullanıldığı bir diğer yöntemdir. Veri setindeki karıştırıcı değişkenlere ve örneklem büyüklüğüne bağlı olarak tabakalar oluşturulabilir. Literatürde en yaygın kullanılan yaklaşım, veri setini beş tabakaya ayırmaktır (10). Eğilim skoruna göre beş tabakaya ayrıldığında yanlılığın %90 oranında azaldığı bildirilmiştir (11). Sungur G.'nin çalışması, ağırlıklandırma ve eşleştirme yöntemlerinin tabakalandırmaya kıyasla yanlılığı daha iyi azalttığını göstermiştir (47). Buna ek olarak, çok sayıda sonuç değişkeni içeren çalışmalarda daha fazla tabaka kullanılması, karıştırıcı değişkenlere bağlı yanlılığı daha da azaltabilir. Özellikle değişken dengesizliğinin belirgin olduğu durumlarda bu strateji önemli olabilir (27). Bununla birlikte, tabaka sayısının ve büyüklük dağılımının belirlenmesi konusunda daha fazla araştırmaya ihtiyaç duyulmaktadır. Örneğin, eşit büyüklükte tabakalar mı tercih edilmelidir? Eğilim puanı dağılımının orta kısmında daha fazla bireyin yer alması, dağılımın uçlarını daha ayrıntılı incelemeye olanak sağlayabilir mi? gibi soruların yanıtlanması gerekmektedir.

Bu çalışmada, tabakalandırma yönteminin performansı senaryolara göre değişkenlik göstermiştir. İlk senaryoda, parsiyel nefrektomi veri setinde, ağırlıklandırma yöntemiyle birlikte diğer yöntemlere kıyasla daha yüksek beta katsayısına sahip olduğu ve ağırlıklandırma yönteminden daha düşük bir standart hataya sahip olduğu belirlenmiştir. Duyarlılık ve özgüllük açısından orta düzeyde bir performans sergilerken, en yüksek EAA değerine sahip yöntem olarak saptanmıştır. Genel olarak, senaryoların

çoğunda diğer yöntemlere kıyasla en yüksek EAA değerini sağlamıştır. Ancak, örneklem sayısı azaldıkça güven aralığı en fazla genişleyen yöntem olduğu gözlemlenmiştir.

Regresyon düzeltmesi, seçim yanlılığını ve potansiyel karıştırıcıları kontrol etmek için kullanılan geleneksel bir yöntemdir. Eğilim skoru yöntemlerinin ortaya çıkmasının ardından kullanımı azalmış olsada, literatürde bu konuda çelişkili bulgular mevcuttur (48). Birçok çalışma, eğilim skoru eşleştirmenin, tedavi edilen ve edilmeyen bireyler arasındaki başlangıçtaki sistematik farklılıkları, eğilim skoruna göre tabakalandırma veya eğilim skoru kullanarak yapılan regresyon düzeltmesine kıyasla daha iyi ortadan kaldırdığını göstermektedir (20,25,49). Bununla birlikte, sonuç değişkeni sürekli olduğunda, regresyon düzeltmesinin eğilim skoru yöntemleri ile benzer sonuçlar verdiği bildirilmiştir (10). Sürekli sonuç değişkeni varlığında eğilim skoru yöntemleri çalışmalarda detaylandırılmıştır (50-51).

Cattone ve arkadaşları, örneklem büyüklüğü 200'ün üzerinde olan veri setlerinde eğilim skoru yöntemlerinin tercih edilmesi gerektiğini vurgulamışlardır. Ayrıca, 80'den az örnekleme sahip veri setlerinde, ağırlıklandırma yönteminin eşleştirmeye kıyasla daha iyi sonuçlar verdiğini belirtmişlerdir (50). Bunun yanı sıra, regresyon düzeltmesinin bu yöntemlere ek olarak kullanılmasıyla tahminlerin hassasiyetinin arttığı ifade edilmiştir. Başka bir çalışmada ise regresyon düzeltmesi ve eşleştirmenin tüm örneklerde iyi performans gösterdiği, ancak eşleştirmenin bazı durumlarda daha geniş güven aralıklarına sahip, dolayısıyla daha az kesin tahminler ürettiği bildirilmiştir (40). Eğilim skoru yöntemlerinin her zaman geleneksel regresyon ayarlamasına üstün olmadığı ve bu nedenle çalışmanın ihtiyaçlarına en uygun yöntemin dikkatli bir şekilde seçilmesi gerektiği vurgulanmıştır (40).

