

## کاربرد جای گذاری چندگانه در تحقیقات پزشکی و اپیدمیولوژی

مجید میرمحمدخانی: دستیار تخصصی، گروه اپیدمیولوژی و آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران  
کوروش هلاکویی نائینی: استاد، گروه اپیدمیولوژی و آمارزیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران - نویسنده رابط:  
holakoin@sina.tums.ac.ir  
تاریخ دریافت: ۱۳۸۹/۱۱/۹ تاریخ پذیرش: ۱۳۹۰/۵/۲۵

### چکیده

گم شدن داده ها در اکثر مطالعات اپیدمیولوژیک امری اجتناب ناپذیر است و با مکانیسم ها و دلایل مختلفی رخ می دهد. به دلیل ماهیت فراگیر آن، بسیاری از روش شناسان از مشکل داده های گم شده به عنوان یکی از مهمترین معضلات طراحی و تحلیلی در پژوهش یاد می کنند. اگرچه در اکثر مطالعات، تحلیل بر اساس موارد کامل انجام می شود، اما این رویکرد ممکن است منجر به برآوردها و استنتاج هایی گردد که به دلیل سوگرایی ناشی از داده های گم شده دور از واقعیت باشد. این در حالی است که با ترمیم داده های گم شده ممکن است بتوان به نتایج دقیق تری دست یافت. روش های معمولی و ساده مانند جای گذاری منفرد، دارای ضعف هایی هستند که استفاده از آنها را در عمل بسیار محدود می کند. اما نشان داده شده که روش های جدید جای گذاری در کنار پیچیدگی نسبی خود، اگر درست و بجا استفاده شوند، می توانند در بسیاری از موارد باعث تقلیل چشمگیر سوگیری گردند. در مقاله حاضر کاربرد جای گذاری چندگانه به عنوان یک شیوه مدرن در برخورد با مشکل فقدان داده ها در تحقیقات پزشکی و سلامت به طور خلاصه بررسی شده است. واژگان کلیدی: جای گذاری چندگانه، اپیدمیولوژی، داده های گمشده

### مقدمه

2004). گرچه این رویکرد که تحلیل موارد کامل (Complete-Case Analysis) نام دارد، در برخی از شرایط به نتایج قابل قبولی منتهی می گردد، اما پژوهشگران میدانند که در بسیاری از موقعیت ها با ترمیم داده ها به نتایج بهتری دست خواهند یافت. برای انجام این ترمیم لازم است با استفاده از یک روش قابل قبول، بجای هر داده گم شده مقداری مناسب برآورد و جایگزین گردد (Ghosh and Pahwa 2008). تا به حال روش های مختلفی برای جای گذاری داده های گم شده معرفی و به کار رفته است. روش های ساده و معمولی جای گذاری، دارای ضعف هایی هستند که استفاده از آنها را سوال برانگیز و در عمل بسیار محدود می کند. اما نشان داده شده که روش های جدید جای گذاری در کنار پیچیدگی نسبی خود، اگر درست و بجا استفاده شوند، می توانند در بسیاری از موارد باعث تقلیل چشمگیر خطا و افزایش دقت مطالعه گردند (Horton and Kleinman 2007). در این مقاله سعی

وجود داده های گم شده (Missing data) در بیشتر مطالعات امری اجتناب ناپذیر است و به دلیل ماهیت فراگیر و تاثیر گذاری که دارد، از آن به عنوان یکی از مهمترین معضلات طراحی و آماری در پژوهش یاد می کنند (Little and Rubin 2002). داده های مطالعات اپیدمیولوژیک نیز بدلائل مختلف از این مشکل مصون نمی باشند (Barnard and Meng 1999). هدف اصلی و بدیهی انجام یک تحقیق و تحلیل های آماری آن، دستیابی به برآورد و استنتاجی درست درخصوص جمعیت مورد نظر می باشد. اما وجود نقص در داده ها در صورت عدم چاره اندیشی مناسب، می تواند لطمات جدی به نتایج پژوهش وارد نماید. با پیروی از یک عادت دیرینه محقق می تواند از موارد ناقص چشم پوشی کند و تنها بر اساس داده های کامل اقدام به تحلیل و نتیجه گیری نماید (Joseph et al.

بدون این که اقدامی در جهت ترمیم داده ها صورت گیرد، امکان پذیر می باشد. گرچه این فرض در بسیاری از موقعیت ها پذیرفتنی نیست، اما حتی در این شرایط نیز داده های گم شده می توانند تاثیر نامطلوبی بر نتایج تحقیق بگذارند. چرا که به دلیل کمتر شدن حجم نمونه، خطای معیار (Standard Error) برآورد های مورد نظر نیز افزایش می یابد و در نتیجه دقت (Precision) مطالعه کاهش پیدا می کند. اگر گم شدن داده ها به طور تصادفی (MAR) یا غیرتصادفی (MNAR) روی دهد، علاوه بر کم شدن دقت مطالعه خطرو وجود اریبی نیز در انجام برآورد های مورد نظر، نتایج مطالعه را تهدید خواهد کرد، به خصوص در شرایطی که گم شدن داده ها با مکانیسم غیر تصادفی (MNAR) اتفاق می افتد. بنابراین روشهای جای گذاری چنانچه درست و بجا مورد استفاده قرار گیرند، می توانند تحت هر یک از شرایط پیش گفت به کم کردن مشکلات تحلیلی ناشی از داده های گم شده کمک نمایند (He 2010; Little and Rubin 2002).

جای گذاری داده ها: ایده اصلی در جای گذاری این است که با استفاده از روابط موجود میان داده ها، مقداری مناسب برای داده های گم شده برآورد و جایگزین گردد. در گروهی از روش های جایگذاری، این کار در یک مرحله انجام می شود. در این صورت به آن جای گذاری منفرد (Single Imputation) می گویند (Ghosh and Pahwa 2008). به عنوان مثال، یکی از ساده ترین شیوه ها در این گروه، محاسبه میانگین مقادیر مشاهدات موجود و جای گذاری آن بجای مقادیر گم شده (Mean Imputation) می باشد که البته در بسیاری از موارد نتایج مطالعه را با خطای زیادی همراه می نماید (Little and Rubin 2002).

