







دانشگاه علوم پزشکی  
و خدمات بهداشتی درمانی تهران

# طرح های پژوهش

تجربی، غیر تجربی

تألیف:

دکتر نورالدین نخستین انصاری

دکتر احسان مقیمی

دکتر عبدالرحمان رستمیان



مرکز تحقیقات آسیب دیدگان جنگ



دفتر امور ایثارگران

Research Center for War-affected People

[www.issar.tums.ac.ir](http://www.issar.tums.ac.ir)



انتشارات دانش شناس



|                              |                                                                                                                                                                                                              |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| سرشناسه :                    | نخستین انصاری، نورالدین، ۱۳۴۰ -                                                                                                                                                                              |
| عنوان و نام پدیدآور :        | طرح‌های پژوهشی: تجربی، غیرتجربی / تألیف نورالدین نخستین انصاری، احسان مقیمی، عبدالرحمان رستمیان؛ [برای] دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران، دفتر امور ایثارگران و مرکز تحقیقات آسیب‌دیدگان جنگ. |
| مشخصات نشر :                 | تهران: دانش شناس، ۱۴۰۰.                                                                                                                                                                                      |
| مشخصات ظاهری :               | ۱۶۸ ص؛ ۱۴/۵×۲۱/۵ س.م.                                                                                                                                                                                        |
| شابک :                       | 978-600-8598-69-5                                                                                                                                                                                            |
| وضعیت فهرست نویسی :          | فیبا                                                                                                                                                                                                         |
| یادداشت :                    | کتابنامه: ص. [۱۷۹] - ۱۸۹.                                                                                                                                                                                    |
| موضوع :                      | طرح‌نویسی در تحقیق                                                                                                                                                                                           |
| Proposal writing in research |                                                                                                                                                                                                              |
| تحقیق -- روش‌شناسی           |                                                                                                                                                                                                              |
| Research -- Methodology      |                                                                                                                                                                                                              |
| شناسه افزوده :               | مقیمی، احسان، ۱۳۴۱ -                                                                                                                                                                                         |
| شناسه افزوده :               | رستمیان، عبدالرحمان، ۱۳۳۹ -                                                                                                                                                                                  |
| شناسه افزوده :               | دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران. دفتر امور ایثارگران و مرکز تحقیقات آسیب دیدگان جنگ                                                                                                          |
| رده بندی کنگره :             | ۵۵/۱۸۰Q                                                                                                                                                                                                      |
| رده بندی دیویی :             | ۴۴/۰۰۱                                                                                                                                                                                                       |
| شماره کتابشناسی ملی :        | ۸۵۴۸۷۲۱                                                                                                                                                                                                      |

### ناشر: انتشارات دانش شناس

نام کتاب: طرح‌های پژوهش تجربی، غیرتجربی

مؤلف: دکتر نورالدین نخستین انصاری؛ دکتر احسان مقیمی؛

دکتر عبدالرحمان رستمیان

محل نشر: تهران تاریخ نشر: ۱۴۰۰

نوبت چاپ: اول؛ ویرایش: اول

شمارگان: ۵۰۰ شابک: ۵-۶۹-۸۵۹۸-۶۰۰-۹۷۸ ص پ: ۶۹۶-۳۱۴۵

<http://www.daneshshenas.com> Email: [daneshshenas@gmail.com](mailto:daneshshenas@gmail.com)

[https://www.instagram.com/daneshshenas\\_publishing](https://www.instagram.com/daneshshenas_publishing)



