مقاله مروری
مجله دانشگاه علوم پزشکی رفسنجان
دوره ۲۱، خرداد ۱۴۰۱، ۳۶۶-۳۴۳

چگونه می‌توان در مطالعات مشاهده‌ای روابط علیتی را سنجید؟ همسان‌سازی با نمره گرایش: یک مقاله آموزشی

مینا طهماسبی فرد[۱]، فرید نجفی[۲]، شهاب رضائیان[۳]، مهدی مرادی نظر[۴]

دریافت مقاله: ۱۴۰۱/۰۲/۱۱ ارسال مقاله به نویسنده جهت اصلاح: ۱۴۰۱/۰۲/۳۱ دریافت اصلاحیه از نویسنده: ۱۴۰۱/۰۳/۲۳ پذیرش مقاله: ۱۴۰۱/۰۳/۲۵

چکیده

زمینه و هدف: در بیش‌تر مطالعات مشاهده‌ای پژوهش‌گران سعی در پیدا کردن روابط علیتی بین مشاهدات را دارند، این ارتباط سنجی به دلیل عدم کنترل متغیرهای مخدوش‌گر با خطا همراه است. یکی از روش‌هایی که برای کنترل متغیرهای مخدوش‌گر کاربردی است، همسان‌سازی با نمره گرایش است. لذا هدف از این مطالعه، توضیح مراحل انجام همسان‌سازی با نمره گرایش است.

مواد و روش‌ها: همسان‌سازی نمره گرایش دارای ۵ مرحله است. مرحله اول برآورد نمره گرایش است که شامل انتخاب مدل و متغیر مناسب می‌باشد. مرحله دوم، بر اساس نمرات گرایش برآورد شده باید تصمیم گرفت از میان روش‌های تطبیق کدام روش را انتخاب کرد. مرحله سوم هم‌پوشانی و پشتیبانی مشترک است، در این مرحله مشاهداتی که خارج از محدوده نمرات تطبیق قرار دارند، حذف می‌شوند. پس از آن کیفیت مطابقت، ارزیابی می‌شود و در نهایت باید حساسیت اثرات برآورد شده را تخمین زد.

یافته‌ها: در تطبیق نمره گرایش، متغیرهای اساسی مشاهده شده بین گروه‌های مواجهه و عدم مواجهه متعادل می‌شوند. با این حال، اگر مدل آماری مورد استفاده برای محاسبه نمره گرایش به درستی انتخاب نشده باشد، عدم تعادل بین ویژگی‌های پایه دو گروه هم‌چنان می‌تواند وجود داشته باشد.

نتیجه‌گیری: همسان‌سازی بر اساس نمره گرایش در مواردی که مخدوش کننده‌های مطالعه زیاد باشند، مفید است. این روش در مطالعات مشاهده‌ای جایگزین تصادفی‌سازی در مطالعات تجربی است که با کاهش تورش انتخاب و کنترل متغیرهای مخدوش‌گر برآوردها را دقیق‌تر می‌کند.

واژه‌های کلیدی: همسان‌سازی نمره گرایش، متغیرهای مخدوش‌گر، مطالعات مشاهده‌ای

۱- دانشجوی کارشناسی ارشد اپیدمیولوژی ، گروه اپیدمیولوژی ، دانشکده بهداشت ، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه ، ایران
۲- استاد، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران
۳- دانشیار، گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران
۴- (نویسنده مسئول) استادیار ،گروه اپیدمیولوژی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران
مرکز تحقیقات بیماری های رفتاری، دانشگاه علوم پزشکی کرمانشاه ایران
تلفن: ۰۸۳-۳۸۲۶۵۲۵۵، دورنگار: ۰۸۳-۳۸۲۶۵۲۵۵، پست الکترونیکی: M.moradinazar@gmail.com

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

مقدمه

بیش‌تر مطالعات انجام شده در سیستم خدمات بهداشتی و درمانی از نوع مطالعات مشاهده‌ای است که در آن پژوهش گران سعی در پیدا کردن روابط علیتی بین مشاهدات را دارند که این ارتباط سنجی معمولاً به دلیل عدم کنترل متغیرهای مخدوش‌گر با خطا همراه است. این نوع مطالعات بیش‌تر در حوزه‌هایی مانند روان‌شناسی و پزشکی مورد استفاده قرار می‌گیرند، چرا که در این حوزه‌ها به خاطر مسائل اخلاقی امکان تصادفی‌سازی وجود ندارد. اعتبار نتایج مطالعات مشاهده‌ای کم‌تر از نتایج مطالعات کنترل شده هستند، چرا که بسیاری از آن‌ها در برابر تورش انتخاب (Selection bias) آسیب پذیر می‌باشند. پژوهش‌گران اغلب تلاش می‌کنند این نقیصه را با به کارگیری روش‌های آماری پیچیده مانند همسان‌سازی نمره گرایش جبران کنند [۱].

همسان‌سازی یا تطبیق بر اساس نمره گرایش (Propensity score matching) توسط Rosenbaum و Rubin به عنوان یک روش جدید کنترل اثر متغیرهای مخدوش‌گر در سال‌های اخیر مورد توجه بسیاری از محققین قرار گرفته است [۲]. تحلیل نمره گرایش بیش‌تر زمانی مناسب است که به هر نحوی تصادفی‌سازی قابل انجام نباشد. در صورت اجرای مناسب تصادفی‌سازی در مطالعات مداخله‌ای تصادفی شده، تخصیص تصادفی تضمین می‌کند که به‌طور متوسط، هیچ تفاوت سیستماتیکی در متغیرهای مخدوش‌گر اعم از مشاهده شده و مشاهده نشده بین گروه‌های مورد مطالعه وجود ندارد [۳].

در مطالعات مشاهده‌ای یا نیمه تجربی، روش تصادفی‌سازی برای تخصیص افراد در گروه‌های مواجهه یافته و عدم مواجهه نمی‌تواند اعمال شود. بنابراین، انتساب افراد به بازوهای مطالعات مشاهده‌ای تصادفی نیست و در نتیجه، برآورد تأثیر مواجهه ممکن است همراه با وجود عوامل مخدوش کننده باشد. بنابراین، از تطبیق نمره گرایش به عنوان راهی برای کاهش سوءگیری و اصلاح "برآورد" اثرات مواجهه می‌توان استفاده کرد [۴].

در مطالعه حاضر، استفاده از همسان‌سازی نمره گرایش به منظور کاهش و کنترل اثر متغیرهای مخدوش‌گر شناخته شده در یک مثال تحت عنوان بررسی تأثیر آنزیم کبدی گاما گلوتامیل ترانسفراز (Gamma-glutamyl transferase) بر روی پرفشاری خون با استفاده از داده‌های کوهورت بزرگ سالان بیماری‌های غیر واگیر روانسر (Ravansar non-communicable chronic disease) انجام شده است. مطالعه کوهورت آینده‌نگر روانسر از مرداد ۱۳۹۳ آغاز شده و برنامه‌ریزی شده است که افراد در فواصل منظم به‌طور سالیانه و به مدت ۱۵ سال مورد پی‌گیری مجدد و ارزیابی قرار گیرند [۵].

مطالعه کوهورت روانسر یکی از ۲۱ مرکز کوهورت پرشین (PERSIAN) می‌باشد که بر روی قوم کرد تحقیق می‌کند. هدف از انجام این کوهورت شناسایی و پیش‌بینی بروز

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

بیماری‌های مزمن غیر واگیر و مرگ‌ومیر مرتبط با آن و پیش گیری از عوامل خطر بیماری‌های مزمن از جمله بیماری‌های قلبی-عروقی، دیابت و فشارخون در شهرستان روانسر است. شهرستان روانسر یکی از شهرستان‌های استان کرمانشاه می‌باشد. ۱۰۰۰۰ نفر از ساکنین شهری و روستایی در سنین ۶۵-۳۵ ساله در زمان اولین دوره جمع‌آوری اطلاعات وارد مطالعه شدند. اطلاعات کامل در خصوص مطالعه کوهورت پرشین که با همکاری کمیته تحقیقات گوارشی و وزارت درمان، آموزش پزشکی صورت گرفته در آدرس اینترنتی: http://persiancohort.com وجود دارد. برای اطلاعات بیش‌تر می‌توان به پروتکل کوهورت روانسر مراجعه کرد [۶].

فشارخون بالا یکی از مهم‌ترین عوامل خطر بیماری‌های قلبی-عروقی در سراسر جهان است. هرچند پرفشاری خون علامت نداشته و علل مختلفی در ایجاد آن نقش دارند اما می‌توان با دارو آن را کنترل نمود. بر اساس گزارش سازمان بهداشت جهانی، فشارخون بالا بعد از دخانیات، به عنوان دومین عامل خطر اصلی مرگ در کشورهای در حال توسعه و سومین عامل خطر اصلی برای بیماری‌های غیرواگیر معرفی شده است. شیوع این بیماری در جوامع گوناگون از ۱۰درصد تا بیش از ۶۰ درصد متغیر می‌باشد [۷].