رویکردهای بهتر و پیچیده تری نیز در میان روشهای جای گذاری منفرد وجود دارد. انواع برآورد های رگرسیونی جزء آنها محسوب می شوند که گاهی به نام جای گذاری میانگین شرطی (Conditional Mean Imputation) نیز خوانده می شوند. در این روش با استفاده از مدل رگرسیونی

شده است که به طور خلاصه به معرفی روش جای گذاری چندگانه (Multiple Imputation) به عنوان یک روش نوین مقابله با داده های گم شده پرداخته و مروری بر کاربردهای آن در مطالعات بهداشتی و اپیدمیولوژی انجام شود. گم شدن داده ها با سه مکانیسم متفاوت رخ می دهد (Little and Rubin 2002):

۱- گم شدن کاملاً تصادفی (Missing Completely At Random (MCAR): این حالت زمانی اتفاق می افتد که در آن فقدان داده ها با متغیرهای مورد مطالعه بی ارتباط باشد. مثلاً وقتی که تعدادی لوله آزمایش حاوی نمونه یا چند برگ از پرسشنامه های تکمیل شده به طور کاملاً تصادفی - مثلاً "حین حمل و نقل - گم یا خراب شوند، به طوری که دیگر نتوان از آنها داده ای را استخراج نمود.

۲- گم شدن تصادفی (Missing At Random (MAR): در این حالت رویه گم شدن داده ها به طور مشروط تصادفی است و این به این معناست که فقدان داده ها با متغیری که در آن نقص رخ داده بی ارتباط است، اما به دیگر متغیرهای مطالعه مربوط می شود. به عنوان مثال فرض کنید در یک تحقیق، جنسیت افراد، یکی از متغیرهای مورد مطالعه باشد. حال اگر به طور کاملاً تصادفی تعدادی از مشاهدات متعلق به افراد مذکر از دست برود، گم شدن داده ها از نوع تصادفی خواهد بود.

۳- گم شدن داده ها بطور غیرتصادفی (Missing Not At Random (MNAR): این وضعیت زمانی روی می دهد که فقدان داده های یک متغیر با مقادیر همان متغیر ارتباط داشته باشد. یک مثال خوب، بی پاسخی افراد دارای درآمد بالا در جواب به سوال در مورد سطح درآمدشان می باشد. از دیدگاه عملی مکانیسم رویداد نقصان در داده نقش تعیین کننده ای در کاربرد و انتخاب روش جای گذاری دارد. اگر بتوان فرض کرد مکانیسم گم شدن داده ها از نوع کاملاً تصادفی (MCAR) باشد و تحلیل در مطالعه بدون جایگذاری و صرفاً بر اساس داده های کامل انجام شود، انتظار نداریم نتایج دچار اریبی (Bias) قابل توجهی گردد و دست یابی به برآورد های قابل قبول

عرصه پزشکی و بهداشت هم از این امر مستثنی نبوده است. (Harel and Zhou 2007b; He 2010; Kenward and Carpenter 2007; Klebanoff and Cole 2008; Newgard and Haukoos 2007; Yu et al. 2007)

این تکنیک اولین بار توسط روبین (Rubin) در سال ۱۹۷۸ توصیف و در سالهای بعد توسط خود او و دیگران کاملتر گردید و همچنان جنبه های نظری و عملی بیشتری از آن در حال معرفی می باشد. (Barnard and Meng 1999). در جایگذاری متعدد نیز مقادیر گمشده، با استفاده از یک مدل مناسب رگرسیونی و بر اساس داده های مشاهده شده برآورد می گردند. اما تفاوت اصلی این روش با روش های جایگذاری منفرد این است که در جایگذاری چندگانه برای مقدار برآورد شده پراکندگی در نظر گرفته می شود و با انجام تکرار در جایگذاری به آن اجازه تغییر داده می شود (Graham et al. 2007). روش جای گذاری چندگانه به طور خلاصه دارای سه مرحله است. در مرحله اول  $M$  بار جای گذاری مستقل انجام و  $M$  سری داده کامل بدست می آید. در مرحله بعد برای هر یک از این  $M$  سری داده کامل، آنالیز استاندارد جداگانه ای انجام و برآورد های مورد نظر محاسبه می گردد. در مرحله آخر نتایج بدست آمده طبق قوانینی با یکدیگر ترکیب و یک نتیجه منفرد حاصل می گردد (Little and Rubin 2002). از نظر تئوری هرچه تعداد جایگذاری بیشتر باشد نتایج بهتری حاصل می شود و هرچه میزان گم شدگی داده ها کمتر باشد تعداد کمتری از جایگذاری کفایت می کند، اما در عمل  $M$  معمولاً بین ۳ تا ۱۰ بار می باشد (Ghosh and Pahwa 2008). روش جای گذاری متعدد تنها روش جدید موجود و مورد قبول نمی باشد. انعطاف پذیری و کارایی و قابلیت استفاده برای انواع داده ها خصوصاً داده های حجیم، شاخص ترین مزایایی هستند که به روش جای گذاری چندگانه مقبولیت بخشیده است (Horton and Kleinman 2007).

کاربرد جایگذاری متعدد در تحقیقات پزشکی و اپیدمیولوژی: در عرصه اپیدمیولوژی و تحقیقات بهداشتی،

و بر اساس مقادیر موجود، هر یک از مقادیر گم شده در یک مرحله برآورد و جایگزین می گردد. این رویکردها گرچه دقیقتر به نظر می رسند، اما هرگز عاری از خطا نیستند و تنها در صورتی قابل استفاده اند که مکانیسم گم شدن داده ها از نوع MCAR باشد. اما در عمل این شرط در اغلب موقعیتهای برآورده نمی شود. مشکل دیگری نیز وجود دارد و آن این است که در جای گذاری منفرد، چون هر یک از مقادیر جایگزین، صحیح ترین مقدار ممکن فرض شده و برای آن پراکندگی در نظر گرفته نمی شود، لذا برآورد های مورد نظر به غلط با اطمینانی بیشتر از مقدار واقعی به دست می آیند. این بدان معنی است که دقت مطالعه به طور کاذب بالا نشان داده می شود. لذا روشهای فوق در عمل دارای ضعف و اشکال هستند (Baccini 2008; Little and Rubin 2002).