Bu tez çalışması literatürle uyumlu olup, senaryolar incelendiğinde eğilim skoru yöntemlerinin her koşulda regresyon düzeltmesine mutlak bir üstünlük sağlamadığı gözlemlenmiştir. Özellikle ikinci senaryoda, yani sonuç değişkeninin sürekli olduğu durumda, regresyon düzeltmesi yöntemi dar bir güven aralığı ile en yüksek beta katsayısına sahip yöntem olarak belirlenmiştir. Ayrıca, örneklem büyüklüğü arttıkça, regresyon düzeltmesinin standart hatasının ağırlıklandırma yöntemiyle birlikte en düşük seviyede olduğu saptanmıştır.

Çalışma birtakım limitasyonlar içermektedir. Bunlar;

- Yöntemlerin karşılaştırılmasında veya uygulanmasında yöntemlerin tek bir alt tip kullanılmıştır. Örneğin eşleştirme için tanımlanan birden fazla yöntem var iken çalışmada en yakın komşuluk (kaliper:0.20) yöntemi tercih edilmiştir. Diğer eşleştirme yöntemlerinin performansı karşılaştırmalara dahil edilmemiştir.
- Aynı durum tabakalandırma ve ters olasılıklı ağırlıklandırma yöntemleri için de mevcuttur. Tabakalandırmada 5 tabaka yöntemi kullanılmış ve tabakaların ortak etkisi hesaplanarak karşılaştırma yapılmıştır. Ağırlıklandırmada ise olası uç değerler sansürlenmemiştir.
- Mevcut veri setleri kullanılarak senaryolar ve değişken sayıları olabildiğince geniş tutulmuş olsada, tüm olasılıkları değerlendirebilmek için daha çeşitli senaryolara ihtiyaç duyulabilir.
- Son olarak modellerin performanslarının değerlendirilmesi lojistik regresyon ve ROC analizi sonuçları ile yapılmıştır. ROC analizinde sonuçlar bireylerin tedaviye atanma olasılıklarını ikili yapıda değerlendirdikleri için EAA değeri sınırlı bir yoruma sahiptir.
- Bu çıktılar daha farklı simülasyonlarla ve güç analizleriyle genişletilebilir.

6 Sonuç ve Öneriler

Çalışmanın en önemli bulgularından biri, randomize olmayan gözlemsel çalışmalarda veya retrospektif olarak toplanan verilerde eğilim skoru yöntemlerinin farklı performanslar sergileyebileceğidir. Araştırmacılar, çalışmalarında genellikle eşleştirme yöntemini tercih etmelerine rağmen, ağırlıklandırma, tabakalandırma ve regresyon düzeltmesi yöntemleri veri setinin özelliklerine bağlı olarak üstün performans gösterebilmektedir. Yöntemlerin avantaj ve dezavantajları değerlendirildiğinde:

- **Eşleştirme yöntemi**, karıştırıcı değişkenlerin neden olduğu dengesizliği gidermede en etkili yöntem olarak belirlenmiştir. Uygulaması ve yorumlanması görece basit olmakla birlikte, örneklem sayısı azaldıkça standart hatasının arttığı ve duyarlılığının düştüğü gözlenmiştir. Özellikle düşük örneklem sayılı veri setlerinde, dengeyi sağlamak adına örneklem kaybına neden olması, yöntemin kullanılabilirliğini kısıtlamaktadır. Sürekli sonuç değişkeni varlığında en yüksek R^2 değerine sahip yöntem olarak bulunmuştur.
- **Ters olasılıklı ağırlıklandırma yöntemi**, eşleştirme yöntemine benzer bir değişken dengesi sunarken, tüm bireyleri analize dahil ederek yalancı bir popülasyon oluşturması önemli bir avantaj olarak görülmüştür. Düşük örneklem sayılı veri setlerinde, özellikle tabakalandırma yöntemi ile birlikte öne çıktığı saptanmıştır. Yüksek duyarlılığa ve düşük özgüllüğe sahip olduğu belirlenmiştir. Ancak, eğilim skoru 0 veya 1'e yaklaştıkça uç değerlerin ortaya çıkması modelde yanlılığa neden olabilmektedir.
- **Tabakalandırma yöntemi**, veri setindeki tüm popülasyonu kullanabilmesi ve tabakalar arasındaki etkileşimleri ayrı ayrı sunabilmesi nedeniyle avantajlıdır. Tüm senaryolarda en yüksek EAA değerine sahip yöntem olarak bulunmuştur. Bununla birlikte, bazı senaryolarda yüksek standart hataya ve geniş güven aralıklarına sahip olması, yöntemin dezavantajları arasında yer almaktadır.
- **Regresyon düzeltmesi**, eğilim skoru yöntemlerinin yaygınlaşmasıyla kullanımı azalmış olsa da, özellikle sürekli sonuç değişkenlerinin bulunduğu senaryoda diğer yöntemlere kıyasla (eşleştirme yöntemi ile birlikte) daha iyi performans