کلینیسین ها، دانشجویان، و پژوهشگران باید برای طرح ریزی پروژه های خود، با اصول پژوهش، متودولوژی، و کاربردهای آنها آشنا باشند. طرح پژوهش (Research design) سازنده ساختار یا برنامه (Plan) پژوهش است. پژوهشگران در تحقیقات خود از ساختارها و پارادایم های پژوهشی مختلفی استفاده می کنند. بسته به پارادایم پژوهشی (کمی، کیفی، مرکب)، طرح هایی برای پیشبرد پژوهش وجود دارد. پارادایم کمی (Quantitative) یا همان پژوهش علمی کلاسیک رویکرد رایج در پژوهش است که هدف کلی و ساختار را برای پژوهش تعیین می کند. در هنگام انتخاب طرح پژوهش، باید به دو موضوع توجه کنیم: چه هدفی داریم و چگونه می خواهیم به آن هدف دست یابیم. این پروسه سازنده متودولوژی تحقیق است که چگونگی انجام پژوهش را مشخص می کند. طرح های پژوهشی بسیاری برای طراحی پروژه های کمی وجود دارد. انتخاب طرح پژوهش، به سوال و هدف پژوهش بستگی دارد. ابزارهای بسیاری مثل متدها، تئوری ها، و غیره وجود دارند که ما برای ساختن طرح پژوهش، از آنها استفاده می کنیم. در طراحی یک پروژه، اطلاعات تئوریک در مورد موضوع، تجربیات حرفه ای، جنبه های اخلاقی، و مهارت عملی نقش دارند. در کتاب حاضر تلاش شده است که طرح های پژوهشی تجربی، غیرتجربی، اپیدمیولوژیک به طور کاربردی، بر مبنای نیاز دانشجویان و پژوهشگران، و به همراه مثال هایی از پژوهش های انجام شده ارائه گردد. فصل پایانی نیز به تئوری های اندازه گیری، انواع متغیرها، ویژگی پایایی و اعتبار اندازه گیری ها، و پاسخ دهی آنها به تغییرات، اختصاص یافته است که بخش قابل توجهی از تحقیقات منتشر شده در این حیطه را تشکیل می دهند.

لازم می دانیم از زحمات سرکار خانم دکتر صوفیا نقدی استاد گروه فیزیوتراپی دانشکده توانبخشی دانشگاه علوم پزشکی تهران در ویرایش متن این کتاب، سپاسگزاری نماییم.



|                                         |    |
|-----------------------------------------|----|
| پیشگفتار.....                           | ۵  |
| فصل اول.....                            | ۱۳ |
| طرح های پژوهش تجربی.....                | ۱۹ |
| طرح های گروهی.....                      | ۲۰ |
| فرضیات طرح های گروهی.....               | ۲۳ |
| کارآزمایی کنترل شده تصادفی.....         | ۲۴ |
| طرح های تجربی تک عاملی.....             | ۲۷ |
| طرح گروه-کنترل پیش آزمون- پس آزمون..... | ۲۸ |
| طرح گروه-کنترل فقط پس آزمون.....        | ۳۰ |
| طرح یک گروه پیش آزمون - پس آزمون.....   | ۳۱ |
| طرح گروه-کنترل نابرابر.....             | ۳۳ |
| طرح سری زمانی.....                      | ۳۵ |
| خلاصه ویژگی های طرح سری های زمانی.....  | ۳۶ |
| طرح اندازه گیری های مکرر.....           | ۳۸ |
| مزایای طرح اندازه گیری مکرر.....        | ۴۴ |
| معایب طرح اندازه گیری مکرر.....         | ۴۵ |

|                                           |    |
|-------------------------------------------|----|
| طرح پژوهش چندعاملی.....                   | ۴۵ |
| طرح های فاکتوریال و آشیانه ای.....        | ۴۹ |
| طرح آشیانه ای.....                        | ۵۳ |
| طرح های کاملاً تصادفی و بلوک تصادفی.....  | ۵۵ |
| طرح های بین گروهی، درون گروهی و مرکب..... | ۵۷ |
| خلاصه.....                                | ۶۱ |
| فصل دوم.....                              | ۶۳ |
| طرح های پژوهش غیرتجربی.....               | ۶۳ |
| پژوهش توصیفی.....                         | ۶۹ |
| پژوهش توصیفی گذشته نگر.....               | ۷۰ |
| پژوهش برای ارزیابی.....                   | ۷۱ |
| پژوهش اپیدمیولوژیک.....                   | ۷۲ |
| پژوهش تاریخی.....                         | ۷۲ |
| پژوهش توصیفی آینده نگر.....               | ۷۴ |
| معاینه.....                               | ۷۵ |
| مصاحبه.....                               | ۷۷ |
| پرسشنامه.....                             | ۷۷ |
| آنالیز روابط.....                         | ۷۹ |
| دلایل بررسی روابط.....                    | ۷۹ |