عوامل خطر پرفشاری خون به دو دسته غیر قابل تغییر مانند جنسیت، سن و سابقه فامیلی و قابل تغییر مانند چاقی، کم تحرکی، استرس، رژیم غذایی نامناسب و مصرف قرص‌های پیش‌گیری از بارداری در زنان تقسیم می‌شوند. طبق مطالعاتی که تاکنون انجام گرفته است فاکتورهای سن، سیگار کشیدن، نمایه توده بدنی، رژیم غذایی پر چرب، مصرف نمک، فعالیت و تحرک بدنی کم، سطح سواد و نوشیدن مشروبات الکلی با پرفشاری خون رابطه داشته‌اند ولی ارتباط بین اختلال عملکرد کبد و ایجاد فشارخون بالا به خوبی شناخته نشده است [۸].

آنزیم کبدی گاما گلوتامیل ترانسفراز به‌عنوان نشان‌گر بیولوژیکی مصرف بیش‌ازحد الکل و استرس اکسیداتیو استفاده می‌شود. هم‌چنین به عنوان شاخص اختصاصی عملکرد کبد به راحتی در سرم قابل اندازه‌گیری بوده و بنابراین تحقیقات بیش‌تر بر روی آنزیم کبدی گاما گلوتامیل ترانسفراز متمرکز شده است و شواهد نشان می‌دهد که ممکن است یک شاخص خطرناک برای پیش‌بینی مرگ و میر ناشی از تمام علل باشد [۹]. لذا در مطالعه حاضر، به تأثیر آنزیم کبدی گاما گلوتامیل ترانسفراز بر روی پرفشاری خون با روش همسان‌سازی نمره گرایش پرداخته شده است. این ارتباط ممکن است تا حد زیادی تحت تأثیر عوامل متعدد مخدوش‌گر باشد. بنابراین، با استفاده از رویکرد همسان‌سازی نمره گرایش به عنوان یکی از راه‌حل‌های ممکن با کنترل اثر متغیرهای مخدوش‌کننده، برای تخمین برآورد مناسب، از اثر این عوامل مخدوش کننده جلوگیری شد.

به‌طور کلی طرح استنباط علّی در مطالعات مشاهده‌ای یا مطالعات بدون تصادفی‌سازی امری چالش برانگیز است و این

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

دلیلی است که متخصصین آمار و اپیدمیولوژیست‌ها را به تحقیق در مورد روش‌های جدید تحلیلی، ترغیب کرده است. چالش‌های این چنینی همانند استخراج استنباط علیتی از مطالعات مشاهده‌ای باعث نشأت گرفتن مطالعات بیش‌تر در زمینه روش‌های جدید تحلیلی و هم‌چنین باعث تمرکز بیش‌تر در مراحل انجام تحلیل نمره گرایش شده است [۱۰]. لذا با توجه به اهمیتی که این روش در سال‌های اخیر پیدا کرده است، در این مقاله سعی شده است که مراحل انجام روش همسان‌سازی نمره گرایش به صورت مختصر و با زبان ساده با استفاده از یک مثال واقعی توضیح داده شود. مثال استفاده شده در این مقاله، یک مطالعه کوهورت آینده‌نگر می‌باشد که در متن مقاله از واژه‌های مواجهه و عدم مواجهه برای همسان‌سازی به روش نمره گرایش استفاده شده است.

چرا، چه موقع و چگونه به انجام همسان‌سازی نمره گرایش نیاز داریم؟

در تحقیقات به‌خصوص در علوم رفتاری، بهداشتی و اجتماعی، کارآزمایی بالینی تصادفی شده، استاندارد طلایی برای تحقیق تلقی می‌شود، اما مطالعات تصادفی همواره از نظر عملی، اخلاقی و حتی مطلوب بودن همیشه ممکن نیستند ولی امکان طراحی مطالعات مشاهده‌ای وجود دارد. در طول ۳۵ سال گذشته، محققان هنگام ارزیابی برنامه‌های مبتنی بر مطالعات مشاهده‌ای نیاز به رویکردهای کارآمد و مؤثرتر برای ارزیابی اثرات مواجهه تشخیص دادند. در پاسخ به این نیاز، تغییرات چشم‌گیری در روش‌های ارزیابی ایجاد شده است [۱۱].

در حال حاضر در مطالعات مشاهده‌ای، عمدتاً برای کاهش سوگیری و افزایش دقت مطالعه و هم‌چنین برای تعیین روابط علیتی با کم‌ترین مقدار خطا، زمانی که متغیرهای مخدوش‌گر مطالعه زیاد باشند می‌توان از روش‌های تطبیق نمره گرایش استفاده نمود [۱۲]. از این روش می‌توان برای متوازن کردن متغیرهای مخدوش‌گر در افرادی که در مواجهه هستند و هم‌چنین افرادی که تحت مواجهه قرار نمی‌گیرند، استفاده کرد. به این صورت که نمره گرایش اجازه می‌دهد اثرات مواجهه را با شبیه‌سازی به یک مطالعه تصادفی به سبک ناپارامتریک (Randomized experiment in a non-parametric fashion) برآورد نماییم، یعنی مشاهدات در گروه مواجهه با مشاهداتی که در گروه عدم مواجهه (گروه کنترل) قرار دارند تطبیق داده می‌شوند [۱۳]. در نهایت، هر مشاهده در گروه مواجهه با یک یا چند مشاهده در گروه کنترل بر اساس مقدار نمره گرایش متناظر می‌شود. نتیجه پیش‌بینی این است که تفاوت‌های موجود در نتایج در میان هر جفت همسان‌سازی شده می‌تواند بیش‌تر به جهت اثر مواجهه باشد و نه به جهت تفاوت‌های قابل مشاهده بین جفت‌ها (Observable differences between pairs) [۱۴]. به‌طور کلی می‌توان کاربرد نمره گرایش در مطالعات مشاهده‌ای را جایگزین تصادفی‌سازی در مطالعات تجربی کرد و هم‌چنین رویکردهای تطبیق، یکی از

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

راه‌حل‌های ممکن برای کاهش خطاهای اندازه‌گیری، کاهش تورش انتخاب و هم‌چنین با کنترل متغیرهای مخدوش‌گر تجزیه‌وتحلیل را راحت‌تر می‌کند [۱۵].

مواد و روش‌ها

روش‌های مختلفی برای همسان‌سازی وجود دارد از جمله: روش همسان‌سازی با نمره گرایش (Propensity score matching)، طبقه‌بندی بر روی نمره گرایش (Stratification on the propensity score)، شیوه معکوس وزن‌دهی (Inverse probability of treatment weighting; IPTW)، تعدیل متغیرها با استفاده از نمره گرایش (Covariate adjustment using the propensity score)، و یکی از رایج‌ترین روش‌ها برای ایجاد گروه‌های همسان استفاده از روش یادگیری ماشین (Machine learning) است [۱۵]، که در این روش می‌توان برای برآورد نمره گرایش دقیق‌تر از دو تکنیک مختلف Machine learning شامل (Least absolute shrinkage and selection operator; LASSO) و (Random forest) استفاده کرد. شواهد نشان می‌دهد که روش LASSO برای برآورد نمره گرایش به‌خوبی عمل می‌کند، ولی استفاده از Random forest برای این منظور ممکن است منجر به نتایج گمراه کننده شود، به این دلیل که Random forest قادر به این نیست که متغیرهای مخدوش‌گر را به درستی بین دو گروه متعادل کند. بنابراین از اثر متغیرهای مخدوش‌گر جلوگیری نمی‌شود و هم‌چنین سوءگیری در مطالعه باقی می‌ماند [۱۶].

از آن‌جایی که به نظر می‌رسد دانستن این‌که کدام یک از روش‌ها می‌تواند به‌طور مناسب‌تر عمل کند دشوار است، یک جایگزین قابل قبول این است از روش‌های Causal machine learning، به‌عنوان مثال Double machine learning پیشنهاد شده توسط Chernozhukov و یا Modified causal forest مطرح شده توسط Lechner، که به‌طور خاص برای تخمین اثر درمان، بهینه شده‌اند استفاده شود. به‌طور کلی تخمین نمره گرایش با روش Machine learning در مقایسه با روش‌های تطبیق مرسوم فعلی به وضوح سودمند نیست. در عوض روش‌های Causal machine learning که مستقیماً برای تخمین اثر درمان بهینه‌سازی شده‌اند، ممکن است جایگزین امیدوارکننده‌تری باشند [۱۷]. همان‌طور که مورد اشاره قرار گرفت روش‌های مختلفی برای همسان‌سازی وجود دارد اما در مقاله حاضر روش همسان‌سازی با نمره گرایش توضیح داده می‌شود.

