

# دانشور

## پزشکی

## چند رویکرد برخورد با مقادیر گمشده متغیرهای کمی و بررسی اثر آنها بر نتایج حاصل از یک کارآزمایی بالینی

نویسندگان: علیرضا اکبرزاده باغبان<sup>۱</sup>، عرفان قاسمی<sup>۲\*</sup>، خسرو خادمی کلانتری<sup>۳</sup>،  
فرید زایری<sup>۴</sup> و سمیه محمودی اقدم<sup>۵</sup>

۱. دانشیار گروه علوم پایه، دانشکده علوم توانبخشی دانشگاه علوم پزشکی شهید بهشتی،  
تهران، ایران

۲. دانشجوی دکتری آمار زیستی گروه آمار زیستی، دانشکده پیراپزشکی دانشگاه علوم پزشکی  
شهید بهشتی، تهران، ایران

۳. استاد گروه فیزیوتراپی، دانشکده علوم توانبخشی دانشگاه علوم پزشکی شهید بهشتی، تهران،  
ایران

۴. دانشیار گروه آمار زیستی، دانشکده پیراپزشکی دانشگاه علوم پزشکی شهید بهشتی، تهران،  
ایران

۵. کارشناس ارشد فیزیوتراپی، دانشکده علوم توانبخشی دانشگاه علوم پزشکی شهید بهشتی،  
تهران، ایران

E-mail: ghasemi.erfan@gmail.com

\* نویسنده مسئول: عرفان قاسمی

### چکیده

مقدمه و هدف: چالشی اساسی که مطالعات طولی را تحت تأثیر قرار می‌دهد، مسئله گمشدگی در داده‌هاست. گمشدگی در داده‌ها ممکن است به از دست رفتن بخشی از اطلاعات منجر شود که این امر، باعث کاهش دقت برآوردگرها و دست‌یافتن به نتایج اریب و نادرست می‌شود لذا ضروری است تا ابتدا داده‌های حاصل از تحقیق‌های طولی از نظر ساختار گمشدگی ارزیابی شده، با در نظر گرفتن این نکته، تحلیلی مناسب با یا بدون جانهی روی داده‌ها انجام شود. هدف از این پژوهش، شناسایی سازوکار گمشدگی موجود در داده‌های حاصل از کارآزمایی بالینی مربوط به اثربخشی دیسترتکشن مفصلی به روش غیرجراحی در بهبود عملکرد بیماران مبتلا به استنئوآرتریت شدید زانو و همچنین بررسی اثر استفاده از رویکردهای مختلف مواجهه با گمشدگی، بر نتایج حاصل از تحلیل این داده‌هاست.

مواد روش‌ها: به منظور به دست آوردن نتایج معتبر در تحلیل داده‌های مرتبط با مطالعات کارآزمایی بالینی طولی که شامل مقادیر گمشده هستند، رویکردهای مختلف برخورد با این‌گونه داده‌ها، از جمله روش داده‌های کامل، روش داده‌های در دسترس، روش LOCF، روش مبتنی بر مدل پیش‌بینی‌کننده و روش‌های جانهی چندگانه از قبیل روش نمره تمایل (propensity score) و روش تطابق میانگین پیش‌بینی‌کننده (Predictive mean matching method) وجود دارند؛ این روش‌ها در داده‌های حاصل از کارآزمایی بالینی مربوط به اثربخشی دیسترتکشن مفصلی به روش غیرجراحی که از دو روش فیزیوتراپی و فیزیوتراپی به علاوه دیسترتکشن مفصلی، در بهبود عملکرد بیماران مبتلا به استنئوآرتریت شدید زانو در سه زمان ابتدای مداخله، انتهای مداخله و یک ماه پس از آن حاصل شده‌است، مورد استفاده قرار گرفته، نتایج حاصل از آنها با یکدیگر مقایسه شدند. از نرم‌افزار SAS ۹.۱ برای برآزش مدل حاشیه‌ای (marginal model) و از نرم‌افزار SOLAS ۴ برای جانهی داده‌ها استفاده شد.

نتایج: با استفاده از رویکرد داده‌های کامل، داده‌های در دسترس و جانهی LOCF، اثر زمان معنی‌دار شد. هر سه روش جانهی چندگانه، نتایجی یکسان داشتند. در هیچ‌یک از رویکردهای مورد استفاده، اثر گروه درمانی، معنادار نشد اگرچه تفاوت‌هایی فاحش در طول فواصل اطمینان و ضرایب مدل وجود داشت.

نتیجه‌گیری: با توجه به ساختار گمشدگی داده‌ها و نتایج به دست آمده از مؤلفه‌های (پارامترهای) خطای معیار، ضرایب مدل و طول بازه اطمینان، استفاده از نتایج رویکردهای جانهی چندگانه مناسب‌تر بوده، روش جانهی LOCF برای داده‌های طولی مداخله‌ای مناسب نیست.