sergilediği görülmüştür. Buna ek olarak, düşük örneklem sayılı veri setlerinde yüksek duyarlılığa sahip olduğu belirlenmiştir.

Bu çalışmada elde edilen sonuçlar, araştırmacıların eğilim skoru yöntemlerini seçerken veri setinin özelliklerini ve çalışmanın gereksinimlerini dikkatlice değerlendirmeleri gerektiğini göstermektedir. Her yöntemin farklı senaryolarda değişen güçlü ve zayıf yönleri olduğundan, uygun yöntem seçimi sonuçların geçerliliğini artırarak çalışmanın bilimsel değerini yükseltebilir. Yanlış yöntem seçimi ise yanlılıklara neden olarak çalışmanın güvenilirliğini ve bilimsel değerini olumsuz yönde etkileyebilir.

7 Kaynaklar

1. **LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F**. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England J Med*. **1997**; 337 :536- 42.

2. **Ebrahim Valojerdi A, Janani L.** A brief guide to propensity score analysis. *Med J Islam Repub Iran.* **2018**; 7: 32-122.

3. **Pirracchio R, Resche-Rigon M, Chevret S.** Evaluation of the propensity score methods for estimating marginal probability ratios in case of small sample size. *BMC Med Res Methodol*. **2012**; 30:12-70.

4. **D'Agostino RB Jr.** Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. **1998** ; 15:2265-81.

5. **Harrell F Jr., Lee KL, Califf RM, et al**. Regression modelling strategies for improved prognostic prediction. *Stat Med*. **1984**; 3:143–52.

6. **Vittinghoff E, McCulloch CE**. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemio*. **2007**; 165:710–8.

7. **Kahan BC, Jairath V, Doré CJ, Morris TP**. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*. **2014**; 23:15-139.

8. **Judea Pearl.** Causal inference in statistics: An overview. *Statist. Surv*. **2009**; 3:96-146.

9. **Rosenbaum PR, Rubin DB**. The central role of the propensity score in observational studies for causal effects. *Biometrika*. **1983**; 70:41–55.

10. **Austin PC.** An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. **2011**; 46:399-424.

11. **Rosenbaum P, Rubin D.** Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. **1984**; 79:516–524.

12. **Williamson, E.J. and Forbes,** Propensity scores**.** ***Respirology*****. 2014**; 19: 625-635.

13. **Austin PC.** A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med.* **2008**; 27:2037–2049.

14. **Lee, B. K., Lessler, J., Stuart, E.** A. Improving propensity score weighting using machine learning. *Statistics in Medicine*. **2010**; 29, 337–346.

15. **McCaffrey, D. F., Ridgeway, G., & Morral, A. R**. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, **2004**; 9:403–425.

16. **Setoguchi, S., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., & Cook, E. F.** Evaluating uses of data mining techniques in propensity score estimation: A simulation study. *Pharmacoepidemiology and Drug Safety*. **2008**; 17: 546–555.

17. **Kuss O, Blettner M, Börgermann J**. Propensity Score: an Alternative Method of Analyzing Treatment Effects. *Dtsch Arztebl Int*. **2016**;113: 597-603.

18. **Baek, S., Park, S. H., Won, E., Park, Y. R. and Kim, H. J.** Propensity score matching: A conceptual review for radiology researchers. *Korean Journal of Radiology*. **2015**; 16: 286-296.

19. **Austin, PC.** Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in Statistics Simulation and Computation*. **2009**; 38: 1228–1234.