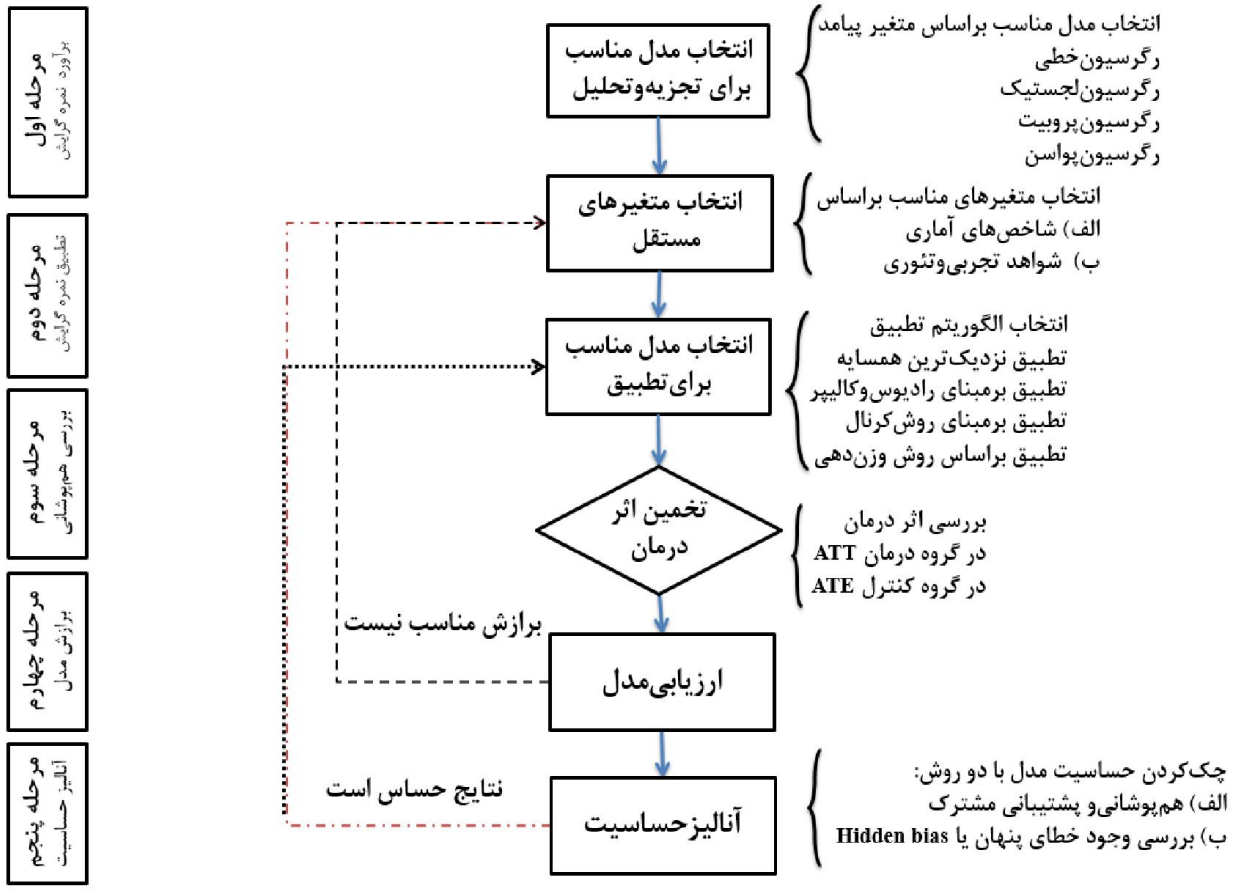
مراحل انجام تطبیق نمره گرایش

همان‌طورکه نمودار ۱ نشان می‌دهد، مراحل انجام همسان‌سازی با نمره گرایش شامل پنج مرحله است. مرحله اول برآورد نمره گرایش است، این مرحله شامل انتخاب مدل مناسب و انتخاب متغیرهای مخدوش‌گر می‌باشد. در مرحله دوم، بر اساس نمرات گرایش برآورد شده (در مرحله اول) باید تصمیم گرفت از میان الگوریتم‌های مختلف، بهترین

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

الگوریتم را انتخاب کنیم. مرحله سوم هم‌پوشانی و پشتیبانی مشترک می‌باشد که در این مرحله مشاهداتی که خارج از محدوده نمرات گرایش قرار دارند، حذف می‌شوند. پس از آن، کیفیت مطابقت ارزیابی می‌شود.

در نهایت، در مرحله آخر باید حساسیت اثرات برآورد شده را تخمین زد. در صورت مناسب نبودن هر مرحله جهت اصلاح به مرحله قبل باز می‌گردیم [۱۸].



نمودار۱ - مراحل انجام روش نمره گرایش [۱۹]

مرحله اول: برآورد نمره گرایش و انتخاب متغیر مناسب

نخستین مرحله در روش همسان‌سازی بر اساس نمره گرایش، برآورد نمره گرایش است. برآورد نمره گرایش در عمل مجهول است که بسته به نوع متغیر پیامد از مدل‌های مختلف رگرسیونی از قبیل رگرسیون لجستیک (Logistic)، پروبیت (Probit) و کاکس (Cox model) استفاده می‌شود. متغیر پیامد در مطالعه حاضر، فشارخون است که به صورت

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

دو حالتی ابتلاء به فشارخون و عدم ابتلاء به فشارخون تعریف می‌شود. با توجه به دو حالتی بودن پیامد مورد بررسی (ابتلاء و عدم ابتلاء به فشارخون) در مطالعه حاضر، مناسب‌ترین مدل برای برآورد بهتر نمره گرایش، استفاده از مدل لجستیک می‌باشد. در برآورد نمره گرایش، مدل پروبیت استفاده کم‌تری دارد و لذا اغلب برای برآورد نمره گرایش، رگرسیون لجستیک به کار گرفته می‌شود. رگرسیون لجستیک چندین مزیت دارد: به‌عنوان یک ابزار آماری آشنا و کاملاً شناخته شده برای محققان می‌باشد، هم‌چنین با استفاده از بسته‌های نرم‌افزاری استاندارد آماری قابل اجرا است. هدف از انتخاب مدل مناسب این است که بتوانیم با انتخاب متغیرهای مناسب، برآورد بهتری از نمره گرایش داشته باشیم [۲۰].

همان‌طور که در نمودار ۱ قابل مشاهده می‌باشد، بعد از انتخاب مدل مناسب برای برآورد نمره گرایش، نوبت انتخاب متغیرهایی است که در مدل گنجانده می‌شود. این قسمت بسیار حائز اهمیت است و یکی از مشکل‌ترین مراحل نمره گرایش است. روش‌های مختلفی برای انتخاب متغیرهای مناسب وجود دارد که از مهم‌ترین آن‌ها می‌توان به بررسی مطالعات پیشین، استفاده از نظرات متخصصین و روش‌های آماری مناسب اشاره نمود [۲۱]. ولی باید در نظر داشت متغیرهای موجود در مسیر علیتی و متغیرهایی که به شدت با پیامد مرتبط هستند با توجه به این‌که ممکن است برآورد اثر مواجهه را تحت تأثیر قرار دهند، نباید در مدل نمره گرایش قرار بگیرند [۲۲]. به‌طورکلی نباید فراموش کرد که نمره گرایش برآوردی از متغیرهای شناخته شده مانند سن و جنسیت را به دست می‌دهد و متغیرهای غیرقابل اندازه‌گیری (شناخته نشده) مانند مارکرهای ژنتیک را برآورد نمی‌کند. بنابراین، ارائه گزارش دقیق از متغیرهای گنجانده شده در مدل نمره گرایش مهم است [۲۳].

مرحله دوم: انتخاب روش تطبیق

پس از برآورد نمره گرایش در مرحله دوم بایستی گروه مواجهه و گروه کنترل با همدیگر تطبیق شوند. هدف اصلی در این مرحله ایجاد دو گروه از شرکت‌کنندگان (بازوهای مطالعه) که از نظر برآورد نمره گرایش تا حد امکان مشابه هم هستند، می‌باشد. بنابراین زمانی که متغیرها بر اساس نمره گرایش مشابه بین دو گروه تقسیم می‌شوند مانند این است که اصل تصادفی‌سازی رعایت شده باشد [۲۴]. در تطبیق به روش نمره گرایش، انتظار می‌رود متغیرهای اساسی مشاهده شده بین گروه‌های مواجهه و عدم مواجهه متعادل شوند. با این حال، اگر مدل آماری مورد استفاده برای محاسبه نمره گرایش به درستی انتخاب نشده باشد عدم تعادل بین ویژگی‌های پایه دو گروه هنوز هم می‌تواند وجود داشته باشد. بنابراین، استفاده از مدل آماری مناسب برای ایجاد تعادل بین متغیرهای گروه‌های مورد مطالعه از اهمیت ویژه‌ای برخوردار است [۲۵]. مزیت استفاده از همسان‌سازی بر مبنای نمره گرایش نسبت به روش‌های همسان‌سازی کلاسیک در این است که در این روش امکان همسان‌سازی

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

به‌طور هم‌زمان بر اساس تعداد زیادی متغیر فراهم می‌شود و هم‌چنین در مواردی که پیامد مورد مطالعه (Outcome) نادر باشد، ولی درمان یا مواجهه معمول باشد نسبت به روش‌های کلاسیک برتری دارد [۲۶].