واژگان کلیدی: کارآزمایی بالینی تصادفی‌شده، مطالعه طولی، جانهی چندگانه.

دوماهنامه علمی-پژوهشی  
دانشگاه شاهد  
سال بیست و دوم- شماره ۱۱۶  
اردیبهشت ۱۳۹۴

دریافت: ۱۳۹۳/۱۲/۱۹  
آخرین اصلاح‌ها: ۱۳۹۴/۰۱/۲۵  
پذیرش: ۱۳۹۴/۰۱/۲۹

## مقدمه

چندانی در تحلیل داده‌ها به وجود نخواهد آمد، ولی در صورتی که این دو به هم مرتبط باشند، امکان اریبی جدی وجود خواهد داشت؛ زیرا داده‌های در دسترس با داده‌های کامل، متفاوت خواهند بود (۳،۷).

لیتل و روبین، دسته‌بندی منسجمی برای سازوکار گمشدگی ارائه دادند و سه سازوکار گمشدگی به طور تصادفی (MCAR: Missing Completely At Random)، تصادفی (MAR: Missing At Random) و غیرتصادفی (MNAR: Missing Not At Random) را برای داده‌های گمشده تعریف کردند و روش‌هایی مانند روش مبتنی بر داده کامل، روش تجدید وزن‌دهی و روش جانپی را برای شیوه‌های اجرای تحلیل آماری با وجود بی‌پاسخی‌های پرسش فهرست کردند (۸).

به طور کلی سه نوع سازوکار (مکانیسم) گمشدگی برای داده‌های گمشده می‌تواند در نظر گرفت: ۱. زمانی که گمشدگی داده‌ها، مستقل از مقادیر مشاهده‌شده و مقادیر مشاهده‌نشده متغیر پاسخ باشد، سازوکار گمشدگی را به طور تصادفی (MCAR) می‌گوییم؛ ۲. گمشدگی به صورت تصادفی (MAR) است، اگر گمشدگی داده‌ها به شرط پاسخ‌های مشاهده‌شده، مستقل از مقادیر مشاهده‌نشده باشد و ۳. گمشدگی غیرتصادفی (MNAR)، زمانی رخ می‌دهد که احتمال گمشدگی به مقادیر مشاهده‌شده و نیز به مقادیر مشاهده‌نشده متغیر پاسخ، وابسته باشد. سازوکارهای گمشدگی MCAR و MAR را گمشدگی‌های قابل چشم‌پوشی می‌گویند، در حالی که سازوکار MNAR، غیرقابل چشم‌پوشی است (۲).

علاوه بر دسته‌بندی بالا، مقادیر گمشده در داده‌های طولی می‌توانند با الگوهای مختلف رخ دهند؛ بر این اساس، گمشدگی‌ها به دو دسته تقسیم می‌شوند: ۱. گمشدگی یک‌نوا یا ریزش‌شده (Monotone missingness or dropout): در این نوع گمشدگی، شخص از یک زمان به بعد به طور کامل از مطالعه خارج شده است و تمامی مقادیر پس از آن، گمشده محسوب می‌شوند و ۲. گمشدگی غیریک‌نوا یا متناوب (Non-monotone or

در تحقیق‌های علوم پزشکی، اغلب برای بررسی اثربخشی یک روش درمانی، از مطالعات طولی استفاده می‌شود که در آن، هر شخص در طول زمان، تحت اندازه‌گیری‌های مکرر قرار می‌گیرد (۱). چالشی اساسی که مطالعات طولی را تحت تأثیر قرار می‌دهد، مسئله گمشدگی در داده‌هاست؛ یعنی این امکان وجود دارد که پاسخ هریک از افراد در هریک از زمان‌های پیگیری در دسترس نباشد (۲). زمانی که در داده‌ها گمشدگی وجود دارد، قسمتی از اطلاعات از دست می‌روند و در نتیجه، دقت برآوردگرها کاهش می‌یابد؛ این کاهش دقت می‌تواند به طور مستقیم به داده‌های گمشده یا به روش‌هایی که برای تحلیل این‌گونه داده‌ها به کار می‌روند، مربوط باشد اما مشکلی اساسی‌تر که داده‌های گمشده می‌توانند به وجود آورند، اریبی در برآوردها و استنباط‌هاست که می‌تواند به نتیجه‌گیری‌هایی نادرست و غیرواقعی منجر شود (۵-۳).

به طور کلی می‌توان گفت: زمانی که در داده‌ها گمشدگی وجود دارد، سه مشکل عمده پدید می‌آیند: اول اینکه داده‌ها را از حالت متعادل (وقتی، همه افراد، تعداد اندازه‌گیری‌هایی یکسان دارند) خارج می‌سازد بنابراین، روش‌هایی را که برای داده‌های کامل به کار می‌روند، نمی‌توان استفاده کرد؛ دوم، کاهش حجم نمونه که باعث کاهش دقت و کارایی برآوردگرها می‌شود و در نهایت، از دست رفتن اطلاعاتی که ممکن است باعث اریبی و استنباط‌های نادرست شود (۶).