آنالیز گذشته نگر روابط ..... ۸۱

آنالیز آینده نگر روابط ..... ۸۱

آنالیز تفاوت ها ..... ۸۴

آنالیز گذشته نگر تفاوت ها ..... ۸۵

آنالیز آینده نگر تفاوت ها ..... ۹۰

خلاصه ..... ۹۱

### فصل سوم ..... ۹۳

طرح های اپیدمیولوژیک ..... ۹۴

پژوهش های مقطعی ..... ۹۶

طرح پژوهش مقطعی مکرر یا سریال ..... ۱۰۰

حجم نمونه ..... ۱۰۱

خلاصه ی ویژگی های پژوهش های مقطعی ..... ۱۰۱

عدم مزایای پژوهش های مقطعی ..... ۱۰۲

پژوهش های کیس-کنترل ..... ۱۰۲

مراحل پژوهش مورد-شاهدی ..... ۱۰۳

مزایای طرح کیس-کنترل ..... ۱۰۶

عدم مزایای طرح کیس-کنترل ..... ۱۰۶

پژوهش های کوهورت ..... ۱۰۷

مراحل پژوهش کوهورت ..... ۱۱۰

|                                      |     |
|--------------------------------------|-----|
| خلاصه.....                           | ۱۱۳ |
| فصل چهارم.....                       | ۱۱۵ |
| تعاریف اندازه گیری.....              | ۱۱۶ |
| ویژگی های متغیر.....                 | ۱۱۷ |
| مقیاس های اندازه گیری.....           | ۱۱۸ |
| مقیاس اسمی.....                      | ۱۱۸ |
| نمونه هایی از مقیاس اسمی.....        | ۱۱۹ |
| مقیاس رتبه ای.....                   | ۱۱۹ |
| نمونه هایی از مقیاس های رتبه ای..... | ۱۲۰ |
| مقیاس های فاصله ای.....              | ۱۲۱ |
| مقیاس های نسبتی.....                 | ۱۲۳ |
| تعیین مقیاس اندازه گیری.....         | ۱۲۵ |
| انواع متغیرها.....                   | ۱۲۵ |
| چهارچوب های اندازه گیری.....         | ۱۲۶ |
| پایایی اندازه گیری.....              | ۱۲۷ |
| تنوری های پایایی.....                | ۱۲۸ |
| اجزای پایایی.....                    | ۱۳۱ |
| پایایی دستگاه.....                   | ۱۳۱ |
| اندازه گیری بیوفیزیولوژیکی.....      | ۱۳۲ |

|                                  |     |
|----------------------------------|-----|
| اندازه گیری های خود اظهاری ..... | ۱۳۳ |
| اندازه گیری های مشاهده ای .....  | ۱۳۵ |
| پایایی ایترا ریتر .....          | ۱۳۶ |
| پایایی ایتتر ریتر .....          | ۱۳۷ |
| پایایی درون فردی .....           | ۱۳۹ |
| سنجش پایایی .....                | ۱۴۰ |
| پایایی نسبی .....                | ۱۴۰ |
| مشکلات ضرایب همبستگی .....       | ۱۴۱ |
| ضریب ICC .....                   | ۱۴۵ |
| پایایی مطلق .....                | ۱۴۷ |
| اعتبار اندازه گیری .....         | ۱۴۹ |
| اعتبار سازه .....                | ۱۵۰ |
| اعتبار محتوی .....               | ۱۵۱ |
| اعتبار ملاکی .....               | ۱۵۲ |
| اعتبار ملاکی همزمان .....        | ۱۵۴ |
| اعتبار پیشینی کننده .....        | ۱۵۴ |
| پاسخ دهی به تغییرات .....        | ۱۵۵ |
| خلاصه .....                      | ۱۵۸ |
| منابع .....                      | ۱۵۹ |