به طور کلی می‌توان با استفاده از یکی از ۵ روش تطبیق شامل: نزدیک‌ترین همسایگان (Nearest neighbors matching) (افراد با نمره گرایش نزدیک به هم)، کالیپر و رادیوس (Caliper and radius)، طبقه‌بندی (Stratification and interval)، کرنل (Kernel and local linear) و وزن‌دهی (Weighting) بر اساس نمره گرایش محاسبه شده، برای برآورد اثر مواجهه (یا اثر متغیر مستقل) با کنترل اثر متغیرهای مخدوش‌گر احتمالی بهره برد [۳]. روش‌های تطبیق می‌توانند به انتخاب افراد در گروه کنترل کمک کنند که بیش‌ترین شباهت را به افراد در گروه مواجهه دارند. در واقع تطبیق، آن دسته از افراد در گروه کنترل که بیش‌ترین شباهت را به افراد در گروه مواجهه دارند، شناسایی می‌کند. با انجام درست تطبیق از سوگیری می‌توان جلوگیری نمود [۲۷]. انتخاب الگوریتم مناسب برای تطبیق، به اهداف تجزیه‌وتحلیل و خصوصیات منحصر به فرد هر مجموعه داده، حجم نمونه، تعداد مشاهدات در گروه کنترل و مواجهه، و در نهایت توزیع نمره گرایش بستگی دارد [۲۸]. توضیحات مربوط به هر روش تطبیق در نمودار ۲ به‌طور خلاصه قابل مشاهده است.

در این مطالعه از روش تطبیق نزدیک‌ترین همسایگان، برای تطبیق افراد در دو گروه با سطح آنزیم کبدی بالا و پایین استفاده شده است. برای این منظور، افرادی که دارای سطح بالای آنزیم کبدی گاما گلوتامیل ترانسفراز هستند به عنوان گروه مواجهه، با افرادی که دارای سطح پایین آنزیم کبدی گاما گلوتامین ترانسفراز هستند به عنوان گروه عدم مواجهه با نمره گرایش نزدیک به‌هم تطبیق داده می‌شوند. استفاده از تطبیق یک به یک (به ازای یک فرد در گروه مواجهه یک کنترل انتخاب شود) متداول‌تر است. در میان الگوریتم‌های تطبیق، روش نزدیک‌ترین همسایگان یکی از پرکاربردترین شیوه‌های این نوع همسان‌سازی است. هم-چنین، با انجام این روش تطبیق می‌توان اثر تعداد متغیرهای مخدوش‌گر بیش‌تری را نسبت به روش کالیپر تعدیل کرد [۱۲]. فرض کنید فردی با سطح بالای آنزیم کبدی گاما گلوتامیل ترانسفراز در گروه مواجهه دارای نمره گرایش ۰/۰۳ با فاصله اطمینان (۰/۰۴۲۰ - ۰/۰۲۸۲) می‌باشد. این فرد با فردی که دارای سطح پایین آنزیم کبدی می‌باشد در گروه کنترل (عدم مواجهه) با نمره گرایش ۰/۰۲ و فاصله اطمینان (۰/۰۴۲۰ - ۰/۰۱۹۱) با روش تطبیق نزدیک‌ترین همسایگان (افراد با نمره گرایش نزدیک به هم) تطبیق داده می‌شود [۲۹].

در ادامه، همان‌طور که در نمودار ۲ ملاحظه می‌شود، زمانی که نمره گرایش بین گروه مواجهه و گروه کنترل (عدم مواجهه) از هم فاصله بیش‌تری دارند روش تطبیق نزدیک-

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

ترین همسایه کارآیی کم‌تری دارد. بنابراین، برای انجام تطبیق مناسب‌تر، بهتر است از روش کالیپر استفاده شود [۳۰]. یکی دیگر از الگوریتم‌های تطبیق، طبقه‌بندی می‌باشد. ایده اصلی استفاده از روش طبقه‌بندی برای توازن داده‌ها توسط Cochran در سال ۱۹۶۸ مطرح شد. طبقه‌بندی در نمره گرایش شامل مقایسه نتایج بین افرادی که مواجهه دارند و افرادی که مواجهه ندارند در لایه‌هایی است که توسط نمره گرایش تعریف شده است. متداول‌ترین روش، استفاده از پنج لایه تقریباً به اندازه یکسان است. بنابراین تأثیر مواجهه بر روی نتایج، در هر طبقه یا لایه تخمین زده می‌شود و سپس اثرات مواجهه خاص هر لایه برای دست‌یابی به یک اثر مواجهه کلی ادغام می‌شوند. به گفته Cochran، طبقه‌بندی به صورت پنج لایه از نمره گرایش تخمین زده شده تقریباً ۹۰ درصد سوءگیری را به دلیل خطاهای اندازه‌گیری شده از بین می‌برد. این روش زمانی که تعداد افراد در گروه کنترل (عدم مواجهه) نسبت به افراد در گروه مواجهه بیش‌تر باشد، بهترین استفاده را دارد [۳۱].

الگوریتم دیگر تطبیق بر پایه کرنال است. این روش توسط Heckman، Ichimura و Todd (۱۹۹۷-۱۹۹۸) توسعه داده شد. در این روش از میانگین وزنی افراد گروه کنترل استفاده می‌شود و برخلاف سه روش قبلی که مورد بحث قرار گرفت، مهم نیست که نمره گرایش آن‌ها نزدیک به هم باشد. هم چنین، از این روش می‌توان برای زمانی که داده‌ها به صورت دو زمانه (به عنوان مثال، تجزیه و تحلیل داده‌های قبل و بعد از مواجهه) هستند، برای نشان دادن تغییر ناشی از یک مداخله به روشی پویا استفاده کرد [۳۲].

در نهایت، آخرین الگوریتم تطبیق بر اساس وزن‌دهی می‌باشد. این روش را می‌توان زیر مدلی از مدل‌های توسعه یافته توسط Rosenbaum و Rubin (۱۹۸۳) دانست. این روش تجزیه و تحلیل چند متغیره را با استفاده از نمرات گرایش به عنوان وزن نمونه توصیف می‌کند. جذابیت بیش‌تر این روش به طور معمول اجازه می‌دهد تا بیش‌تر شرکت کنندگان در نتیجه‌گیری تحقیق حفظ شوند. به همین دلیل، این روش به طور گسترده در علوم اجتماعی و بهداشتی مورد استفاده قرار گرفته است [۳۳].

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]



نمودار ۲- الگوریتم‌های تطبیق [۱۹]

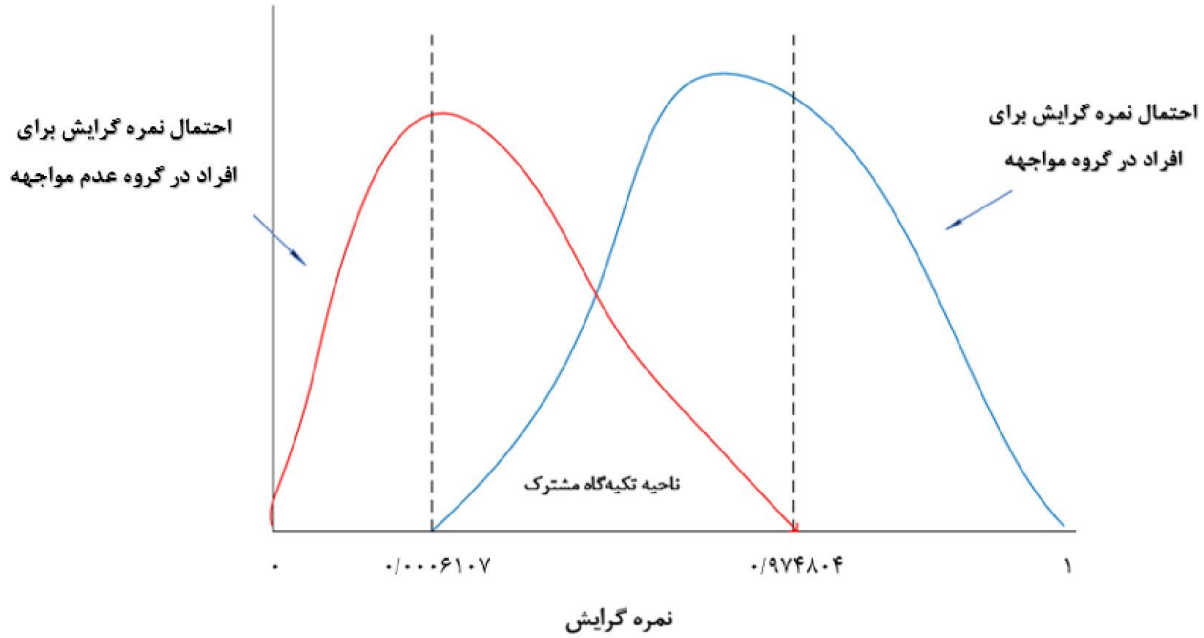
مرحله سوم: بررسی هم‌پوشانی و پشتیبانی مشترک