شرایطی که به گمشدگی داده‌ها منجر شده‌اند، می‌توانند باعث ایجاد اریبی در برآوردها و رسیدن به نتایج نادرست و غیرواقعی شوند؛ بنابراین به فرضیاتی درباره دلایل گمشدگی نیاز داریم که اغلب «مکانیزم داده‌های گمشده» نامیده می‌شوند؛ این فرضیات در واقع در پاسخ به این پرسش که «چرا گمشدگی رخ داده؟ یا به طور خاص‌تر، آیا مقادیر گمشده با پرسش‌های کاربردی تحقیق ارتباطی دارند یا نه؟» شکل می‌گیرند. اگر دلایل گمشدگی، با پیامد مورد نظر ارتباطی نداشته باشند، مشکل

انتخاب شدند. با خطای نوع اول آزمون  $0/05$  و خطای نوع دوم آزمون  $0/2$  و برآورد اندازه اثر از مطالعات مشابه پیشین، تعداد ۲۰ بیمار در هر یک از دو گروه درمانی قرارگرفتند (درمجموع ۴۰ بیمار)؛ گروه اول، درمان فیزیوتراپی و گروه دوم، درمان فیزیوتراپی همراه با دیسترکشن مفصلی دریافت کردند. بیماران ده جلسه تحت درمان قرارگرفتند و برای ارزیابی توانایی عملکردی بیماران، از همه بیماران، در ابتدای جلسه اول، پس از اتمام جلسه دهم و یک ماه پس از اتمام درمان، تست 6MW گرفته شد.

برای تحلیل صحیح داده‌های حاصل از این تحقیق، به دلیل وجود گمشدگی در داده‌ها از روش‌های مناسب برای تحلیل داده‌های گمشده استفاده شد؛ به همین دلیل ابتدا سازوکار گمشدگی در گروه‌های مختلف درمانی بررسی شد و سپس مدل‌سازی مناسب بر مبنای جانمایی مقادیر گمشده صورت گرفت. رویکردهایی مختلف که در این تحقیق برای برخورد با گمشدگی در داده‌ها مورد استفاده قرارگرفتند، عبارت‌اند از:

الف) روش مورد کامل (Complete-case method): در این روش، تمامی واحدهایی که دارای مقادیر گمشده در هر یک از متغیرهای حاضر در تحلیل هستند، کنارگذاشته می‌شوند؛ این روش، یکی از ساده‌ترین و شاید متداول‌ترین روش‌ها برای حل مشکل داده‌های ناقص است؛ هرچند که این رویکرد به‌طور گسترده مورد استفاده قرار می‌گیرد و کاربرد روش‌های آماری مستلزم داده‌های کامل را میسر می‌سازد، اما دلایلی وجود دارند مبنی بر اینکه این رویکرد در حالت کلی، شیوه‌ای کارآمد و مطلوب تلقی نمی‌شود؛ دلیل اول، این است که با حذف همه واحدهای دارای داده‌های گمشده، اندازه نمونه موجود، به‌صورتی قابل ملاحظه کوچک شده، دقت مطالعه کم می‌شود؛ دلیل دوم نیز این است که اگر افرادی که از تحلیل حذف می‌شوند با آنهایی که باقی می‌مانند، تفاوت زیادی در متغیرهای مورد مطالعه داشته باشند، ممکن است برآوردهای حاصل به شدت اریب شوند (۹،۱۰)؛ در این روش، تنها زمانی که

(intermittent missingness): هرگاه گمشدگی از نوع ریزش شده نباشد، متناوب خواهد بود؛ به عبارتی، شخص به‌طور غیریک‌نوا در بعضی از زمان‌های اندازه‌گیری حضور داشته، در بعضی از زمان‌ها حضور نداشته است و ممکن است میان دو مقدار مشاهده شده، گمشدگی داشته باشیم (۲).

با توجه به این مسئله که از دست دادن اطلاعات مربوط به افراد حاضر در یک کارآزمایی بالینی طولی، بسیار شایع و قابل توجه است و نادیده گرفتن مسئله گمشدگی، در تحلیل داده‌ها باعث نتیجه‌گیری‌های اریب و نادرستی خواهد شد، در این تحقیق، تأثیر رویکردهای مختلف برخورد با داده‌های گمشده را در داده‌های حاصل از کارآزمایی بالینی مربوط به اثربخشی دیسترکشن مفصلی به روش غیرجراحی در بهبود عملکرد بیماران مبتلا به استئوآرتریت شدید زانو، مورد بررسی قرار دادیم.