در دهه های اخیر و در راستای غلبه بر محدودیتهای فوق، تلاشهای زیادی جهت معرفی و بکارگیری برخی شیوه های جدید (Modern Techniques) برای جایگذاری صورت گرفته است (Ghosh and Pahwa 2008). مهمترین وجه تمایز و برتری این روشها نسبت به روش های مرسوم این است که در صورتی که با آگاهی و به درستی به کار گرفته شوند، هم در مواردی که مکانیسم گم شدن داده ها از نوع کاملاً تصادفی (MCAR) است و هم درحالتی که فقدان داده ها با مکانیسم تصادفی (MAR) رخ می دهد، می توانند نتایج ناریب داشته باشند (Little and Rubin 2002; Patrician 2002). در این میان روش جای گذاری چندگانه - به علت کارایی نسبتاً بالای خود - بیشتر مورد توجه و استفاده پژوهشگران قرار گرفته است (Freedman and Wolf 1995; Schafer 1999; Sinharay et al. 2001; Zhou et al. 2004). در سال های اخیر، با ورود برنامه های نرم افزاری کاربردی برای جای گذاری متعدد، استفاده عملی از آن در فعالیتهای پژوهشی باز هم رواج و مقبولیت بیشتری یافته است (Horton and Kleinman 2007). تحقیقات در

است (Soullier et al. 2008). در تعدادی از مقالات نشان داده شده است، استفاده از روش جای گذاری چندگانه در مطالعات تحلیل بقا، دقت برآورد ها را نسبت به تحلیل بر اساس داده های کامل افزایش می دهد. به عنوان نمونه می توان به موردی اشاره کرد که این تحلیل برای مطالعه بقای بیماران مبتلا به سرطان پستان انجام شده است (Baneshi and Talei 2010).

مشکل فقدان داده ها در سایر مطالعات اپیدمیولوژیک نیز مهم و قابل توجه است. به عنوان نمونه می توان به داده های ارزشمند سیستم مراقبت های بهداشتی اشاره کرد. این داده ها به دلیل ماهیت ذاتی خود ممکن است دارای نقصان باشند. بنابراین استفاده از متدهای نوین در ترمیم آنها یک ایده سودمند و کارساز به نظر می رسد (Faris et al. 2002). اهمیت این مطلب باعث شد Rubin خود از اولین کسانی باشد که روش جای گذاری چندگانه را در بستر داده های مراقبت های بهداشتی پیشنهاد می نمایند (Rubin and Schenker 1991).

روش جای گذاری چندگانه در مطالعات مقطعی (Cross-Sectional) و بررسی (Survey) نیز کاربرد داشته و بهره گیری از آن در عرصه بهداشت عمومی (Public Health) مورد پیشنهاد و بحث قرار گرفته است (Zhou et al. 2001).

در این گونه مطالعات مهمترین دلیل گم شدن داده ها بی پاسخی (Non-response) است که می تواند یافته ها را دچار اربیبی نماید (Powers et al. 2005). به عنوان مثال در یک طرح پژوهشی با استفاده از جای گذاری چندگانه نسبت به برآورد میزانهای ابتلا به بیماری اوستئوپروز در جامعه اقدام شده است (Kmetz et al. 2002). همچنین تاکنون از این روش، به منظور بررسی وضعیت سلامت قلب و عروق (Arnold and Kang et al. 2003)، اندازه فشارخون (Kronmal 2003)، شکایات اورولوژیک (Taylor et al. 2002)، بررسی وضع سلامت زنان و شاخص توده بدنی آنها (Mishra and Dobson 2004)، مطالعه سلامت مادران

استفاده از جای گذاری چندگانه بیشتر به منظور تقلیل سوگیری ناشی از فقدان داده ها بوده است. این روش به منظور مدل سازی (Ferguson et al. 2008; Gelman et al. 2005; Yu and Tiwari 2006) و تحلیل سوگیری (Bias analysis) (Greenland 2009) نیز پیشنهاد شده یا به کار رفته است. همچنین نشان داده اند که با رویکرد جایگذاری چندگانه، می توان مکانیسم گم شدن داده ها را نیز تعیین و بررسی نمود (Fraser and Yan 2007).

فقدان داده ها و تأثیری که نبود این مقادیر روی صحت و دقت نتایج می گذارد، در پژوهشهای طولی (Longitudinal) و نیز مطالعات تحلیل بقا (Survival Analysis) دارای اهمیت مضاعفی است و به همین دلیل تاکنون حجم نسبتاً زیادی از متون و مقالات مربوط به تحلیل داده های گمشده و روشهای مقابله با آنها با این نوع مطالعات ارتباط دارد. از میان آنها در چندین مقاله متد جای گذاری چندگانه به عنوان یک روش مناسب جهت تقلیل آثار نامطلوب فقدان داده ها معرفی و به کار رفته است. یکی از اولین موارد استفاده از روش جایگذاری چندگانه در مطالعات اپیدمیولوژیک، به کار گیری آن در یک مطالعه همگروهی (cohort) برای بررسی توزیع فاصله زمانی میان تشخیص سرمی عفونت HIV و شروع بیماری ایدز بوده است (Taylor et al. 1990). از این شیوه برای برآورد بروز بیماری ایدز نیز استفاده کرده اند (Brancato et al. 1997). در یک مطالعه جالب نشان داده اند که با استفاده از روش جای گذاری چندگانه می توان به برآورد دقیقتری از میزان موفقیت باروری مصنوعی (IVF) In Vitro Fertilization دست یافت. بسیاری از زوجها پس از ناکامی در اولین اقدام ممکن است از انجام مجدد آن منصرف شوند و لذا نتیجه کار نامشخص و جزء داده های گم شده خواهد بود. در این مطالعه محقق پس از به کار گیری روش جای گذاری چندگانه برای مقابله با مشکل فقدان این داده ها اقدام به برآورد صحیح تری از میزان موفقیت عمل مذکور نموده

داده های ناشی از بی پاسخی و عدم همکاری افراد نیز استفاده کرده اند. (Taylor and Zhou 2009)

بررسی متون نشان از حضور جای گذاری چندگانه در مطالعات مورد شاهدهی (Case-Control) نیز دارد، گرچه این حضور نسبتاً "چشمگیر نبوده است" (Cordell et al. 2006; Schroeder and Weinberg 2001).