مرحله سوم همسان‌سازی به روش نمره گرایش، مرحله هم‌پوشانی و پشتیبانی مشترک (Overlap and common support) نامیده می‌شود. این مرحله گام مهمی برای بررسی هم‌پوشانی و پشتیبانی مشترک بین دو گروه مواجهه و عدم مواجهه به حساب می‌آید. در این مرحله حداقل و حداکثر نمره گرایش برآورد می‌شود. همان‌طور که در نمودار ۳ مشخص شده است، افرادی که نمره گرایش آن‌ها در گروه مواجهه و عدم مواجهه بزرگ‌تر از حداکثر و کم‌تر از حداقل نمره گرایش برآورد شده باشد، از مدل حذف می‌شوند و افرادی که نمره گرایش آن‌ها در ناحیه تکیه‌گاه مشترک قرار دارد، به‌عنوان افرادی که نمره گرایش آن‌ها بین مقادیر حداکثر و حداقل قرار دارد در مدل باقی می‌مانند. در واقع معیار اصلی این روش حذف کلیه افرادی است که نمره گرایش آن‌ها در گروه مقابل بزرگ‌تر از حداکثر و کم‌تر از حداقل است. هم‌چنین، اجرای شرایط پشتیبانی مشترک تضمین می‌کند که هر ترکیبی از خصوصیات مشاهده شده در گروه مواجهه می‌تواند در میان گروه عدم مواجهه یا کنترل نیز مشاهده شود [۲۴].

در مطالعه انجام گرفته در کوهورت روانسر در استان کرمانشاه، حداکثر و حداقل نمره گرایش برآورد شده به ترتیب برابر با ۰/۹۷۴۸۰۴ و ۰/۰۰۰۶۱۰۷ است. مشاهداتی

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

که نمره گرایش آن‌ها بزرگ‌تر یا کم‌تر از این مقادیر بودند، حذف شده اند. هم‌چنین، در این مرحله مقدار میانگین اثر درمان (اثر مواجهه) در گروه درمان (Average treatment effect on the treated; ATT) و میانگین اثر درمان (Average treatment effect; ATE) برآورد شد. البته قبل از اعتماد و استفاده از مقدار برآورد شده ATT باید برازش روش همسان‌سازی نمره گرایش بررسی شود که یکی از راه‌های آن استفاده از مقدار سوءگیری استاندارد شده می‌باشد که مقدار آن بعد از همسان‌سازی باید کم‌تر از ۵ درصد باشد [۳۴].



نمودار ۳- ناحیه هم‌پوشانی و پشتیبانی مشترک [۳۵]

مرحله چهارم: شاخص برازش مدل

مرحله چهارم به شاخص برازش مدل (Goodness-of-fit test; GOF) معروف است. موضوعاتی که در این مرحله مورد ارزیابی قرار می‌گیرد، عبارت است از این‌که آیا تعادل در نمره گرایش بین دو گروه مواجهه و عدم مواجهه وجود دارد یا خیر. هم‌چنین، آیا تطبیق انجام شده از کیفیت مناسب برخوردار است و در نهایت اثرات مواجهه تخمین زده شده از سوءگیری انتخاب برخوردار هستند یا خیر؟

شاخص‌های مختلفی برای برازش مدل وجود دارد که از مهم‌ترین آن‌ها عبارتند از: تورش استاندارد شده (Standardized bias)، آزمون تی‌تست (T-test)، آزمون طبقه‌بندی (Stratification test) و در نهایت سودو آردو

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

(Pseudo-R^2) که شاخص مناسب‌تری می‌باشد. مقدار Pseudo-R^2 قبل و بعد از تطبیق با هم مقایسه می‌شود، چنان‌چه بعد از انجام تطبیق، مقدار Pseudo-R^2 کم شده باشد بدین معنی است که با کاهش سوءگیری انتخاب، اختلاف منظمی در توزیع متغیرهای بین دو گروه وجود ندارد و تعادل بین دو گروه برقرار است. هم‌چنین، تطبیق از کیفیت مناسب برخوردار است [۳۶].

در ادامه هم‌چنین دو معیار از جمله (Akaike information criterion; AIC) و Bayesian information criterion; BIC) برای سنجش برازش مدل مورد استفاده قرار می‌گیرد. این شاخص‌ها تعادلی میان دقت مدل و پیچیدگی آن برقرار می‌کنند، هم‌چنین برای انتخاب بهترین مدل آماری پیشنهاد می‌شوند [۳۷]. همان‌طور که در جدول ۱ قابل ملاحظه می‌باشد، مقدار AIC و BIC بعد از تطبیق مدل کاهش یافته‌اند. در واقع هر قدر این دو شاخص کوچک‌تر باشند، مدل مورد نظر نسبت به بقیه مدل‌ها بهتر و مناسب‌تر است. معمولاً AIC و BIC هم‌جهت هستند ولی اگر هم جهت نبودند، BIC قدرت پیش‌بینی کننده بهتری دارد [۳۸].

یکی دیگر از شاخص‌های برازش مدل، ضریب تعیین یا ضریب تشخیص (Coefficient of determination) می‌باشد که با علامت R^2 نشان داده می‌شود. این شاخص یکی از شاخص‌های برازش مدل است که قدرت پیش‌بینی متغیر وابسته (ملاک) بر اساس متغیرهای مستقل (پیش‌بین) را نشان می‌دهد. مقدار این شاخص بین صفر تا یک می‌باشد و اگر از ۰/۶ بیش‌تر باشد، نشان می‌دهد متغیرهای مستقل تا حد زیادی توانسته‌اند تغییرات متغیر وابسته را تبیین کنند [۳۹]. همان‌طور که در جدول ۱ قابل ملاحظه است، R^2 بعد از تطبیق و در مدل دوم برابر با ۰/۹ شده که نشان‌دهنده این است متغیرهای مستقل تا حد زیادی توانسته‌اند متغیر پیامد را پیش‌بینی کنند.

مرحله پنجم: آنالیز حساسیت

در نهایت حساسیت (Sensitivity analysis) اثرات برآورد شده مورد بررسی قرار می‌گیرد. Rosenbaum شرح مختصر و منظمی از تجزیه و تحلیل حساسیت ارائه می‌دهد، به این صورت که آیا متغیرهای مخدوش‌گر اثرات برآورد شده را تحت تأثیر قرار داده‌اند یا خیر. به‌طور کلی در این مرحله با بررسی حساسیت مدل متوجه خواهیم شد که هرگونه تغییر منجر به چه چیزی خواهد شد و یا بهترین تغییر چه تغییری خواهد بود [۴۰].

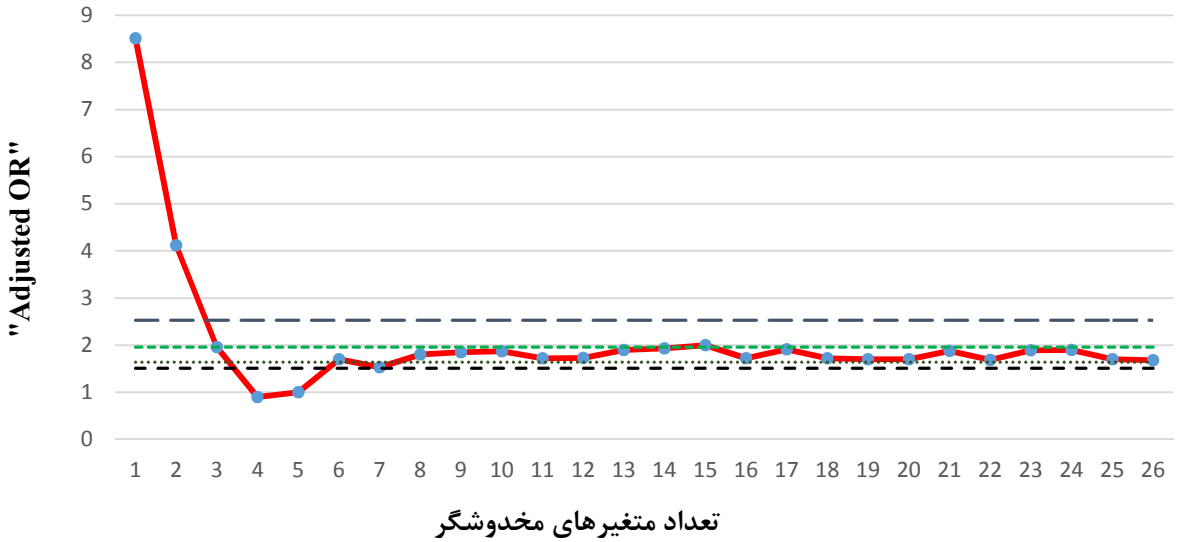
در حساسیت مدل با دو روش، تحت عنوان خطای پنهان (Hidden bias) و پشتیبانی مشترک می‌توان موارد اشاره شده را بررسی نمود. اگر متغیرهای مشاهده نشده‌ای وجود داشته باشند که بر روی متغیر پیامد تأثیر گذارند، ممکن است یک "خطای پنهان" ایجاد شود. روشن است که برآوردهای تطبیق در برابر این "خطای پنهان" قوی نیستند. بنابراین می‌توان با بررسی و جایگزین کردن الگوریتم‌های تطبیق این خطا را کاهش داد [۴۱].