## روش کار

در این تحقیق، بیماران خانم با دامنه سنی ۴۵ تا ۷۵ سال و مبتلا به استئوآرتریت اولیه درجه ۳ و ۴ کلوگرن و لارنس (طبق تشخیص پزشک متخصص ارتوپدی) و نداشتن مشکلاتی مانند «بدخیمی در زانو، عفونت در اطراف زانو، وجود بیماری‌های متابولیک و اندوکراین، تزریق کورتون طی سی روز گذشته، آغاز مصرف مسکن کمتر از یک ماه، سابقه شکستگی زانو، جراحی مفصل زانو در شش ماه گذشته، درد کمتر از یک سال، هایپرموبیلیتی زانو و بی‌ثباتی لیگامانی، تورم شدید اطراف زانو، استئوپروز و مشکلات روحی روانی»، در مطالعه شرکت کردند. در صورتی که افراد، قادر نبودند، تست ۶ دقیقه راه رفتن (6MW) را انجام دهند، بیش از دو جلسه پیاپی غیبت داشتند یا به هر دلیلی نمی‌توانستند درمان را تکمیل کنند از مطالعه خارج شدند. نمونه‌ها به‌صورت غیرتصادفی از بیماران مراجعه‌کننده به کلینیک فیزیوتراپی دانشکده توانبخشی دانشگاه علوم پزشکی شهید بهشتی و کلینیک فیزیوتراپی درمانگاه ۱۷ شهریور

گمشدگی براساس سازوکار MCAR صورت گرفته باشد، برآوردهای نارایب از میانگین پاسخ در طول زمان خواهیم داشت؛ در این حالت، نمونه‌های موجود در واقع، زیر نمونه‌ای تصادفی از جمعیت اصلی ما هستند؛ با وجود این، حتی در صورتی که سازوکار گمشدگی MCAR باشد، این نوع تحلیل، چندان مناسب نیست زیرا با این روش، تعداد زیادی از افراد را از دست داده، حجم نمونه، بسیار کاهش می‌یابد. به‌طور خلاصه می‌توان گفت که تحلیل برمبنای داده‌های کامل، بسیار ناکاراست و به توانی پایین منجر خواهد شد (۹).

ب) روش مورد در دسترس (Available-case method): رویکردی دیگر که برای داده‌های گمشده وجود دارد، رویکرد داده‌های در دسترس است؛ در این رویکرد از تمامی داده‌های موجود استفاده می‌شود؛ این رویکرد از روش داده‌های کامل، کارتر است زیرا در این رویکرد از اطلاعات جزئی که بعضی افراد در قسمتی از مطالعه ارائه داده‌اند، نیز استفاده می‌شود. روش‌های مبتنی بر داده‌های موجود در صورتی که گمشدگی‌ها MCAR نباشند، نتایجی اریب را برای برآورد متوسط پاسخ در طول زمان نتیجه می‌دهند.

ج) روش‌های مبتنی بر جانهای: رویکردی دیگر که در مواجهه با گمشدگی در داده‌ها به‌کار می‌رود، استفاده از یکی از روش‌های جانهای برای پرکردن جاهای خالی در داده‌هاست؛ به این صورت که جاهای خالی حاصل از گمشدگی را با مقادیر مناسب جانهای شده پرمی‌کنیم؛ یکی از مزایای این رویکرد، این است که با جای‌گذاری مقادیر گمشده، مجموعه داده‌ها به‌صورت کامل درآمده، می‌توان از روش‌های مناسب برای داده‌های کامل استفاده کرد (۹).

جانهای با روش LOCF (Last observation carried forward): در این روش، مقدار گمشده با آخرین مقداری که پیش از آن فرد مورد نظر مشاهده شده است، جای‌گذاری می‌شود؛ این روش برای گمشدگی‌های یک‌نوا به‌طور فراوان مورد استفاده قرار می‌گیرد؛ با این حال، معایبی نیز دارد؛ در واقع، فرض اینکه مقادیر

گمشده پس از آخرین مقدار مشاهده شده، ثابت باقی می‌مانند، کمی غیرمعمول به‌نظر می‌رسد و شاید بتوان گفت: تنها حالتی که این فرضیه برقرار می‌شود، هنگامی است که گمشدن به دلیل بهبودی یا درمان کامل صورت گرفته باشد (۹، ۲)؛ تغییرهایی که می‌توان در این روش اعمال کرد، عبارت‌اند از: جایگزین کردن مقدار پایه (Baseline) یا بدترین مقدار هر فرد به جای مقادیر گمشده. جایگزین کردن بدترین مقدار در کارآزمایی‌هایی که یک درمان را با دارونما مقایسه می‌کنند، روش محتاطانه است؛ این روش‌ها بسیار مشکوک بوده، نمی‌توانند برآوردهای نارایب از مؤلفه‌ها (پارامترها) را به‌دست دهند؛ با این حال در این روش، برآوردهای واریانس، مقادیری کوچک‌تر را نسبت به روش‌هایی که تنها از داده‌های کامل یا از داده‌های موجود استفاده می‌کنند، به‌دست می‌دهد (۲).