## فهرست شکل ها

- شکل ۱-۱: انواع طرح های پژوهش تجربی، تک عاملی، چند عاملی، کارآزمایی کنترل شده تصادفی..... ۲۲
- شکل ۲-۱: پژوهش کنترل شده تصادفی، استاندارد طلایی در پژوهش های بالینی است. ۲۶.....
- شکل ۳-۱: طرح یک گروه پیش آزمون-پس آزمون برای اثر امواج ماوراء صوت پیوسته بر رینوسینوزیت مزمن..... ۳۲
- شکل ۴-۱: طرح تک موردی با اندازه گیری های متعدد قبل از درمان، حین درمان، و پس از درمان..... ۳۷
- شکل ۵-۱: طرح سری زمانی با ۴ اندازه گیری قبل و بعد از درمان..... ۳۷
- شکل ۶-۱: انتخاب ترتیب درمان ها برای چهار سطح متغیر مستقل..... ۴۱
- شکل ۷-۱: طرح درمان مکرر با یک گروه و دوره پاک شدگی..... ۴۳
- شکل ۸-۱: طرح کارآزمایی کنترل شده تصادفی با طرح متقاطع..... ۴۴
- شکل ۹-۱ الف: طرح فاکتوریال  $2 \times 3$ ..... ۵۰
- شکل ۹-۱ ب: خطوط موازی بیانگر عدم تعامل بین درمانگر و نوع درمان است. ۵۱
- شکل ۱۰-۱ الف: نمونه طرح فاکتوریال چند عاملی با تعامل بین متغیرها..... ۵۲
- شکل ۱۰-۱ ب: خطوط غیرموازی بیانگر تعامل بین درمانگر و نوع درمان است. ۵۳
- شکل ۱۱-۱: دیاگرام شماتیک طرح سه فاکتوری آشیانه ای..... ۵۴

شکل ۱-۱۲: طرح بلوک تصادفی. اعداد، تعداد افراد زن و مرد را در گروه ها نشان می دهد. ۵۷.....

شکل ۲-۱: طرح پژوهش تجربی و طرح پژوهش غیرتجربی ۶۶.....

شکل ۲-۲: ماتریکس انواع پژوهش که سه بعد طرح پژوهش را نشان می دهد: هدف پژوهش (ردیف ها)، زمان جمع آوری داده ها (ستون ها)، و مانیپولاسیون (سلول ها). ۶۹.....

شکل ۲-۳: اهداف پژوهش ۷۰.....

شکل ۲-۴: معاینه کننده دستش را روی یکی از دست های فرد نگه می دارد. ۷۶.....

شکل ۲-۵: پژوهش توصیفی نرماتو برای تعیین نیروی عضلات اندام، ۸۳.....

شکل ۲-۶: طرح مورد-شاهد (کیس-کنترل) ۸۷.....

شکل ۲-۷: طرح کوهورت ۸۸.....

شکل ۳-۱: طرح های اپیدمیولوژیک با زمان اندازه گیری ها و همچنین جهت پیشرفت طرح مشخص می شوند. ۹۵.....

شکل ۳-۲: تریمینولوژی مورد استفاده در پژوهش های اپیدمیولوژیکی ۹۶.....

شکل ۳-۳: طرح پژوهش مقطعی ۹۸.....

شکل ۳-۴: پژوهش مقطعی برای بررسی رابطه بین پیامد و علل ۹۸.....

شکل ۳-۵: طرح پژوهشی کیس-کنترل ۱۰۴.....

شکل ۳-۶: طرح پژوهش کیس-کنترل برای بررسی علت سرطان ریه ۱۰۵.....

شکل ۳-۷: نسبت شانس ابتلا به سرطان ریه. ۱۰۷.....

شکل ۳-۸: طرح پژوهش کوهورت ۱۰۸.....