با توجه به این‌که یکی از مفروضات آنالیز حساسیت نبود متغیر مخدوش‌گر است، بررسی حساسیت به روش پشتیبانی

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

مشترک به خوبی به بررسی این موضوع می‌پردازد. با استفاده از روش پشتیبانی مشترک، برآورد اثر درمان (ATT) را در زمانی که متغیر مخدوش‌گری در داخل مدل نیست و هم چنین زمانی که متغیر مخدوش‌گر در داخل مدل قرار دارد به‌دست می‌آوریم. به عنوان مثال، در مطالعه کوهورت بزرگ سالان روانسر این مقدار برای هر دو حالت (مقدار برآورد اثر درمان، زمانی که هیچ متغیر مخدوش‌گری در داخل مدل نیست و زمانی که متغیر مخدوش‌گر در داخل مدل قرار دارد) برابر با ۰/۰۰۲ شد. بنابراین، با توجه به مساوی بودن مقدار برآورد اثر درمان در هر دو حالت می‌توان گفت با تعدیل کردن متغیرهای مخدوش‌گر از اثر آن‌ها جلوگیری شده است. البته همیشه مقدار برآورد اثر درمان در هر دو حالت با هم برابر نیستند ولی هر چقدر به هم‌دیگر نزدیک‌تر باشند نشان‌دهنده این است که با تعدیل متغیرهای مخدوش‌گر اثر آن‌ها بر روی ارتباط به‌دست آمده کنترل شده است [۴۲].

در مطالعه بررسی ارتباط بین آنزیم کبدی گاما گلوتامیل ترانسفراز و پرفشاری خون، برداری از متغیرهای مخدوش‌گر مورد بررسی قرار گرفت. مطابق با نمودار ۴، در این نمودار، محور عمودی نسبت شانس تعدیل شده گزارش شده است و محور افقی تعداد متغیرهای مخدوش‌گری است که وارد مدل شده‌اند. همان‌طور که در نمودار قابل ملاحظه است با تعدیل کردن تعداد متغیرها، برآورد نسبت شانس با ثبات‌تر خواهد شد و در نهایت به یک نسبت شانس تقریباً ثابتی دست می‌یابیم. به‌طور کلی نسبت شانس متغیرها به برآورد کلی نزدیک می‌شود [۴۳].



نمودار ٤- تعدیل اثر متغیرهای مخدوش‌گر

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

نتایج

در مطالعه کوهورت حاضر، بعد از اعمال معیارهای ورود و خروج از ۱۰۰۰۰ نفر، ۸۲۶۷ مورد انتخاب شد که بر اساس تعریف ۳۱۹۵ فرد با سطح آنزیم کبدی بالا (گروه مواجهه) و ۵۰۷۲ فرد با سطح آنزیم کبدی پایین (گروه عدم مواجهه) وارد مطالعه شدند. بعد از تعدیل کردن متغیرهای مخدوش-گر شامل: سن، جنسیت، وضعیت تأهل، نمایه توده بدنی، سطح تحصیلات، محل سکونت، مصرف سیگار، مصرف الکل، مصرف نمک، وضعیت اقتصادی و اجتماعی، بعد خانوار، فعالیت فیزیکی، شاخص تغذیه سالم، افسردگی، دیابت، بیماری قلبی-عروقی، سابقه بیماری، وضعیت شغلی، نوع روغن مصرفی، پروتئینوری، آلبومین ادرار، گلوکز سرم، کلسترول تام، تری گلیسیرید، لیپوپروتئین با چگالی پایین (Low-density lipoprotein; LDL) و لیپوپروتئین با چگالی بالا (High-density lipoprotein; HDL)، تنها دو متغیر جنسیت و گروه سنی معنی‌دار شدند که نتایج این دو متغیر معنی‌دار به‌طور جداگانه در جدول ۱ مورد اشاره قرار گرفته است. بنابراین نتایج حاصل از تجزیه و تحلیل رگرسیون لجستیک چند متغیره نشان می‌دهد که افراد با سطح بالای آنزیم کبدی گاما گلوتامیل ترانسفراز (گروه مواجهه)، شانس ابتلاء به فشارخون بالاتری نسبت به افراد در گروه با سطح پایین آنزیم کبدی گاما گلوتامیل ترانسفراز (گروه عدم مواجهه) به صورت OR = ۱/۲۸، ۱/۳۳ – ۱/۰۲ :فاصله اطمینان ۹۵ درصد دارند.

ولی بر اساس روش نمره گرایش، ۲۰۶۰ فرد با نمره گرایش نزدیک به هم در بین دو گروه قرار گرفتند. یعنی ۱۰۳۰ فرد با سطح آنزیم کبدی بالا و ۱۰۳۰ فرد با سطح آنزیم کبدی پایین وارد مطالعه شدند و سایر افراد به دلیل نبود گروه تطبیق مناسب از مدل خارج شدند که ارتباط بین آنزیم کبدی گاما گلوتامیل ترانسفراز و پرفشاری خون به صورت OR = ۱/۳۲، ۳/۱۲ – ۱/۳۰ :فاصله اطمینان ۹۵ درصد، به‌دست آمد که نسبت به مدل رگرسیون لجستیک چندمتغیره متفاوت است. این مطالعه نشان می‌دهد سطح آنزیم کبدی گاما گلوتامیل ترانسفراز با پرفشاری خون ارتباط مثبت و معنی‌داری دارد، با توجه به این‌که بعد از استفاده از تطبیق بر اساس نمره گرایش نیز این ارتباط هم‌چنان باقی مانده است.

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

جدول ۱- برآورد نسبت شانس (Adjusted odds ratio; AOR) با روش تطبیق نمره گرایش و مدل رگرسیون لجستیک چندمتغیره در رابطه بین آنزیم‌های کبدی گاما گلوتامیل ترانسفراز و پرفشاری خون

متغیر		مدل کامل با روش رگرسیون لجستیک چندمتغیره		تطبیق با روش نمره گرایش	
		AOR (فاصله اطمینان ۹۵٪)	مقدار p	AOR (فاصله اطمینان ۹۵٪)	مقدار p
جنسیت	زن (رفرنس)	۱	۱	۱	۱
	مرد	۱/۳۶ (۱/۰۷ – ۱/۷۴)	۰/۰۱۴	۱/۱۵ (۱/۰۳ – ۱/۲۶)	۰/۰۱۲
گروه سنی (سال)	۴۵-۳۵ (رفرنس)	۱	۱	۱	۱
	۵۵-۴۶	۱/۰۷ (۱/۰۳ – ۱/۲۹)	۰/۰۳۳	۱/۰۳ (۱/۰۱ – ۱/۱۲)	۰/۰۳۲
	۶۵-۵۶	۱/۱۲ (۱/۰۹ – ۱/۳۵)	۰/۰۴۲	۱/۰۹ (۱/۰۴ – ۱/۱۷)	۰/۰۴۰
گاما گلوتامیل ترانسفراز	عدم مواجهه (رفرنس)	۱	۱	۱	۱
	مواجهه	۱/۲۸ (۱/۰۶ – ۱/۵۴)	۰/۰۰۲	۱/۳۲ (۱/۰۲ – ۱/۴۲)	۰/۰۰۱
شاخص‌های برازش مدل	Pseudo-R^2	۰/۰۹۱		۰/۰۸۳	
	R^2	۰/۵۳۰		۰/۹۵۰	
	AIC	۳۵۱۵/۰۰۹		۱۴۹۷/۹۶۷	
	BIC	۳۷۱۱/۵۷۰		۱۵۳۳/۱۶۸	

p<۰/۰۵ارتباط معنی‌دار

متغیرهای مخدوش‌گر موجود در هر دو مدل (نمره گرایش و رگرسیون لجستیک چندمتغیره) شامل: سن، جنسیت، وضعیت تأهل، نمایه توده بدنی، سطح تحصیلات، محل سکونت، مصرف سیگار، مصرف الکل، مصرف نمک، وضعیت اقتصادی و اجتماعی، بعد خانوار، فعالیت فیزیکی، شاخص تغذیه سالم، افسردگی، دیابت، بیماری قلبی-عروقی، سابقه بیماری، وضعیت شغلی، نوع روغن مصرفی، پروتئینوری، آلبومین ادرار، گلوکز سرم، کلسترول تام، تری گلیسیرید، HDL و LDL، می‌باشد.