جانهای چندگانه: روش‌هایی که در بالا مورد بحث قرار گرفتند، به جای هر مقدار گمشده، یک مقدار جانهای شده تک را قرار می‌دهند و تحلیل حاصل، با مقدار جانهای شده به همان طریقی رفتاری کند که با مقادیری که به‌واقع اندازه‌گیری شده‌اند، رفتاری شود و در تحلیل، عدم حتمیت درباره‌ی این مقدار گمشده در نظر گرفته نمی‌شود؛ تا حدودی به دلیل همین مسئله، روبین (۱۱) و سایر پژوهشگران، طی دهه‌های گذشته، فنی به نام جانهای چندگانه را ابداع کرده‌اند؛ این روش، جای هر مقدار گمشده را با دو یا چند مقدار جانهای شده پرمی‌کند و هر مجموعه از داده‌های به‌دست آمده را با استفاده از روش‌های مربوط به داده‌های کامل تحلیل می‌کند؛ جای‌گذاری با تعدادی میان پنج تا ده مورد کافی به‌نظر می‌رسد؛ با ترکیب نتایج این تحلیل‌ها، استنباط‌هایی به‌دست می‌آیند که می‌توانند عدم حتمیت درباره مقادیر گمشده را در نظر بگیرند (۲، ۱۱).

در این روش به تعداد جانهای صورت گرفته، مجموعه داده‌های کامل تولید می‌شود و برمبنای هریک از این مجموعه داده‌ها، برآوردی از مؤلفه‌ها س مدل و واریانس آنها حاصل می‌شود سپس در انتها این مقادیر با

(۱۴)، شش روش یادشده در برخورد با گمشدگی اعمال شده، مؤلفه‌های مدل حاشیه‌ای به روش GEE برآورد و با هم مقایسه شدند. از نرم‌افزار SAS ۹.۱ برای برازش مدل حاشیه‌ای و از نرم‌افزار ۴ SOLAS برای جانهی داده‌ها استفاده شد.

#### یافته‌ها

جدول ۱، الگوی گمشدگی در داده‌ها را نشان می‌دهد و از این جدول، مشخص است که الگوی گمشدگی داده‌ها یک‌نواست.

جدول ۱. الگوی گمشدگی در کل داده‌های طرح

تعداد	جلسه پیگیری	جلسه آخر	جلسه اول
۲۹	مشاهده	مشاهده	مشاهده
۱۰	گمشده	مشاهده	مشاهده
۱	گمشده	گمشده	مشاهده
۴۰			

در جدول شماره ۲، نتایج حاصل از برازش مدل حاشیه‌ای، روی داده‌های کامل (با حذف موارد دارای گمشدگی)، داده‌های در دسترس، داده‌های حاصل از جانهی LOCF، جانهی چندگانه مدل پیش‌بینی‌کننده، نمره تمایل و تطابق میانگین پیش‌بینی‌کننده را مشاهده می‌کنید. با توجه به جدول شماره ۲ ملاحظه می‌شود که معنی‌داری اثر زمان در رویکردهای جانهی چندگانه با سایر رویکردها تفاوت دارد، اگرچه در هیچ‌یک از روش‌های جانهی اثر گروه درمانی معنی‌دار نیست.

هم ترکیب می‌شوند تا برآوردی کلی حاصل شود و واریانس کلی که به دست می‌آید نیز، این عدم اطمینان حاصل از جای‌گذاری مقادیر داده‌های گمشده را دربردارد. از روش‌های رایج جانهی چندگانه می‌توان به روش مبتنی بر مدل پیش‌بینی‌کننده ( Predictivemodel based method)، روش نمره تمایل ( Propensity score method) و روش تطابق میانگین پیش‌بینی‌کننده ( Predictive mean matching method)، اشاره کرد.

مدل‌هایی مختلف برای تحلیل داده‌های طولی پیشنهاد شده‌اند که یکی از پرکاربردترین این مدل‌ها، مدل حاشیه‌ای است. یک مدل حاشیه‌ای را می‌توان در سه مشخصه زیر بیان کرد:

امید ریاضی پاسخ با یک تابع ربط به متغیرهای کمکی وابسته است:

$$g(\mu_{ij}) = x'_{ij}\beta$$

واریانس پاسخ‌ها به میانگین حاشیه‌ای وابسته است:

$$\text{var}(Y_{ij}) = \phi v(\mu_{ij})$$

ارتباط میان مشاهدات تکراری درون فردی به میانگین‌های حاشیه‌ای و مؤلفه‌های اضافی  $\alpha$ ، وابسته است.

این مدل، میانگین پاسخ و ارتباط درون فردی را به‌طور جداگانه مدل‌بندی می‌کند و مؤلفه‌های آن را می‌توان به روش معادلات برآوردی تعمیم‌یافته (GEE) برآورد کرد (۱۲، ۱۳).