ترمیم داده های گم شده در حوزه مراقبتهای ویژه و طب اورژانس بسیار حایز اهمیت می باشد. از چالشهای مهم در این عرصه، نقصان داده های مربوط به عوامل موثر بر پیش آگهی بیماران می باشد. به عنوان نمونه در یک مطالعه جالب نشان داده شد استفاده از روش جای گذاری چندگانه در تحقیقات تروما، می تواند مناسبترین راه حل برای رفع مشکل نقصان اطلاعات در خصوص نمره گلاسکو برای ارزیابی کما (Glasgow Coma Scale (GCS باشد (Moore et al. 2005). در این راستا محققین روش جای گذاری چندگانه را با موفقیت به کار برده اند و آن را توصیه نموده اند (Moore et al. 2009a; Moore et al. 2009b; Perez et al. 2002). از روش مورد بحث برای ترمیم داده های مراکز ثبت حوادث (Trauma Registry Data) نیز می توان بهره گرفت (Hayes and Groner 2008).

برخورد با مشکل گم شدن داده ها در مطالعات سرطان نیز اهمیت زیادی دارد. با مرور متون و مقالات منتشر شده می بینیم جای گذاری چندگانه در جنبه های مختلفی از مطالعات انکولوژی مورد استفاده قرار گرفته است. برای مثال در یک مطالعه با هدف بررسی هزینه- اثربخشی شیمی درمانی در مقایسه با روش های معمول درمان تسکینی در بیماران مبتلا به سرطان سلول کوچک ریه، اجزای ناکامل داده های مربوط به هزینه ها با استفاده از رویکرد جایگذاری متعدد ترمیم گردیده است (Burton et al. 2007). اخیراً، جای گذاری چندگانه به عنوان یک ابزار سودمند برای افزایش صحت داده های مربوط به بانکهای اطلاعاتی مراکز ثبت بیماران سرطانی (Cancer Registry Databases) معرفی شده است (He et al.

و کودکان (Mao and Li 2005)، بررسی پوشش واکسیناسیون (Santibanez et al. 2005)، افسردگی در افراد طلاق گرفته (Sbarra and Emery 2005) و برآورد شیوع دیابت (Heidarian Miri 2010) استفاده شده است. در یک مطالعه با هدف بررسی میزان مناسب بودن انجام عمل جراحی کاتاراکت، از روش جای گذاری چندگانه برای اصلاح داده ها استفاده شده است (Choi et al. 2004). نشان داده اند با بهره گیری از روش مذکور می توان نسبت به ترمیم داده های ناقص مربوط به درآمد افراد نیز در تحقیقات مربوط به مراقبت سلامت جامعه محور (health surveillance Population-based) اقدام نمود (Zeng 2009). از دیدگاه نظری، وقتی اندازه حجم یا پیچیدگی یک مجموعه از داده ها افزایش می یابد، استفاده از روش های جایگذاری دچار محدودیت می شوند. یکی دیگر از موارد کاربرد این روش در مطالعات اپیدمیولوژیک، استفاده از آن در ترمیم داده های بزرگ (large data sets) و پیچیده می باشد (He et al. 2009; Stuart et al. 2009).

از جای گذاری چندگانه در مطالعات کارآزمایی بالینی (Clinical Trials) نیز استفاده کرده اند. به دلایل چالشها و محدودیت های اخلاقی، یکی از مهمترین مشکلات در کارآزماییهای بالینی، نیاز به انجام مطالعه روی کمترین حجم نمونه یا در کوتاه ترین زمان ممکن می باشد. یکی از موارد کاربرد جایگذاری چندگانه، متوقف کردن زود هنگام کارآزماییهای بالینی دشوار و پر چالش بوده است (Betensky 1998). به این ترتیب که در مراحل از انجام مطالعه که بدلائل اخلاقی یا تکنیکی امکان اندازه گیری اثر مداخله وجود ندارد، نتیجه اندازه گیری مشاهده نشده به عنوان داده گمشده در نظر گرفته می شود و با استفاده از جایگذاری چندگانه مقدار مناسب برآورد می گردد. جایگذاری متعدد در انجام کارآزماییهای بالینی با حجم نمونه کم نیز سودمندی خود را نشان داده است (Barnes et al. 2006). از آن برای ترمیم نقصان

برخی شرایط ممکن است بررسی نمونه موجود منتهی به نتایج قاطع نگردد و این در حالی است که اخذ مجدد نمونه ناممکن، دشوار و یا غیر اخلاقی است. در این حالت نتیجه بررسی نمونه می تواند یک داده گم شده تلقی گردد و در تحلیل با آن مانند یک داده گم شده رفتار شود. با عنایت به این که در بررسی آزمون ها همیشه نتیجه آزمون استاندارد طلایی در دسترس نیست، محققان نشان داده اند که می توان با کمک رویکرد جایگذاری چندگانه، ابتدا نتایج نامعلوم و گم شده برآورد و سپس اقدام به ارزیابی صحت (Accuracy) روشهای تشخیصی نظیر تستهای آزمایشگاهی و نیز آزمونهای غربالگری نمود (Albert 2007; Harel and Zhou 2007a; et al. 2007).

نقش تحلیل داده های گمشده در مطالعات ژنتیک نیز بسیار قابل توجه است. در سالهای اخیر به موازات پیشرفت بشر در عرصه بیولوژی سلولی-مولکولی و افزایش توانایی در تعریف و تعیین مارکرهای ژنی قابل اندازه گیری، بسیاری از مطالعات در حوزه ژنتیک پزشکی به سمت بررسی ارتباط میان این مارکرها و پیامدهای سلامت معطوف گردیده است. اما همواره حجم مترکم اطلاعات مربوط به مارکرها و جایگاههای ژنی، کامل نیستند و این در حالی است که هنوز تعیین و بررسی مارکرهای ژنتیکی نیازمند روشهای نسبتاً پرهزینه می باشد. لذا ژنتیک حیطة ای است که استفاده از تکنیکهای کار آمد جایگزینی داده ها در آن دارای اهمیت و ضرورت بوده و از روش جای گذاری چندگانه برای ترمیم داده ها به طور فزاینده ای در مطالعات مربوط استفاده شده است. به عنوان نمونه، از جای گذاری چندگانه در مطالعاتی که بر اساس نقشه های پرتراکم پلی مورفیسمهای تک نوکلئوتیدی (High-density Single-Nucleotide Polymorphism, SNP Maps Croiseau et al. 2007) انجام می شوند، با موفقیت استفاده شده است (et al. 2007). این شیوه به عنوان یک ابزار کارآمد برای تقلیل خطای اندازه گیری (Measurement Error) در برخی مطالعات ژنتیکی نیز معرفی گردیده است (Padilla et al. 2009). علاوه بر آن، جای گذاری چندگانه در