20. **Austin PC**. Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. *Int J Biostat*. **2009**; 5:13.

21. **Amoah J, Stuart EA, Cosgrove SE, Harris AD, Han JH, Lautenbach E, Tamma PD**. Comparing Propensity Score Methods Versus Traditional Regression Analysis for the Evaluation of Observational Data: A Case Study Evaluating the Treatment of Gram-Negative Bloodstream Infections. *Clin Infect Dis.* **2023**;71(9):497-505.

22. **Ho, D., Imai, K., King, G., and Stuart, E.** Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference, *Political Analysis*. **2007**:15, 199–236.

23. **A. Olmos and P. Govindasamy**. "Propensity scores: a practical introduction using R," *Journal of MultiDisciplinary Evaluation*, pp. 68-88, **2015**.

24. **Beşpınar, E.** (**2018**). Eğilim skoru kullanılarak eşleştirme yöntemlerinin performanslarının karşılaştırılması. Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

25. **Austin, PC.** Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for improvement. *Journal of Thoracic and Cardiovascular Surgery*. **2007**: 134, 1128–1135.

26. **Cochran, W. G.** The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society*. **1965**: 7(3), 134–155.

27. **Cochran, W. G., and CHambers, S. P.** The planning of observational studies of human populations. *Journal of the Royal Statistical Society*. **1965**: 128(2), 234-266.

28. **Rosenbaum, P. R**. Optimal matching for observational studies. *Journal of the American Statistical Association*. **1989**: 84(408), 1024–1032.

29. **Rosenbaum, P. R**. A characterization of optimal designs for observational studies. *Journal of the Royal Statistical Society.* **1991**:53(3), 597–610.

30. **Dehejia, R. H. and Wahba, S.** Propensity Score-Matching Methods For Nonexperimental Causal Studies. *The Review of Economics and Statistics*. **2002**; 84(1), 151-161.

31. **Cochran, W. G.** The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. **1968**; 24, 295–313.

32. **Huppler Hullsiek, Louis, T. A.** Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics*. **2002**; 3, 179–193.

33. **Imbens, G. W.** Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*. **2004**; 86, 4–29.

34. **Lunceford, J. K., & Davidian, M.** Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*.**2004**; 23, 2937– 2960.

35. **Williamson, E. J. and Forbes, A.** Introduction to propensity scores. *Respirology*. **2014**; 19: 625-635.

36. **Anıl H, Yıldız A, Güzel A, Akdemir S, Karamık K, Arslan M.** Comparison of Posterior and Antero-Lateral Renal Tumors in Retroperitoneal Laparoscopic Partial Nephrectomy: A Propensity Score Matching Analysis. *J Kidney Cancer VHL*. **2023**; 10;10(3):9-16.

37. **Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, Fremes SE.** A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. *Can J Cardiol*. **2016**;32(2):259-65.

38. **Streiner DL, Norman GR.** The pros and cons of propensity scores. *Chest*. **2012**;142(6):1380-1382.

39. **Rubin DB.** On principles for modeling propensity scores in medical research. *Pharmacoepidemiol Drug Saf*. **2004**;13(12):855-7.

40. **Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, Nichols M, Stone GW, Pocock SJ.** Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *J Am Coll Cardiol*. **2017**; 69(3):345-357.

41. **Robins J.** A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Math Model*. **1986**; 7: 1393–1512.

42. **Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C, Jager KJ.** An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J*. **2021**; 26;15(1):14-20.

43. **Heinze G, Jüni P.** An overview of the objectives of and the approaches to propensity score analyses. *Eur Heart J*; **2011**;32:1704–8.

44. **Austin PC**. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Making*. **2009**; 29:661–77.

45. **Lunceford JK, Davidian M.** Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med*.**2004**;23:2937–60.

46. **Austin PC**, "The Performance of Different Propensity Score Methods for Estimating Marginal Hazard Ratios," *Statistics in Medicine*. **2013**: 2837–2849.

47. **Gurel S.** "The Performance of Propensity Score Methods to Estimate the Average Treatment Effect in Observational Studies with Selection Bias: A Monte Carlo Simulation Study." PhD diss., University of Florida, **2012**.

48. **Glynn RJ, Schneeweiss S, Stürmer T.** Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol*. **2006**; 98:253–9.

49. **Austin P.C., Mamdani M.M.** A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine*. **2006**;25:2084–2106. doi: 10.1002/sim.2328.