لذا در این مقاله روی داده‌های مربوط به کارآزمایی بالینی است مقایسه اثربخشی دو روش درمانی فیزیوتراپی و دیسترکشن مفصلی به همراه فیزیوتراپی در بهبود عملکرد بیماران مبتلا به استئوآرتریت شدید زانو

جدول ۲. نتایج حاصل از برازش مدل حاشیه‌ای در رویکردهای مختلف برخورد با گمشدگی

p	طول فاصله اطمینان	فاصله اطمینان		خطای استاندارد	ضریب رگرسیونی	متغیر	رویکرد برخورد با گمشدگی
۰/۹۲۳۴	۱۰۶/۷۷	-۵۰/۷۶۲۴	۵۶/۰۰۱۹	۲۷/۲۳۶۳	۲/۶۱۹۷	گروه درمانی*	داده‌های کامل
۰/۰۲۴۲**	۱۷/۷۳	۱/۳۲۶۴	۱۹/۰۵۲۹	۴/۵۲۲۱	۱۰/۱۸۹۷	زمان	
۰/۹۳۹۸	۹۴/۰۱	-۴۵/۱۹۲۳	۴۸/۸۱۵۴	۲۳/۹۸۲۰	۱/۸۱۱۵	گروه درمانی*	داده‌های در دسترس
۰/۰۰۷۲**	۱۶/۵۰	۳/۰۶۵۸	۱۹/۵۶۹۰	۴/۲۱۰۱	۱۱/۳۱۷۴	زمان	
۰/۹۰۷۵	۹۴/۱۱	-۴۴/۲۶۴۷	۴۹/۸۴۲۷	۲۴/۰۰۷۴	۲/۷۸۹۰	گروه درمانی*	جانهای با روش LOCF
۰/۰۰۴۸**	۱۳/۷۵	۳/۰۱۳۹	۱۶/۷۶۱۱	۳/۵۰۷۰	۹/۸۸۷۵	زمان	
۰/۹۶۵۱	۵۶/۸۰	-۲۹/۰۳۳۸	۲۷/۷۶۴۳	۱۴/۴۸۹۱	-۰/۶۳۴۸	گروه درمانی*	جانهای چندگانه با روش مدل پیش‌بینی‌کننده
۰/۲۸۳۶	۳۴/۹۶	-۷/۹۲۲۴	۲۷/۰۳۷۴	۸/۹۰۷۸	۹/۵۵۷۵	زمان	
۰/۸۵۲۸	۵۶/۷۶	-۲۵/۶۹۷۴	۳۱/۰۶۶۹	۱۴/۴۶۲۹	۲/۶۸۴۸	گروه درمانی*	جانهای چندگانه با روش نمره تمایل
۰/۱۷۹۷	۳۴/۱۴	-۵/۳۹۳۷	۲۸/۷۴۳۷	۸/۶۹۴۴	۱۱/۶۷۵۰	زمان	
۰/۸۸۷۶	۵۷/۱۶	-۲۶/۵۱۹۴	۳۰/۶۳۸۶	۱۴/۵۶۵۳	۲/۰۵۹۶	گروه درمانی*	جانهای چندگانه با روش تطابق میانگین پیش‌بینی‌کننده
۰/۱۶۰۶	۳۳/۵۱	-۴/۷۶۱۶	۲۸/۷۵۱۶	۸/۵۴۸۰	۱۱/۹۹۵۰	زمان	

\* گروه کنترل به عنوان مرجع در نظر گرفته شده است و ضرایب با تعدیل روی متغیرهای سن و BMI محاسبه شده‌اند. \*\* در سطح ۰/۰۵ معنی‌دار است.

## بحث

پیش‌بینی‌کننده استفاده شد. جدول ۲، برازش مدل حاشیه‌ای را بر داده‌های حاصل از رویکردهای یاد شده نشان می‌دهد؛ مشاهده می‌شود که رویکردهای مختلف، نتایج متفاوت را از لحاظ معنی‌داری اثر متغیرها نشان می‌دهند.

در جدول ۲ در برازش مدل حاشیه‌ای به داده‌های کامل، تنها اثر متغیر زمان، معنی‌دار شده است. در استفاده از داده‌های در دسترس و همچنین جانهای LOCF نیز اثر زمان، معنی‌دار شده است با این تفاوت که در رویکردهای اخیر، علاوه بر زمان، متغیر سن افراد نیز معنی‌دار شده است (نتایج نشان‌دهنده نشده‌اند).

سه روش جانهای چندگانه، نتایج متفاوت نسبت به سه روش بالا داشته‌اند. با توجه به نتایج جدول ۲ ملاحظه می‌شود که اثر زمان و گروه درمانی معنی‌دار

در این پژوهش، مدل حاشیه‌ای برای بررسی اثر دیسترنشن مفصلی در بهبود توانایی عملکردی بیماران مبتلا به استئوآرتریت شدید زانو به کار برده شد. وجود گمشدگی در داده‌های حاصل از تحقیق‌های پزشکی از جمله مطالعه حاضر، باعث ایجاد پیچیدگی‌هایی در تحلیل این گونه داده‌ها می‌شود (۱۵). برای فائق آمدن بر مشکلات ناشی از وجود گمشدگی، راه‌حلهایی مختلف پیشنهاد شده‌اند (۱۱، ۱۶).