- شکل ۳-۹: طرح پژوهش کوهورت در مقایسه با طرح کیس-کنترل..... ۱۱۱
- شکل ۳-۱۰: طرح پژوهش کوهورت..... ۱۱۲
- شکل ۴-۱: مقیاس اسمی ضعیف ترین و مقیاس نسبتی قویترین مقیاس های اندازه گیری هستند..... ۱۲۴
- شکل ۴-۲: مقیاس رتبه ای برای "تندی"..... ۱۲۴
- شکل ۴-۳: هر اندازه گیری شامل یک جزء واقعی و یک جزء خطا است..... ۱۲۹
- شکل ۴-۴: واریاسیون در اندازه گیری وضعیت به جلو آمده سر می تواند ناشی از خطای اندازه گیری (تئوری کلاسیک) یا عوامل مختلف (تئوری تعمیم پذیری) باشد..... ۱۳۰
- شکل ۴-۵: پایایی آزمون-بازآزمون با دو اندازه گیری در دو زمان مختلف..... ۱۳۳
- شکل ۴-۶: دینامومتر و وضعیت تست..... ۱۳۴
- شکل ۴-۷: نسخه فارسی پرسشنامه شاخص درجه بندی عملکردی..... ۱۳۵
- شکل ۴-۸: پایایی اینتراریتر، همبستگی بین نمرات یک ریتر که در دو زمان مختلف اندازه گیری شده..... ۱۳۷
- شکل ۴-۹: پایایی اینتراریتر، همبستگی بین نمرات دو ریتر اندازه گیری شده در یک زمان..... ۱۳۸
- شکل ۴-۱۰: منابع خطا در پایایی آزمون-بازآزمون..... ۱۴۰
- شکل ۴-۱۱: نمونه ای از تست Infant Motor Performance در ارزیابی اعتبار محتوی اکولوژیکی..... ۱۵۳



## فهرست جدول ها

---

- جدول ۱-۱: داده های فرضی برای طرح مرکب بررسی اثر دارو بر افسردگی، با در نظر گرفتن "زمان" و "دارو" به عنوان متغیرهای مستقل و نمره افسردگی به عنوان متغیر وابسته. .... ۶۰
- جدول ۱-۲: انواع پژوهش های غیر تجربی ..... ۶۷
- جدول ۱-۴: تست های آماری مورد استفاده در مقیاس های اسمی،  
رتبه ای، فاصله ای، و نسبتی ..... ۱۲۳
- جدول ۲-۴: ضرایب همبستگی ..... ۱۴۱
- جدول ۳-۴: داده های اندازه گیری مکرر حداکثر فشار دمی با پایایی نسبی خوب. ۱۴۳
- جدول ۴-۴: داده های اندازه گیری مکرر حداکثر فشار دمی با پایایی نسبی ضعیف. .... ۱۴۴
- جدول ۵-۴: داده های اندازه گیری مکرر حداکثر فشار دمی با ضریب همبستگی پیرسون بالا، اما پایایی مطلق ضعیف. .... ۱۴۶



## فصل اول

### طرح های پژوهش تجربی

بخش های فصل

طرح های گروهی

فرضیات طرح های گروهی

کارآزمایی کنترل شده تصادفی

طرح های تجربی تک عاملی (طرح گروه-کنترل پیش آزمون-پس آزمون، طرح گروه-

کنترل فقط پس آزمون، طرح گروه-کنترل نابرابر، طرح سری زمانی، طرح های اندازه

گیری مکرر یا درمان مکرر)

طرح های تجربی چند عاملی (طرح های فاکتوریال در مقابل آشیانه ای، طرح های کاملاً

تصادفی در مقابل بلوک-تصادفی، طرح های بین گروهی، درون گروهی، و مرکب)

خلاصه

## طرح های گروهی

پژوهش های تجربی، با دستکاری کنترل شده یک متغیر مستقل فعال یا چند متغیر، توسط پژوهشگران مشخص می شوند. این دستکاری کنترل شده می تواند در گروه های معینی از افراد انجام شود. مفهوم "گروه" اشاره به ویژگیهای های افراد دارد، نه این که درمان چگونه اعمال شده است.

در ادامه، پژوهش کنترل شده تصادفی و سپس چندین طرح پژوهش تک عاملی (Single-factor) و چند عاملی (Multiple-factor) توضیح داده می شود (شکل ۱-۱). باید توجه داشت که "تجربه" نیازمند حداقل یک متغیر مستقل است یعنی تغییری که پژوهشگر می تواند کنترل کند.