(2008). از این روش در بررسی عوامل خطر سرطان ریه (Altenburg et al. 2002)، ترمیم داده های ناقص کیفیت زندگی در بیماران سرطانی در مراحل پایانی (Morita et al. 2005)، ترمیم موارد نقص بیومارکر ها در مطالعه مربوط به سرطان پستان (Emerson et al. 2009) و نیز جهت ارزیابی صحت روشهای رادیوگرافیک در تشخیص سرطان معده (Albert 2007) با موفقیت استفاده شده است. کاربرد روش فوق در عرصه داده های رادیولوژی محدود به بیماران سرطانی نمی شود. به عنوان نمونه در یک مطالعه از روش مورد بحث برای ترمیم داده های مربوط به تصویر برداری تشدید مغناطیسی قلب (Cardiac Magnetic Resonance Imaging) در بیماران که به دلایلی برای ایشان این تصویر برداری انجام نگرفته، استفاده شده است (Data Delaney et al. 2009).

علاوه بر مطالعات مربوط به تصویر برداری، بررسی متون نشان می دهد روش جای گذاری چندگانه در ترمیم داده های گمشده آزمایشگاهی نیز مفید بوده است. به عنوان مثال در یک مطالعه در حیطه اپیدمیولوژی بیماریهای عفونی با بهره گیری از روش مذکور مقادیر گم شده آلبومین سرم بیماران به طور موفق ترمیم گردید (Mulla et al. 2009). در مطالعه ای دیگر، به منظور ترمیم داده های گم شده در خصوص سطح سرمی فاکتور نکروز تومور در بیماران مبتلا به آرتریت روماتوئید از روش مذکور استفاده شده است (Schiattino et al. 2005).

تعیین حساسیت و ویژگی روشها و تستهای نوین تشخیصی در دامنه تحقیقات بهداشتی و بالینی نیازمند در اختیار داشتن روشهای استاندارد طلایی معتبر و قابل اتکایی است که عموماً پرهزینه یا تهاجمی هستند. به عنوان مثال می توان به روش های نوین یا غیرتهاجمی تشخیص بیماریهای سرطانی اشاره کرد. برای تعیین حساسیت و ویژگی این آزمون ها اغلب نیاز به اطلاع از نتیجه بررسی نمونه آسیب شناسی افراد وجود دارد. اما در

آموزش دادن باشد و یا الگوریتم های عملیاتی برای آن تعریف نگردد، عملاً جایگاهی در حل مشکلات نخواهد داشت. بسیاری از پژوهشگران، به دلیل عدم شناخت کافی از مبانی تئوریک آن و نیز کمبود نمونه های عملی و کاربردی، به روش های مدرن جای گذاری به دیده ابزاری برای داده پردازی (Data makeup) نگاه می کنند. اما در واقع هدف از به کار گیری روش جای گذاری چندگانه، نه داده پردازی، بلکه استفاده هر چه بیشتر از مشاهدات واقعی است. تحلیل بر اساس داده های کامل تنها در صورتی دارای وجهت علمی خواهد بود که مکانیسم فقدان داده ها MCAR باشد. اما حقیقت این است که اگر نتوان برآورده شدن این شرط مهم را نشان داد بی اعتنایی به مشاهدات مفقود شده و تحلیل بر اساس داده های کامل، مصداق قابل قبول تری از داده پردازی به نظر می رسد و به تبع آن بروز سوگیری بیشتری در نتایج قابل انتظار خواهد بود. البته این بدان معنا نیست که در همه شرایط روشهای جدید جای گذاری مناسب و کافی هستند. بلکه این روشها نیز در صورت استفاده نابجا و غلط می توانند منجر به نتایج نادرست و دارای سوگیری گردند و اگر به اندازه کافی در این خصوص دقت و احتیاط نشود به کارگیری آنها می تواند خطرناک باشد (Ghosh and Pahwa 2008). این مسئله در حال حاضر یک چالش مهم محسوب می گردد و بایستی وجهت استفاده از این روش در شرایط مختلف با انجام برخی مطالعات تخصصی ارزیابی شود. در تعدادی از مقالات که در سالهای اخیر منتشر شده اند، نقاط قوت و ضعف و نیز کاربرد این روش در مطالعات پزشکی و اپیدمیولوژیک در مقایسه با سایر روشها خصوصاً "روش تحلیل داده های کامل، مورد بررسی قرار گرفته است (Horton et al. 2010; Sterne et al. 2009) از مطالعات تحلیل حساسیت (Sensitivity Analysis) نیز برای نشان دادن کارایی متد جای گذاری چندگانه، استفاده گردیده است (Carpenter et al. 2007). در این مقالات محقق شرایط مختلف گمشدگی را روی داده های واقعی یا تولید شده شبیه سازی می کند و سپس با استفاده از روش

مطالعات بررسی رابطه بر اساس وابستگی های خانوادگی (Family-based Association Studies) نیز کاربرد داشته است. هدف از انجام این مطالعات بررسی خطر افزایش یافته بیماری در وابستگان و خویشاوندان می باشد (Croiseau et al. 2007).

## بحث

با بررسی مقالات و اقدامات پژوهشی صورت گرفته در عرصه پزشکی و اپیدمیولوژی در ارتباط با روش جایگذاری چندگانه ملاحظه می گردد عمدتاً سه هدف اساسی برای محققین مطرح بوده است: معرفی و آشنایی باتکنیک جایگذاری متعدد و جنبه های مختلف نظری آن به عنوان یک شیوه نوین در برخورد با داده های گمشده، نشان دادن سودمندی این روش به عنوان یک شیوه موثر در مقابله با مشکلات و چالشهای ناشی از نقصان داده ها و استفاده کاربردی از آن به عنوان یک ابزار تحلیلی در بسترهای مختلف مطالعاتی. نمونه هایی از این مقالات را می توان در جدول ۱، مشاهده نمود.