50. **Imai K, Ratkovic M**. Covariate balancing propensity score. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2014**; 76(1):243-263

51. **Fong C, Hazlett C, Imai K**. Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *The Annals of Applied Statistics*, **2018**; 12(1):156-177.

52. **Cottone F, Anota A, Bonnetain F, Collins GS, Efficace F**. Propensity score methods and regression adjustment for analysis of nonrandomized studies with health-related quality of life outcomes. *Pharmacoepidemiol Drug Saf*. **2019**;28(5):690-699.

EKLER

Ek-1: Parsiyel nefrektomi veri setinin R kodları

```
library(tidyverse)
library(broom)
library(MatchIt)
library(survey)
library(pROC)
library(ggplot2)
library(flextable)
library(officer)
library(cobalt)
library(ggplot2)
# Propensity Score Modeli (Lojistik Regresyon)
ps_model <- glm(Grup ~ Yaş + BMI + KitleBoyutu + Nefrometriskoru + GFRkategori + Evre,
                family = binomial(), data = TezVeri)
ps_summary <- tidy(ps_model, conf.int = TRUE) %>%
mutate(OR = exp(estimate),  # Odds oranı hesapla
       CI_Lower = exp(conf.low),
       CI_Upper = exp(conf.high)) %>%
  mutate(across(where(is.numeric), ~ round(.x, 3)))  # Tüm sayıları 3 ondalık basamağa yuvarla

# Word çıktısı oluşturma
doc <- read_docx()
doc <- body_add_par(doc, "Propensity Score Modeli Sonuçları", style = "heading 1")
doc <- body_add_flextable(doc, flextable(ps_summary))

# Word dosyasını kaydetme
print(doc, target = "PS_Model_Sonuclari.docx")
```

```
# 1. NAIVE MODEL
naive_model <- glm(Trifekta ~ Grup, family = binomial(), data = TezVeri)
TezVeri$naive_prob <- predict(naive_model, type = "response")
naive_roc <- roc(TezVeri$Trifekta, TezVeri$naive_prob)
naive_summary <- tidy(naive_model, conf.int = TRUE) %>%
  mutate(Method = "Naive")

# 2. EŞLEŞTİRME (Matching)
match_model <- matchit(Grup ~ Yaş + BMI + KitleBoyutu + Nefrometriskoru +
GFRkategori + Evre,
                       method = "nearest", caliper = 0.2, data = TezVeri)
mdata <- match.data(match_model)
match_result <- glm(Trifekta ~ Grup, family = binomial(), data = mdata)
mdata$match_prob <- predict(match_result, type = "response")
match_roc <- roc(mdata$Trifekta, mdata$match_prob)
match_summary <- tidy(match_result, conf.int = TRUE) %>%
  mutate(Method = "Matching")
labels <- c("distance" = "Mesafe",
            "BMI" = "Vücut Kitle İndeksi",
            "Evre" = "Klinik Evre",
            "GFRkategori" = "KBH Evre",
            "KitleBoyutu" = "Kitle Boyutu",
            "Nefrometriskoru"="Nefrometri skoru")

# Love Plot (Balance Assessment)
love_plot <- love.plot(match_model, threshold = 0.1, abs = TRUE, var.order =
"unadjusted",
                       colors = c("red", "blue"), sample.names = c("Öncesi", "Sonrası"),
                       var.names = labels) +
  labs(title = "Değişkenlerin Dengesi", x = "Standardize Edilmiş Ortalama Farklar") +
  theme(legend.position = "bottom", legend.title = element_blank())
```

```
# Love Plot'u göster
print(love_plot)

# 3. TABAKALAMA (Stratification)
TezVeri$ps_strata <- cut(TezVeri$ps, breaks = quantile(TezVeri$ps, probs = seq(0, 1,
0.2)), include.lowest = TRUE)
strata_model <- glm(Trifekta ~ Grup + ps_strata, family = binomial(), data = TezVeri)
TezVeri$strata_prob <- predict(strata_model, type = "response")
strata_roc <- roc(TezVeri$Trifekta, TezVeri$strata_prob)
strata_summary <- tidy(strata_model, conf.int = TRUE) %>%
mutate(Method = "Stratification")

library(ggplot2)
library(dplyr)

# Boxplot oluştur ve göster
p <- ggplot(TezVeri, aes(x = ps_strata, y = ps, fill = as.factor(Grup))) +
  geom_boxplot() +
  scale_fill_manual(values = c("#E74C3C", "#1ABC9C")) +  # Kırmızı ve turkuaz tonları
  labs(title = "Eğilim Skorunun Tabakalara Göre Dağılımı",
       x = "Tabakalar",
       y = "Eğilim Skoru",
       fill = "Grup") +
  theme_minimal() +
  theme(legend.position = "right",
        text = element_text(size = 14))

# Figürü göster
print(p)
```

```
# 4. TERS OLASILIK AĞIRLIKLANDIRMA (IPTW)
TezVeri$iptw_w <- ifelse(TezVeri$Grup == 1, 1 / TezVeri$ps, 1 / (1 - TezVeri$ps))