یک استراتژی پژوهشی بسیار رایج، بررسی مقادیر متغیر وابسته در گروه های غیرقابل دستکاری مثل جمعیت های بیمار نسبت به سالم، یا جمعیت های جوان نسبت به سالمند، یا مقایسه گروه های مختلف قومی می باشد. در این پژوهشها، چون پژوهشگر نمی تواند طبقات را دستکاری کند (برای مثال نمی تواند فردی که به علت سکتة مغزی دچار اختلال زبانی یا آفازی شده است را در گروه غیرآفازیک قرار دهد یا به عکس)، پژوهش از نوع تجربی نیست، بلکه پژوهشهایی برای بررسی همبستگی هستند و روابط علت و معلولی را نشان نمیدهند.

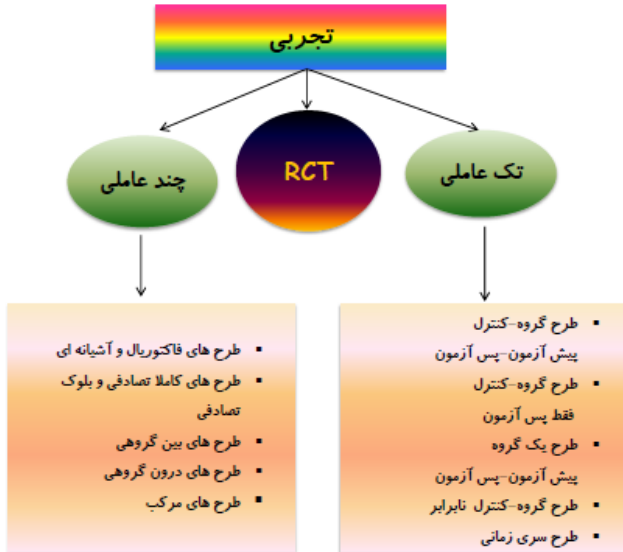
در این فصل، ترمینولوژی و اطلاعات لازم برای درک ساختار بندی طرح های تجربی گروهی توضیح داده می شود. برای طراحی پژوهش های معتبر، بجز تقسیم افراد به گروه ها، اندازه گیری، و مداخله، موارد زیاد دیگری باید در نظر گرفته شوند. روش های کنترل عوامل مخدوش کننده و تقسیم افراد در گروه ها باید به دقت انجام شوند تا عوامل تهدید کننده اعتبار پژوهش کنترل گردند. ابزارهای آماری نیز باید بطور صحیح استفاده شوند. سرانجام، گزارش پژوهش که به صورت انتشار مقاله علمی است، باید اطلاعات مورد نیاز برای ارزیابی کیفیت پژوهش را در بر داشته باشد.

در یک تقسیم بندی کلی، طرح پژوهش به دو دسته "تجربی" و "مشاهده ای" (غیر تجربی) تقسیم می شود. طرح پژوهش تجربی شامل پژوهش کنترل شده تصادفی است. طرح مطالعه مشاهده ای شامل کیس سری (Case series)، مقطعی، مورد-شاهدی (Case-control) و کوهورت است.

انتخاب طرح پژوهش، به سوال پژوهش بستگی دارد و باید در راستای هدف و فرضیه پژوهش باشد. هدف، طراحی پژوهش های معتبر است. پژوهش تجربی، همان پژوهش علمی کلاسیک، و آشنا ترین نوع طرح پژوهش است. طرح پژوهش تجربی یک رویکرد علمی به پژوهش است که در آن یک یا چند متغیر مستقل دستکاری گردیده و اثر آن بر متغیر وابسته بررسی می شود.

پژوهش باید در یک مدت زمان معین انجام شود تا نتایج معقولی برای رابطه بین دو نوع متغیر (مستقل و وابسته) ثبت گردد.

پژوهش تجربی، ممکن است که گروهی یا فردی باشد. در پژوهش گروهی، داده های افرادی که درمان (متغیر مستقل) را دریافت کرده اند، "گروه بندی" شده و آنالیز میشوند. در پژوهش های تجربی بر اندازه گیری تاکید می شود.