پژوهش حاضر با مقایسه نتایج حاصل از رویکردهای مختلف به دنبال یافتن بهترین رویکرد در مواجهه با گمشدگی است؛ در این پژوهش از رویکردهای داده‌های کامل، داده‌های در دسترس، داده‌ها با جانهای LOCF و جانهای چندگانه به روش‌های مبتنی بر مدل پیش‌بینی‌کننده، نمره تمایل و روش تطابق میانگین

۲، خطای معیار برآوردگر ضریب متغیر گروه درمانی، به طور تقریبی، برابر ۲۷ و طول بازه اطمینان، به طور تقریبی، برابر ۱۰۶ واحد است که هر دو، مقادیری بزرگ و نامعقول‌اند؛ درحالی‌که این دو مقدار مربوط به جانهای چندگانه، به ترتیب، به طور تقریبی، برابر ۱۴ و ۵۶ هستند؛ به عبارتی، هر دو مقدار نصف‌شده‌اند که معرف دقت بالاتر این روش‌های جانهای است.

نتیجه‌گیری: نتایج ارائه‌شده در مطالعه حاضر، نشان‌دهنده برتری روش‌های جانهای چندگانه بر جانهای تکی به روش LOCF هستند؛ درحالی‌که این روش به‌نوبه‌خود، بهتر از روش بدون جانهای کامل و داده‌های در دسترس است؛ از طرف دیگر، روش LOCF شاید برای مطالعات طولی، مشاهده‌ای قابل قبول‌تر باشد که در آنها شرایط مطالعه دستکاری نمی‌شوند و جای‌گذاری مقداری نامشخص با مقدار مشاهده پیشین، غیرمعقول نیست؛ اما در مطالعاتی مانند مطالعه حاضر (که [در آنها] با داده‌های طولی مداخله‌ای سروکار داریم و بیماران در جلسات درمانی مختلف، تحت درمان قرار گرفته‌اند و در نتیجه، در مقادیر متغیر وابسته آنها تغییرهایی ایجاد می‌شود)، انجام این کار چندان درست به نظر نمی‌رسد.

### سپاس و قدردانی

این طرح با شماره 1391-1-93-10088 و با کمک مالی دانشکده توانبخشی دانشگاه علوم پزشکی و خدمات بهداشتی- درمانی شهید بهشتی به‌انجام رسیده‌است که بدینوسیله از مسئولان محترم دانشکده تقدیر و تشکر می‌شود.

نشده‌است. طول فواصل اطمینان نیز در روش‌های جانهای با سایر رویکردها، تفاوتی قابل‌ملاحظه دارد به طوری‌که طول فاصله اطمینان برای متغیر گروه درمانی کاهش‌داشته ولی برای متغیر زمان، افزایش پیدا کرده‌است.

به‌طور کلی، رویکرد استفاده از داده‌های کامل، روشی مناسب برای حل مشکل گمشدگی در داده‌ها نیست زیرا هم باعث کاهش شدید حجم نمونه و در نتیجه، کارایی برآوردها می‌شود و هم ممکن است باعث اریبی در داده‌ها شود (۱۰، ۱۷). استفاده از روش داده‌های در دسترس، نیز زمانی قابل‌توجیه است که سازوکار گمشدگی به صورت به‌طور کامل تصادفی (MCAR) باشد؛ در غیر این صورت، نتایج حاصل از این رویکرد، ممکن است دارای اریبی باشند؛ همچنین برای استفاده از روش LOCF، باید توجیه منطقی داشت که بتوان از آخرین داده مشاهده‌شده به جای داده‌های گمشده استفاده کرد که اغلب، توجیهی مناسب برای آن وجود ندارد؛ در این میان، استفاده از روش‌های جانهای چندگانه، روشی معقول‌تر به نظر می‌رسد؛ این روش‌ها را می‌توان برای گمشدگی‌های تصادفی (MAR) و به‌طور کامل تصادفی (MCAR) به‌کاربرد.

در پژوهش حاضر مشاهده می‌شود که نتایج حاصل از جانهای چندگانه از نتایج به‌دست‌آمده از رویکردهای دیگر، متفاوت است ولی هر سه رویکرد جانهای چندگانه، نتایجی مشابه، نسبت به یکدیگر دارند.

با توجه به اینکه گمشدگی موجود در داده‌ها به صورت یک‌نوا (ریزش) است، درست بودن فرض به‌طور کامل تصادفی سازوکار گمشدگی، چندان منطقی به نظر نمی‌رسد و فرض گمشدگی تصادفی (MAR) برای داده‌ها مناسب‌تر است؛ بنابراین، استفاده از رویکردهای جانهای چندگانه برای این داده‌ها مناسب‌تر و نتایج حاصل از آنها قابل‌استنادتر است.