گرچه اهداف مذکور در بسیاری از موارد از یکدیگر قابل تفکیک نیستند، اما مرور مقالات منتشر شده نشان از این واقعیت دارد که رویکرد جایگزینی چندگانه، علی رغم کارایی و انعطافی که دارد، آن طور که باید و شاید در حوزه های مختلف تحقیقات بهداشتی و اپیدمیولوژی مورد توجه و استفاده پژوهشگران قرار نگرفته است. در واقع، هنوز اغلب محققین ترجیح می دهند به جای استفاده از روش های مدرن جای گذاری، از روش تحلیل بر اساس داده های کامل استفاده نمایند. شاید مهمترین عاملی که در متون برای این امر ذکر شده، این باشد که در حوزه مطالعات بهداشتی مواردی که بتوانند به عنوان راهنمای عملیاتی برای انجام این متد مورد استفاده محققین قرار بگیرد نسبتاً کم تعداد هستند (Stuart et al. 2009). در توضیح این مطلب باید گفت اگر یک شیوه خواص بسیار عالی داشته باشد اما به دلایلی نظیر پیچیدگی یا کمبود افراد آشنا یا متخصص غیر قابل

علی رغم همه چالش ها و دشواری ها، به نظر می رسد استفاده فراگیر از رویکرد جایگذاری چندگانه و درس هایی که به طور عملی از آنها حاصل خواهد شد، در آینده نزدیک، جایگاه مهم و تعریف شده ای را در عرصه های مختلف تحقیقات پزشکی و اپیدمیولوژی به آن اختصاص دهد (Kenward and Carpenter 2007).

های مختلف اقدام به تحلیل می نماید. با این کار محقق می تواند نشان دهد که استفاده از روش جایگذاری چندگانه در چه شرایطی نتایج خوب و قابل قبول دارد و در کدام حالات قابل استفاده نیست.

## نتیجه گیری

جدول ۱- نمونه هایی از کاربرد روش جای گذاری چندگانه در مطالعات انجام شده در عرصه تحقیقات پزشکی و اپیدمیولوژی

شرح مختصر کاربرد	نویسنده و سال انتشار مقاله*
ارزیابی صحت روشهای رادیوگرافیک تشخیصی سرطان معده	Albert 2007
ترمیم داده ها در بررسی وضعیت ابتلا به بیماری های قلب و عروق	Arnold and Kronmal 2003
در تحلیل بقای بیماران سرطان پستان	Baneshi and Talei 2010
انجام کارآزمایی بالینی با حجم نمونه کم	Barnes et al. 2006
متوقف کردن زود هنگام مطالعات کارآزمایی بالینی	Betensky 1998
ترمیم داده ها در مطالعه روی پلی مورفیسمهای تک نوکلئوتیدی	Croiseau et al. 2007
ترمیم موارد نقص بیومارکرها در سرطان پستان	Emerson et al. 2009
کاربرد جای گذاری چندگانه در مدل سازی	Ferguson et al. 2008
کاربرد جای گذاری چندگانه در تعیین مکانیسم گم شدن داده ها	Fraser and Yan 2007
کاربرد جای گذاری چندگانه در تحلیل بایاس	Greenland 2009
ترمیم داده های مراکز ثبت حوادث	Hayes and Groner 2008
ترمیم داده های مراکز ثبت سرطان	He et al. 2008
ترمیم داده ها در بررسی شیوع دیابت	Heidarian Miri 2010
ترمیم داده ها در بررسی وضعیت ابتلا به پر فشاری خون	Kang et al. 2003
ترمیم داده ها در بررسی وضعیت ابتلا به اوستئوپروز	Kmetic et al. 2003
ترمیم داده ها در بررسی وضع سلامت مادران و کودکان	Mao and Li 2005
ترمیم داده ها در بررسی نمایه توده بدنی زنان	Mishra and Dobson 2004
ترمیم اجزای نمره گلاسکو برای ارزیابی کما	Moore et al. 2009,2005
ترمیم داده های ناقص کیفیت زندگی در بیماران سرطانی	Morita et al. 2005
ترمیم مقادیر گم شده آلبومین سرم	Mulla et al. 2009
ترمیم داده ها در بررسی پوشش واکسیناسیون	Santibanez et al. 2005
ترمیم داده ها در بررسی افسردگی در افراد طلاق گرفته	Sbarra and Emery 2005
ترمیم داده ها در تعیین میزان موفقیت باروری مصنوعی	Soullier et al. 2008
ترمیم داده ها در بررسی فاصله زمانی میان تشخیص HIV و شروع بیماری	Taylor et al. 1990
ترمیم داده ها در بررسی وضعیت شکایات اورولوژیک	Taylor et al. 2002
ترمیم داده های ناقص مربوط به درآمد	Zeng 2009

• مقالات بر اساس نام نویسنده اول و به ترتیب حروف الفبا در جدول آورده شده اند.