بحث

همان‌طور که در داده‌های کوهورت روانسر دیده شد، همسان‌سازی با نمره گرایش یک روش آماری است که در آن محقق هر مشاهده در گروه مواجهه را با یک یا چند مشاهده در گروه عدم مواجهه (کنترل) با ویژگی‌های مشابه تطبیق می‌کند. بنابراین، با استفاده از همسان‌سازی محقق می‌تواند تأثیر یک مداخله یا مواجهه را تخمین بزند. در واقع همسان‌سازی نمره گرایش یک روش مفید در تجزیه‌وتحلیل داده‌ها برای تخمین تأثیر یک برنامه یا رویدادی است که تصادفی‌سازی برای آن از نظر اخلاقی امکان‌پذیر نیست.

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

در روش همسان‌سازی با نمره گرایش انتخاب مدل مناسب و متغیرهای مناسب اهمیت زیادی دارد. هر چند انتخاب مدل مناسب بر اساس متغیر پیامد صورت می‌گیرد ولی از بین مدل‌های مختلف، بیش‌تر محققان از روش‌های رگرسیون لجستیک استفاده می‌کنند. البته مدل پروبیت نیز زمانی که توزیع تجمعی متغیر وابسته برحسب متغیرهای مستقل نرمال باشد، استفاده می‌شود [۴۴].

پس از برآورد نمره‌گرایش در مرحله دوم بایستی گروه درمان و کنترل با یکدیگر تطبیق شوند، ۵ روش وجود دارد که از بین آن‌ها دو روش Matching nearest neighbor و Weighting on propensity از بقیه روش‌ها بهتر پاسخ می‌دهد. البته استفاده از روش Kernel and local linear matching برای تطبیق در سال‌های اخیر بیش‌تر شده است. بعد از این‌که بر اساس یکی از الگوریتم‌ها تطبیق انجام گرفت، مرحله سوم متغیرهایی که خارج از تطبیق هستند حذف می‌شوند و مقدار میانگین اثر درمان در گروه درمان (ATT) و میانگین اثر درمان (ATE) برآورد می‌شود. مرحله چهارم شاخص برازش مدل می‌باشد، موضوعاتی که مورد ارزیابی قرار می‌گیرد این است که آیا تعادل در تحلیل نمره گرایش که در مرحله اول محاسبه شد بین دو گروه مواجهه و عدم مواجهه وجود دارد یا خیر. هم‌چنین آیا تطبیقی که انجام گرفت از کیفیت مناسب برخوردار است و در مرحله نهایی، حساسیت اثراتی که تخمین زده می‌شود، مورد بررسی قرار می‌گیرد [۴۵].

از جمله نقاط قوت مطالعه به این موارد می‌توان اشاره کرد که همسان‌سازی یا تطبیق بر اساس نمره گرایش به طور گسترده در تجزیه‌وتحلیل‌های آماری، به‌ویژه در حوزه پزشکی مورد استفاده قرار می‌گیرد. با توجه به بالا بودن هزینه‌های مطالعات کارآزمایی بالینی تصادفی شده، محققان بیش‌تری به مطالعات مشاهده‌ای به عنوان ابزاری برای انجام تحقیقات کم هزینه‌تر روی می‌آورند. به نظر می‌رسد که روش تطبیق نمره گرایش زمانی که به توان آن را در مراحل طراحی مطالعات (از طریق تطبیق یا طبقه‌بندی) ادغام کرد، بیش‌ترین مزایا را به همراه دارد. این مزایا شامل ارائه برآوردهای دقیق‌تر از اثرات واقعی مواجهه و هم‌چنین صرفه‌جویی در زمان و هزینه است. در نهایت، مهم است توجه داشته باشیم که استفاده از نمره گرایش تنها در تجزیه‌وتحلیل مطالعات مشاهده‌ای توصیه نمی‌شود، بلکه استفاده از نمره گرایش در کنار روش‌های سنتی تحلیل پیشنهاد می‌شود [۱۲].

با این حال، چندین محدودیت باید در مطالعه حاضر پذیرفته شود. همسان‌سازی یا تطبیق بر اساس نمره گرایش فقط متغیرهای قابل مشاهده (قابل اندازه‌گیری) را مورد بررسی قرار می‌دهد، در حالی که متغیرهای دیگری که به عنوان متغیرهای غیرقابل اندازه‌گیری (Unmeasured variable) شناخته می‌شوند در این روش جای نداشته و خطای برآورد حاصل از آن‌ها حتی بعد از تطبیق نیز باقی می‌ماند. بنابراین، لازم است مدل‌های تطبیق نمره گرایش شامل مجموعه بزرگی از متغیرها باشند [۴۶]. برای استفاده

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

از روش همسان‌سازی نمره گرایش لازم است پایگاه داده‌ای (Data base) از داده‌های گروه مواجهه و عدم مواجهه در اختیار داشت تا بتوان به ازای هر فرد مواجهه، یک یا چند غیر مواجهه را انتخاب نمود. لیکن در خیلی از مطالعات مشاهده‌ای، اساساً چنین پایگاه داده‌ای وجود ندارد و محقق مجبور است نمونه‌ای از جامعه آماری (بیمار/سالم و یا مواجهه/عدم مواجهه) انتخاب نماید و داده‌های خود را از روی نمونه جمع‌آوری نماید و بر مبنای آن در مورد جامعه استنباط‌هایی انجام دهد. بنابراین، انجام تطبیق نمره گرایش به نمونه‌های بزرگ با هم‌پوشانی قابل توجه نیاز دارد. با این وجود، لازم است تا جایی که امکان دارد حجم نمونه مورد مطالعه را بالا در نظر گرفت [۴۷].

اگرچه در رابطه با روش همسان‌سازی نمره گرایش تحقیقاتی وجود دارد، اما برای تأیید نتایج کنونی نیاز به مطالعات بیش‌تر است. با این وجود، از مطالعه حاضر می‌توان به‌عنوان مبنایی مناسب برای مطالعات آینده استفاده نمود.

نتیجه‌گیری

با توجه به این‌که پژوهش‌گران درحوزه‌های علوم پزشکی، اجتماعی و روان‌شناسی بیش‌تر اوقات در مطالعات خود به دنبال روابط علیتی می‌باشند، بنابراین لازم است تمرکز بیش‌تر بر روی روش‌های آماری از جمله همسان‌سازی با نمره گرایش انجام گیرد تا بتوان با انجام درست آن از مطالعات مشاهده‌ای نیز به نتایج علیتی دست یافت. روش همسان‌سازی با نمره گرایش می‌تواند به‌عنوان یکی از راه‌حل‌های ممکن به‌منظور انجام مطالعات بیش‌تر با خطای اندازه‌گیری و تورش انتخاب کم‌تر و هم‌چنین کنترل متغیرهای مخدوش کننده، برای تخمین برآورد مناسب مورد استفاده قرار گیرد.

تشکر و قدردانی

نویسندگان از پژوهش‌گران طرح PERSIAN به دلیل حمایت ارزشمندشان در طراحی روش‌ها و تهیه و تدوین پرسش‌نامه و هم‌چنین مصاحبه‌گران، کارکنان RaNCD و جمعیت روانسر که در جمع‌آوری داده‌ها و مشارکت در این تحقیق همکاری داشته‌اند، عمیقاً قدردانی می‌کنند..

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

References

[1] Dehejia RH, Wahba S. Propensity score-matching methods for nonexperimental causal studies. *Rev Econ Stat* 2002; 84(1): 151-61.

[2] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55.

[3] Janani L, Moradian S, Ahmadieh H, Mohammad K, Golestan B, Forouzanfar M. Propensity score analysis for comparing primary vitrectomy with and without encircling band in Pseudophakic/Aphakic retinal detachment. *Bina Journal of Ophthalmology* 2011; 16(2): 100-7.

[4] Cavuto S, Bravi F, Grassi M, Apolone G. Propensity score for the analysis of observational data: an introduction and an illustrative example. *Drug Dev Res* 2006; 67(3): 208-16.

[5] Poustchi H, Eghtesad S, Kamangar F, Etemadi A, Keshtkar A-A, Hekmatdoost A, et al. Prospective epidemiological research studies in Iran (the PERSIAN Cohort Study): rationale, objectives, and design. *American Journal of Epidemiology* 2018; 187(4): 647-55.

[6] Pasdar Y, Najafi F, Moradinazar M, Shakiba E, Karim H, Hamzeh B, et al. Cohort profile: Ravansar Non-Communicable Disease cohort study: the first cohort study in a Kurdish population. *Int J Epidemiol* 2019; 48(3): 682-3f.

[7] Boutouyrie P, Chowienczyk P ,Humphrey JD, Mitchell GF. Arterial stiffness and cardiovascular risk in hypertension. *Circ Res* 2021; 128(7): 864-86.