ipw_design <- svydesign(ids = ~1, weights = ~iptw_w, data = TezVeri)
ipw_result <- svyglm(Trifekta ~ Grup, family = quasibinomial(), design = ipw_design)
TezVeri$ipw_prob <- predict(ipw_result, type = "response")
ipw_roc <- roc(TezVeri$Trifekta, TezVeri$ipw_prob)
ipw_summary <- tidy(ipw_result, conf.int = TRUE) %>%
mutate(Method = "IPTW")
histogram_plot <- ggplot(TezVeri, aes(x = iptw_w, fill = as.factor(Grup))) +
geom_histogram(binwidth = 0.1, alpha = 0.6, color = "black", position = "identity") +
scale_fill_manual(values = c("red", "blue"), labels = c("Grup Posterior", "Grup Anterior")) +
  theme_minimal() +
  labs(title = " Ters Olasılıklı Ağırlıklandırma Puanı Dağılımı", x = "Ağırlıklandırma Puanları", y = "Frekans", fill = "Grup")



labels <- c("distance" = "Mesafe",
            "BMI" = "Vücut Kitle İndeksi",
            "Evre" = "Klinik Evre",
            "GFRkategori" = "KBH Evre",
            "KitleBoyutu" = "Kitle Boyutu",
            "Nefrometriskoru"="Nefrometri skoru")


# Love Plot (Balance Assessment for IPTW)
love_plot_iptw <- love.plot(Grup ~ Yaş + BMI + KitleBoyutu + Nefrometriskoru + GFRkategori + Evre,
                            data = TezVeri, weights = TezVeri$iptw_w, method = "weighting",
                            abs = TRUE, var.order = "unadjusted",
                            colors = c("red", "blue"), sample.names = c("Öncesi", "Sonrası"),
                            s.d.denom = "pooled", var.names = labels) +
```

```
  labs(title = "Ağırlıklandırma Modeli İçin Değişken Dengesi", x = "Standardize Edilmiş
Ortalama Farklar") +
  theme(legend.position = "bottom", legend.title = element_blank())
# Histogram ve Love Plot'u göster
print(histogram_plot)
print(love_plot_iptw)

# 5. PROPENSITY SCORE REGRESYONU (Adjustment)
adj_model <- glm(Trifekta ~ Grup + ps, family = binomial(), data = TezVeri)
TezVeri$adj_prob <- predict(adj_model, type = "response")
adj_roc <- roc(TezVeri$Trifekta, TezVeri$adj_prob)
adj_summary <- tidy(adj_model, conf.int = TRUE) %>%
mutate(Method = "Regression Adjustment")

# **Tüm sonuçları tek bir tabloda birleştirelim**
combined_results <- bind_rows(naive_summary, match_summary, strata_summary,
ipw_summary, adj_summary) %>%
  mutate(OR = exp(estimate),  # Odds oranı hesapla
         CI_Lower = exp(conf.low),  # %95 güven aralığı alt
         CI_Upper = exp(conf.high)) %>%  # %95 güven aralığı üst
  select(Method, term, estimate, std.error, p.value, OR, CI_Lower, CI_Upper) %>%
  rename(Beta = estimate, SE = std.error, P_Value = p.value)

# **ROC Eğrisi ve AUC Değerleri**
roc_results <- data.frame(
  Method = c("Naive", "Matching", "Stratification", "IPTW", "Regression Adjustment"),
  AUC = c(auc(naive_roc), auc(match_roc), auc(strata_roc), auc(ipw_roc), auc(adj_roc)),
  Sensitivity = c(coords(naive_roc, "best", ret = "sensitivity")$sensitivity,
                  coords(match_roc, "best", ret = "sensitivity")$sensitivity,
                  coords(strata_roc, "best", ret = "sensitivity")$sensitivity,
                  coords(ipw_roc, "best", ret = "sensitivity")$sensitivity,
                  coords(adj_roc, "best", ret = "sensitivity")$sensitivity),
```

```
  Specificity = c(coords(naive_roc, "best", ret = "specificity")$specificity,
                  coords(match_roc, "best", ret = "specificity")$specificity,
                  coords(strata_roc, "best", ret = "specificity")$specificity,
                  coords(ipw_roc, "best", ret = "specificity")$specificity,
                  coords(adj_roc, "best", ret = "specificity")$specificity)
)

# **Sonuçları yazdır**
print(combined_results)
write.csv(combined_results, "PSM_results.csv", row.names = FALSE)
write.csv(roc_results, "ROC_results.csv", row.names = FALSE)

# **Flextable ile tabloyu oluştur**
ft <- flextable(combined_results) %>%
  autofit() %>%
  theme_vanilla()

doc <- read_docx() %>%
  body_add_flextable(ft) %>%
  body_add_par("Tablo: Propensity Score Analiz Sonuçları", style = "heading 1")

print(doc, target = "PSM_results.docx")
```