## References

- Albert, P.S., 2007. Imputation approaches for estimating diagnostic accuracy for multiple tests from partially verified designs. *Biometrics*, **63**(3), pp. 947-57.
- Altenburg, H.P., Agudo, A., Berrino, f., Boshuizen, H.C., Bueno-de-Mesquita, H.B., Janzon, L., Le Marchand, L., Linseisen, J., Lukanova, A., Rasmuson, T., Vineis, P., Riboli, E. and Miller, A., 2002. Using multiple imputation methods to estimate relative risks in small EPIC lung cancer subsets. *IARC Sci Publ*, 156, pp.53-4.
- Arnold, A.M. and Kronmal, R.A., 2003. Multiple imputation of baseline data in the cardiovascular health study. *Am J Epidemiol*, **157**(1), pp.74-84.
- Baccini, M., 2008. Multiple imputation for missing data: a brief introduction. *Epidemiol Prev*, **32**(3), pp.162-3.
- Baneshi, M.R. and Talei, A.R., 2010. Impact of Missing Data on Estimation of Survival Rates: An Example in breast Cancer. *Iran J Cancer Prev*, **3**(3), pp.127-131.
- Barnard, J. and Meng, X.L., 1999. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res*, **8**(1), pp.17-36.
- Barnes, S.A., Lindborg, S.R. and Seaman, J.W., 2006. Multiple imputation techniques in small sample clinical trials. *Stat Med*, **25**(2), pp.233-45.
- Betensky, R.A., 1998. Multiple imputation for early stopping of a complex clinical trial. *Biometrics*, **54**(1), pp.229-42.
- Brancato, G., Pezzotti, P., E. Rapiti, E., Perucci, C.A., Abeni, D., Babbalacchio, A. and Rezza, G., 1997. Multiple imputation method for estimating incidence of HIV infection. The Multicenter Prospective HIV Study. *Int J Epidemiol*, **26**(5), pp.1107-14.
- Burton, A., Billingham, L.J. and Bryan, S., 2007. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clin Trials*, **4**(2), pp.154-61.
- Carpenter, J.R., Kenward, M.G. and White, I.R., 2007. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Stat Methods Med Res*, **16**(3), pp.259-75.
- Choi, Y. J., Nam, C.M. and Kwak, M.J., 2004. Multiple imputation technique applied to appropriateness ratings in cataract surgery. *Yonsei Med J*. **45**(5): pp. 829-37.
- Cordell, H.J., 2006. Estimation and testing of genotype and haplotype effects in case-control studies: comparison of weighted regression and multiple imputation procedures. *Genet Epidemiol*, **30**(3), pp.259-75.
- Croiseau, P., Bardel, C. and Genin, E., 2007. Efficiency of multiple imputation to test for association in the presence of missing data. *BMC Proc*, 1 Suppl 1: S24.
- Croiseau, P., Genin, E. and Cordell, H.J., 2007. Dealing with missing data in family-based association studies: a multiple imputation approach. *Hum Hered*, **63**(3-4), pp.229-38.
- Delaney, J.A., McClelland, R.L., Brown, E., Bluemke, D.A., Vogel-Claussen, J., Lai, S. and Heckbert S.R., 2009. Multiple imputation for missing with cardiac magnetic resonance imaging data: results from the Multi-Ethnic Study of Atherosclerosis (MESA). *Can J Cardiol*, **25**(7), pp.232-5.
- Emerson, J.W., Dolled-Filhart, M., Harris, L., Rimm, D.L. and Tuck, D.P., 2009. Quantitative assessment of tissue biomarkers and construction of a model to predict outcome in breast cancer using multiple imputation. *Cancer Inform*, 7, pp.29-40.
- Faris, P.D., Ghali, W.A., Brant, R., Norris, C.M., Galbraith, P.D. and Knudtson, M.L., 2002. Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. *J Clin Epidemiol*, **55**(2), pp.184-91.
- Ferguson, M. K., Siddique, J. and Karrison, T., 2008. Modeling major lung resection outcomes using classification trees and multiple imputation techniques. *Eur J Cardiothorac Surg*, **34**(5), pp.1085-9.

- Fraser, G. and Yan, R., 2007. Guided multiple imputation of missing data: using a subsample to strengthen the missing-at-random assumption. *Epidemiology*, **18**(2), pp.246-52.
- Freedman, V.A. and Wolf, D.A., 1995. A case study on the use of multiple imputation. *Demography*, **32**(3), pp.459-70.
- Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D.F. and Meulders, M., 2005. Multiple imputation for model checking: completed-data plots with missing and latent data. *Biometrics*, **61**(1), pp.74-85.
- Ghosh, S. and Pahwa, P.P., 2008. Assessing Bias Associated With Missing Data from Joint Canada/U.S. Survey of Health: An Application. *Biometrics Section*, pp.3394-401.
- Graham, J.W., Olchowski, A.E. and Gilreath, T.D., 2007. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci*, **8**(3), pp.206-13.
- Greenland, S., 2009. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *Int J Epidemiol*, **38**(6), pp.1662-73.
- Harel, O., Hofer, S.M., Hoffman, L., Pedersen, N.L. and Johansson, B., 2007. Population inference with mortality and attrition in longitudinal studies on aging: a two-stage multiple imputation method. *Exp Aging Res*, **33**(2), pp.187-203.
- Harel, O. and Zhou, X.H., 2007a. Multiple imputation for the comparison of two screening tests in two-phase Alzheimer studies. *Stat Med*, **26**(11), pp.2370-88.
- Harel, O. and Zhou X.H., 2007b. Multiple imputation: review of theory, implementation and software. *Stat Med*, **26**(16), pp.3057-77.
- Hayes, J.R. and Groner, J.I., 2008. Using multiple imputation and propensity scores to test the effect of car seats and seat belt usage on injury severity from trauma registry data. *J Pediatr Surg*, **43**(5), pp. 924-7.
- He, Y., 2010. Missing data analysis using multiple imputation: getting to the heart of the matter. *Circ Cardiovasc Qual Outcomes*, **3**(1), pp.98-105.
- He, Y., Yucel, R. and Zaslavsky, A.M., 2008. Misreporting, Missing Data, and Multiple Imputation: Improving Accuracy of Cancer Registry Databases. *Chance (N Y)*, **21**(3), pp.55-58.
- He, Y., Zaslavsky, A., Landrum, M., D. Harrington, D. and Catalano, P., 2009. Multiple imputation in a large-scale complex survey: a practical guide. *Stat Methods Med Res*.
- Heidarian Miri, H., 2010. *Study of prevalence and risk factors of diabet type II in Iran*. M.Sc. Tehran University of Medical Sciences. [In Persian]
- Horton, N.J. and Kleinman, K.P., 2007. Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data Regression Models. *The American Statistician*, **61**(1), pp.79-90.
- Horton, N.J., White, I.R. and Carpenter, J., 2010. The performance of multiple imputation for missing covariates relative to complete case analysis. *Stat Med*, **29**(12), pp.1357.
- Joseph, L., Belisle, P., Tamim, H. and Sampalis, J.S., 2004. Selection bias found in interpreting analyses with missing data for the prehospital index for trauma. *J Clin Epidemiol*, **57**(2), pp.147-53.
- Kang, T., Kraft, P., Gauderman, W.J. and Thomas, D., 2003. Multiple imputation methods for longitudinal blood pressure measurements from the Framingham Heart Study. *BMC Genet*, 4 Suppl 1: S43.
- Kenward, M.G. and Carpenter, J., 2007. Multiple imputation: current perspectives. *Stat Methods Med Res*, **16**(3), pp.199-218.
- Klebanoff, M.A. and Cole, S.R., 2008. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol*, **168**(4), pp.355-7.
- Kmetz, A., Joseph, L., Berger, C. and Tenenhouse A., 2002. Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. *Epidemiology*. **13**(4), pp.437-44.