[8] Rahman S, Islam S, Haque T, Kathak RR, Ali N. Association between serum liver enzymes

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

[Downloaded from journal.rums.ac.ir on 2026-11-21]

[DOR: 20.1001.1.17353165.1401.21.3.6.7]

[DOI: 10.52547/jrums.21.3.343]

and hypertension: a cross-sectional study in Bangladeshi adults. *BMC Cardiovasc Disord* 2020; 20(1): 1-7.

[9] Kunutsor SK, Apekey TA, Cheung BM. Gamma-glutamyltransferase and risk of hypertension: a systematic review and dose–response meta-analysis of prospective evidence. *Journal of Hypertension* 2015; 33(12): 2373-81.

[10] Dehejia RH, Wahba S. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *J Am Stat Assoc* 1999; 94(448): 1053-62.

[11] Imbens GW. The role of the propensity score in estimating dose-response functions. *Biometrika* 2000; 87(3): 706-10.

[12] D'Agostino Jr RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; 17(19): 81-2265.

[13] Ali M, Prieto-Alhambra D, Lopes L, Ramos D, Bispo N, Ichihara M. Propensity score methods in health technology assessment: principles, extended applications, and recent advances. *Front Pharmacol* 2019; 10(2): 973-833.

[14] Elston JA, Hofler R, Lee J. Dividend policy and institutional ownership: Empirical evidence using a propensity score matching estimator. *Journal of Accounting and Finance* 2011; 11(1): 89-102.

[15] Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; 46(3): 399-424.

[16] Goller D, Lechner M, Moczall A, Wolff J. Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany's programmes for long term unemployed. *Labour Econ* 2020; 65(10): 18–55.

[17] Knaus MC, Lechner M, Strittmatter A. Machine learning estimation of heterogeneous causal effects: Empirical monte carlo evidence. *Econom J* 2021; 24(1): 134-61.

[18] Garrido MM, Kelley AS, Paris J, Roza K, Meier DE, Morrison RS, et al. Methods for constructing and assessing propensity scores. *Health Services Research* 2014; 49(5): 1701-20.

[19] Li M. Using the propensity score method to estimate causal effects: A review and practical guide. *Organ Res Methods* 2013; 16(2): 188-226.

[20] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics* 2000; 28(2): 407-337.

[21] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *American Journal of Epidemiology* 2006; 163(12): 1149-56.

[22] Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health* 2001; 22(1): 189-212.

[23] Ali MS, Groenwold RH, Belitser SV, Pestman WR, Hoes AW, Roes KC, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of clinical epidemiology* 2015; 68(2): 122-31.

[24] Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv* 2008; 22(1): 31-72.

[25] Heckman JJ, Vytlacil EJ. Local instrumental variables and latent variable models for identifying and bounding treatment effects.

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

[Downloaded from journal.rums.ac.ir on 2026-11-21]

[DOR: 20.1001.1.17353165.1401.21.3.6.7]

[DOI: 10.52547/jrums.21.3.343]

Proceedings of the National Academy of Sciences 1999; 96(8): 4730-4.

[26] Braitman LE, Rosenbaum PR. Rare outcomes, common treatments: analytic strategies using propensity scores . *Annals of Internal Medicine* 2002; 137(8): 693-5.

[27] Rothman EF, Corso PS. Propensity for intimate partner abuse and workplace productivity: Why employers should care. *Violence Against Women* 2008; 14(9): 1054-64.

[28] Ali MS, Groenwold RH, Klungel OH. Best (but oft-forgotten) practices: propensity score methods in clinical nutrition research. *Am J Clin Nutr* 2016; 104(2): 247-58.

[29] Szekér S, Vathy-Fogarassy Á. Weighted nearest neighbours-based control group selection method for observational studies. *Plos One* 2020; 15(7): e0236531.

[30] Lunt M. Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. *American Journal of Epidemiology* 2014; 179(2): 226-35.

[31] Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. *Stat Med* 2010; 29(20): 2137-48.

[32] Berg GD. An application of kernel-based versus one-to-one propensity score matching for a nonexperimental causal study: example from a disease management program evaluation. *Econ Lett* 2011; 18(5): 439-47.

[33] Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Stat Med* 2010; 29(3): 337-46.

[34] Abadie A, Imbens GW. Large sample properties of matching estimators for average

treatment effects. *Econometrica* 2006; 74(1): 235-67.

[35] Hayatbakhsh MM, Oghabian Z, Conlon E, Nakhaee S, Amirabadizadeh AR, Zahedi MJ, et al. Lead poisoning among opium users in Iran: an emerging health hazard. *Subst Abuse Treat Prev Policy* 2017; 12(1): 43.

[36] Veall MR, Zimmermann KF. Pseudo-R2 measures for some common limited dependent variable models. *J Econ Surv* 1996; 10(3): 241-59.

[37] Pho K-H, Ly S, Ly S, Lukusa TM. Comparison among Akaike information criterion, Bayesian information criterion and Vuong's test in model selection: A case study of violated speed regulation in Taiwan . *Journal of Advanced Engineering and Computation* 2019; 3(1): 293-303.

[38] Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods* 2012; 17(2): 228.

[39] Piepho HP. A coefficient of determination (R2) for generalized linear mixed models. *Biom Z* 2019; 61(4): 860-72.

[40] Rosenbaum PR. Sensitivity analysis for m-estimates, tests, and confidence intervals in matched observational studies. *Biometrics* 2007; 63(2): 456-64.

[41] Kyu HH, Shannon HS, Georgiades K, Boyle MH. Caesarean delivery and neonatal mortality rates in 46 low-and middle-income countries: a propensity-score matching and meta-analysis of Demographic and Health Survey data. *Int J Epidemiol* 2013; 42(3): 781-91.

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

[42] Nannicini T. Simulation-based sensitivity analysis for matching estimators. *The Stata Journal* 2007; 7(3): 334-50.

[43] Cerulli G. Data-driven sensitivity analysis for matching estimators. *Economics Letters* 2019; 185(1): 108-749.

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

[45] Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, et al. A review of propensity-score methods and their use in cardiovascular research. *Canadian Journal of Cardiology* 2016; 32(2): 259-65.

[46] 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(6): 661-77.

[47] Rubin DB. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *J Am Stat Assoc* 1979; 33(7):.

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]

How Can Causal Relationships Be Measured in Observational Studies? Propensity Score Matching: A Tutorial Article

Mina Tahmasebi Fard[1], Farid Najafi[2], Shahab Rezaeian[3], Mehdi Moradinazar[4]

Received: 01/05/22 Sent for Revision: 21/05/22 Received Revised Manuscript: 13/06/22 Accepted: 15/06/22

Background and Objectives: In most observational studies, researchers try to find causal relationships between observations. This correlation is associated with error due to lack of controlling confounding variables. One of the methods used to control the confounding variables is propensity score matching. Therefore, the purpose of this study is to explain the steps of propensity score matching.

Materials and Methods: Propensity score matching has 5 steps. The first step is to estimate the propensity score, which includes selecting the appropriate model and variables. The second step is to select an appropriate matching method based on the estimated propensity scores. The third step is overlap and common support. At this stage, observations that are outside the range of matching scores are removed. The quality of the match is then evaluated, and finally the sensitivity of the estimated effects must be estimated.

Results: In the propensity score matching, the observed basic variables are balanced between the exposure and non-exposure groups. However, if the statistical model used to calculate the propensity score is not correctly selected, an imbalance between the basic characteristics of the two groups can still exist.

Conclusion: Propensity score matching is useful in cases where the confounding variables of study are high. In observational studies, this method can be used as an alternative to randomization in experimental studies, which makes the estimates more accurate by reducing the selection bias and controlling confounding variables.

Key words: Propensity score matching, Confounding variables, Observational studies

Funding: This study did not have any funds.

Conflicts of interest: None declared.

Ethical approval: The Ethics Committee of Kermanshah University of Medical Sciences approved the study (IR.KERMANSHAH.KUMS.REC.1399.1168).

How to cite this article: Tahmasebi Fard Mina, Najafi Farid, Rezaeian Shahab, Moradinazar Mehdi. How Can Causal Relationships Be Measured in Observational Studies? Propensity Score Matching: A Tutorial Article. *J Rafsanjan Univ Med Sci* 2022; 21 (03): 343-66. [Farsi]

1- MSc Student in Epidemiology, Dept. of Epidemiology, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran
2- Prof., Dept.of Epidemiology, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran
3- Associate Prof., Dept. of Epidemiology, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran
4- Assistant Prof., Dept. of Epidemiology, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran Behavioral Diseases Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran, ORCID: 0000-0001-7033-6755
(Corresponding Author) Tel: (083) 38265255, Fax: (083) 38265255, E-mail: M.moradinazar@gmail.com

[Downloaded from journal.rums.ac.ir on 2026-11-21]
[DOR: 20.1001.1.17353165.1401.21.3.6.7]
[DOI: 10.52547/jrums.21.3.343]