Ek-2 : Osteoartrit veri seti R kodları

```
install.packages("MatchThem")  # Paket yüklü değilse yükle
library(MatchThem)             # Paketi aç
data("osteoarthritis")
str(osteoarthritis)  # Veri yapısını incele
summary(osteoarthritis)
osteoarthritis <- na.omit(osteoarthritis)
```

```
# 1. HAM MODEL

naive_model <- glm(KOA ~ OSP, family = binomial(), data = osteoarthritis)
osteoarthritis$naive_prob <- predict(naive_model, type = "response")
naive_roc <- roc(osteoarthritis$KOA, osteoarthritis$naive_prob)
naive_summary <- tidy(naive_model, conf.int = TRUE) %>%
mutate(Method = "Naive"


# 2. EŞLEŞTİRME (Matching)

match_model <- matchit(OSP ~ AGE + BMI + SEX + RAC + SMK,
                       method = "nearest", caliper = 0.2, data = osteoarthritis)

mdata <- match.data(match_model)
match_result <- glm(KOA ~ OSP, family = binomial(), data = mdata)
mdata$match_prob <- predict(match_result, type = "response")
match_roc <- roc(mdata$KOA, mdata$match_prob)
match_summary <- tidy(match_result, conf.int = TRUE) %>%
  mutate(Method = "Matching")

# 3. TABAKALAMA (Stratification)

osteoarthritis$ps_strata <- cut(osteoarthritis$ps, breaks = quantile(osteoarthritis$ps,
probs = seq(0, 1, 0.2)), include.lowest = TRUE)
strata_model <- glm(KOA ~ OSP + ps_strata, family = binomial(), data = osteoarthritis)
osteoarthritis$strata_prob <- predict(strata_model, type = "response")
strata_roc <- roc(osteoarthritis$KOA, osteoarthritis$strata_prob)
strata_summary <- tidy(strata_model, conf.int = TRUE) %>%
mutate(Method = "Stratification")
```

```
# 4. TERS OLASILIK AĞIRLIKLANDIRMA (IPTW)

osteoarthritis $weight <- ifelse(osteoarthritis $OSP == 1, # osteoarthritis
                    1 / osteoarthritis $ps, # osteoarthritis
                    1 / (1 - osteoarthritis $ps)) # osteoarthritis
ipw_model <- glm(KOA ~ OSP,
                 family = quasibinomial
                 data = osteoarthritis, # osteoarthritis
                 weights = weight)


# 5. REGRESYON DÜZELTMESİ (Covariate Adjustment)

adj_model <- glm(KOA ~ OSP + ps, family = binomial(), data = osteoarthritis)
osteoarthritis$adj_prob <- predict(adj_model, type = "response")
adj_roc <- roc(osteoarthritis$KOA, osteoarthritis$adj_prob)
adj_summary <- tidy(adj_model, conf.int = TRUE) %>%
mutate(Method = "Regression Adjustment")
```

ÖZGEÇMİŞ