- Little, R.A. and Rubin, D.B., 2002. *Statistical Analysis with Missing Data*, John Wiley and Sons, New York.
- Mao, Q.X. and Li, X.S., 2005. Markov Chain Monte Carlo Method of multiple imputation for longitudinal data with missing values in the survey of maternal and children health. *Sichuan Da Xue Xue Bao Yi Xue Ban*, **36**(3), pp.422-5.
- Mishra, G.D. and Dobson, A.J., 2004. Multiple imputation for body mass index: lessons from the Australian Longitudinal Study on Women's Health. *Stat Med*, **23**(19), pp. 3077-87.
- Moore, L., Hanley, J.A., Lavoie, A. and Turgeon, A., 2009a. Evaluating the validity of multiple imputation for missing physiological data in the national trauma data bank. *J Emerg Trauma Shock*, **2**(2), pp.73-9.
- Moore, L., Hanley, J.A., Turgeon, A.F., Lavoie, A. and Emond, M., 2009b. A multiple imputation model for imputing missing physiologic data in the national trauma data bank. *J Am Coll Surg*, **209**(5), pp. 572-9.
- Moore, L., Lavoie, A., LeSage, N., Liberman, M., Sampalis, J.S., Bergeron, E. and Abdous, B., 2005. Multiple imputation of the Glasgow Coma Score. *J Trauma*, **59**(3), pp.698-704.
- Morita, S., Kobayashi, K., Eguchi, K., Matsumoto, T., Shibuya, M., Yamaji, Y. and Ohashi, Y., 2005. Analysis of incomplete quality of life data in advanced stage cancer: a practical application of multiple imputation. *Qual Life Res*, **14**(6), pp.1533-44.
- Mulla, Z.D., Seo, B., Kalamegham, R. and Nuwayhid, B.S., 2009. Multiple imputation for missing laboratory data: an example from infectious disease epidemiology. *Ann Epidemiol*, **19**(12), pp.908-14.
- Newgard, C.D. and Haukoos, J.S., 2007. Advanced statistics: missing data in clinical research--part 2: multiple imputation. *Acad Emerg Med*, **14**(7), pp.669-78.
- Padilla, M.A., Divers, J., Vaughan, L.K., Allison, D.B. and Tiwari, H.K., 2009. Multiple imputation to correct for measurement error in admixture estimates in genetic structured association testing. *Hum Hered*, **68**(1), pp.65-72.
- Patrician, P.A., 2002. Multiple imputation for missing data. *Res Nurs Health*. **25**(1), pp.76-84.
- Perez, A., Dennis, R.J., Gil, J.F., Rondon, M.A. and Lopez, A., 2002. Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Colombia. *Stat Med*, **21**(24), pp.3885-96.
- Powers, J.R., Mishra, G. and Young, A.F., 2005. Differences in mail and telephone responses to self-rated health: use of multiple imputation in correcting for response bias. *Aust N Z J Public Health*, **29**(2), pp.149-54.
- Rubin, D.B. and Schenker, N., 1991. Multiple imputation in health-care databases: an overview and some applications. *Stat Med*, **10**(4), pp.585-98.
- Santibanez, T.A., Barker, L.E. and Shaw, K.M., 2005. Measurement of vaccination coverage at age 24 and 19-35 months: a case study of multiple imputation in public health. *Popul Health Metr*, **3**, P. 6.
- Sbarra, D.A. and Emery, R.E., 2005. Coparenting conflict, nonacceptance, and depression among divorced adults: results from a 12-year follow-up study of child custody mediation using multiple imputation. *Am J Orthopsychiatry*, **75**(1), pp.63-75.
- Schafer, J.L., 1999. Multiple imputation: a primer. *Stat Methods Med Res*, **8**(1), pp.3-15.
- Schiattino, I., Villegas, R., Cruzat, A., Cuenca, J., Salazar, L., Aravena, O., Pesce, B., Catalan, D., Llanos, C., Cuchacovich, M. and Aguillon, J.C., 2005. Multiple imputation procedures allow the rescue of missing data: an application to determine serum tumor necrosis factor (TNF) concentration values during the treatment of rheumatoid arthritis patients with anti-TNF therapy. *Biol Res*, **38**(1), pp.7-12.
- Schroeder, J.C. and Weinberg, C.R., 2001. Use of missing-data methods to correct bias

- and improve precision in case-control studies in which cases are subtyped but subtype information is incomplete. *Am J Epidemiol*, **154**(10), pp.954-62.
- Sinharay, S., Stern, H.S. and Russell, D., 2001. The use of multiple imputation for the analysis of missing data. *Psychol Methods*, **6**(4), pp.317-29.
- Soullier, N., Bouyer, J., Pouly, J.L., Guibert, J. and de La Rochebrochard, E., 2008. Estimating the success of an in vitro fertilization programme using multiple imputation. *Hum Reprod*, **23**(1), pp.187-92.
- Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.J., Wood, A.M. and Carpenter J.R., 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338: b2393.
- Stuart, E.A., Azur, M., Frangakis, C. and Leaf, P., 2009. Multiple imputation with large data sets: a case study of the Children's Mental Health Initiative. *Am J Epidemiol*, **169**(9), pp.1133-9.
- Taylor, J.M., Cooper, K.L., Wei, J.T., Sarma, A.V., Raghunathan, T.E. and Heeringa, S.G., 2002. Use of multiple imputation to correct for nonresponse bias in a survey of urologic symptoms among African-American men. *Am J Epidemiol*, **156**(8), pp.774-82.
- Taylor, J.M., Munoz, A., Bass, S.M., Saah, A.J., Chmiel, J.S. and Kingsley, L.A., 1990. Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation. Multicentre AIDS Cohort Study. *Stat Med*. **9**(5), pp.505-14.
- Taylor, L. and Zhou, X.H., 2009. Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics*, **65**(1), pp.88-95.
- Yu, B. and Tiwari, R.C., 2006. Multiple imputation methods for modelling relative survival data. *Stat Med*, **25**(17), pp.2946-55.
- Yu, L.M., Burton, A. and Rivero-Arias, O., 2007. Evaluation of software for multiple imputation of semi-continuous data. *Stat Methods Med Res*, **16**(3), pp.243-58.
- Zeng, Z., 2009. Multiple imputation for missing income data in population-based health surveillance. *J Public Health Manag Pract*, **15**(6), pp.12-21.
- Zhou, X.H., Eckert, G.J. and Tierney, W.M., 2001. Multiple imputation in public health research. *Stat Med*, **20**(9-10), pp.1541-9.
- Zhou, Y.B., Jiang, Q.W. and Zhao, G.M., 2004. A method to manage of miss data--multiple imputation. *Zhonghua Yu Fang Yi Xue Za Zhi*, **38**(6), pp.424-6.