این نتیجه‌گیری از روی مقادیر خطای معیار برآورد مؤلفه گروه درمانی و طول فاصله اطمینان مربوط در جدول ۲ نیز قابل‌حصول است. براساس یافته‌های جدول

### منابع

1. Yang X, Li J, Shoptaw S. Imputation-based strategies for clinical trial longitudinal data with nonignorable missing values. *Statistical in Medicine* 2008; 27(15): 2826-49.
2. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. Wiley-IEEE: 2004.
3. Fitzmaurice GM, Davidian M, Verbeke G, Molenberghs G. *Longitudinal Data Analysis*. Chapman & Hall/CRC: 2009.
4. Twisk JWR. *Applied longitudinal data analysis for epidemiology: a practical guide*. Cambridge University Press: 2003.
5. Hedeker D, Gibbons RD. *Longitudinal data analysis*. New York: Wiley; 2006.
6. Horton NJ, Kleinman KP. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*. 2007; 61(1): 79-90.
7. Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford University Press: 2nd Edition, 2002.
8. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). New York: Wiley; 2002.
9. Fitzmaurice GM. Methods for handling dropouts in longitudinal clinical trials. *Statistica Neerlandica*. 2003; 57(1): 75-99.
10. Molenberghs G, Beunckens C, Sotito C, Kenward MG. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B*. 2008; 70(2): 371-88.
11. Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley; 1987.
12. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*. 1991; 59(1): 25-35.
13. Carrière I, Bouyer J. Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons. *BMC Medical Research Methodology*. 2002; 2(1): 2-15.
14. Aghdam MS, Kalantari KK, Baghban AA, Rezaee M, Rahimi A and Naeimee S. Effect of joint traction on functional improvement and quality of life in patients with severe knee osteoarthritis. *The Scientific Journal Of Rehabilitation Medicine*. 2012; 1(2): 1-7.
15. Myers WR. Handling missing data in clinical trials: An overview. *Drug Information Journal*. 2000; 34(2): 525-33.
16. Baghban, A. A., Ghasemi, E., Zayeri, F., Asgary, S., & Namdari, M. Analysis of longitudinal binary outcomes in clinical trials with low percentage of missing values. *Journal of Paramedical Sciences*. 2013; 4: 33-41.
17. Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2007; 7(2): 147-77.



## Different approaches to dealing with missing values in quantitative variables and survey of their effects on the results of a clinical trial

Alireza Akbarzade Baghban<sup>1</sup>, Erfan Ghasemi<sup>2\*</sup>, Khosro Khademi Kalantari<sup>3</sup>, Farid Zayeri<sup>4</sup>, Somaye Mahmoudi Aghdam<sup>5</sup>

1. Associate Professor - Dept. Basic Sciences, Faculty of Rehabilitation Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
2. Dept. Of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
3. Professor - Dept. Physical Therapy, Faculty of Rehabilitation Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
4. Associate Professor - Dept. Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
5. Faculty of Rehabilitation Sciences, Shahid beheshti University of Medical Sciences, Tehran, Iran.

E-mail: ghasemi.erfan@gmail.com

### Abstract

**Background and Objectives:** A major challenge that affects the longitudinal studies is the problem of missing data. Missing in the data may result in the loss of part of the information which reduces the accuracy of the estimator and obtain the results will be biased and inaccurate. Therefore, it is necessary to evaluate the missing data mechanism from a longitudinal research and to consider this fact, an analysis should be performed on the data with or without imputation. The purpose of this study was to identify the Missing data mechanism from a clinical trial Related to the efficacy of non-surgical method of joint distraction in improving performance in patients with severe knee osteoarthritis, and the effect of different approaches to dealing with the missing on the results the analysis of this data set.

**Methods:** In order to obtain reliable results in the analysis of the data associated with the longitudinal clinical studies involving missing values, there are different approaches for dealing with such data, including complete case data approach, available data, LOCF and multiple imputation methods such as predictive model based method, propensity score method (propensity score) and predictive mean matching method (Predictive mean matching method). This approaches were used on the data obtained from clinical trial of the effectiveness of non-surgical joint distraction by two methods physiotherapy and joint distraction in addition to physiotherapy in improving patients performance with severe knee osteoarthritis were measured in three times, beginning of treatment, end of treatment and one month later and the results were compared. SAS 9.1 software was applied for fitting the marginal model and SOLAS software version 4 for missing data imputation.

**Results:** By using the complete data, available data and LOCF imputation approaches time effect was significant. All the three multiple imputation methods had the same results. In none of the approaches used, the effect of treatment group was not significant, although there were significant differences in the length of confidence intervals and coefficients of the model.

**Conclusion:** According to the missing data mechanism and results obtained from coefficients of the model, standard errors and confidence interval length, to use of multiple imputation approach in the data is more appropriate and LOCF imputation method is not appropriate for interventional longitudinal data.

**Key words:** Randomized clinical trials, Longitudinal study, Multiple imputation