شکل ۱-۱: انواع طرح های پژوهش تجربی، تک عاملی، چند عاملی، کارآزمایی کنترل شده تصادفی

طرح پژوهش تجربی نیازمند حداقل یک متغیر مستقل فعال است یعنی متغیری که پژوهشگر می تواند آن را کنترل کند. در طرح پژوهش تجربی، دستکاری کنترل شده یک یا چند متغیر مستقل توسط پژوهشگر و در افراد یا گروهی از افراد انجام می شود. در طرح های تجربی گروهی، میانگین داده های گروه ها از طریق آمار استنباطی مقایسه میشود، زیرا تاکید بر داده های گروه ها است و فرض این است که تفاوت معنی دار آماری، نشانه تفاوت واقعی بین گروه ها است. در پژوهش های تجربی گروهی، در صورت انتخاب تصادفی و تقسیم تصادفی افراد در گروه ها، نتایج پژوهش می تواند به کل جمعیت تعمیم یابد.

یکی از نگرانی های اصلی در پژوهش های تجربی و به طور مشخص در انسان، تغییر پذیری نتایج است. در پژوهش های انسانی، همواره نتایج مشابه بدست نخواهید آورد. تغییرپذیری نتایج در پژوهش های انسانی، علت های مختلفی دارد. در طرح های تجربی

گروهی فرض بر این است که بجز خطای اندازه گیری، تغییرپذیری بیشتر ناشی از ویژگیهای خود انسان ها (مانند سن، هوش، انگیزه، وضعیت سلامتی و...) است، تا عوامل خارج از نمونه های پژوهش. البته کنترل همه فاکتورها مشکل است و پژوهشگر ممکن است نتواند همه این فاکتورها را به خوبی کنترل کند. یکی از دلایل این که پژوهشگران باید میانگین گروه ها را مقایسه کنند، همین است. در طرح پژوهش تجربی تک عاملی، پژوهشگر فقط یک متغیر را دستکاری می کند. اما در طرح های چندعاملی، پژوهشگر اثر بیش از یک متغیر را بررسی می کند و به تعاملات بین متغیرها و اثر توأم چند متغیر علاقه مند است.

## فرضیات طرح های گروهی

یکی از نگرانی های اصلی در پژوهش های تجربی و به طور مشخص در انسان ها، تغییرپذیری است که در داده ها نمایان می شود. پژوهشگران گزارشی نخواهند یافت که در آن همه نمونه های انسانی نتایج یکسانی تولید کرده باشند. علل این تغییرپذیری چیست؟ یک علت این است که کنترل همه عوامل ممکن نیست و این امکان هست که نتایج، حاصل چیزی غیر از دستکاری متغیرهای مستقل باشد. جدا از خطای تجربی، لازم است ما فرضیاتی در باره منابع تغییرپذیری هم ارائه کنیم.

در طرح های تجربی گروهی، فرض بر این است که تغییرپذیری مربوط به نمونه های انسانی است نه منابع خارج از نمونه های پژوهش. این عوامل برای نمونه می توانند شامل: اختلافات سنی، هوش، انگیزه، وضعیت سلامتی، و غیره باشند. نکته مهم این است که اینها عواملی هستند که پژوهشگران نمی توانند آنها را بخوبی کنترل کنند. این امر موجب شده است که اگر پژوهشگری تعداد نمونه کافی در پژوهش داشته باشد و این نمونه ها بطور تصادفی از جمعیت انتخاب شده و بطور تصادفی در گروه های پژوهش تقسیم شوند، همه ی عوامل داخلی بطور متعادل بین گروه های تجربی و کنترل توزیع

خواهند شد، و به این ترتیب فقط اثرات درمان یا متغیرمستقل (بجز خطاهای تجربی) باقی می ماند که اندازه گیری می شود.

با استفاده از این فرآیند، روشن می شود که به علت تاکید بر داده های هریک از گروه ها در یک پژوهش، پژوهشگران باید میانگین داده های گروه ها را مقایسه کنند، و تفاوت بین میانگین گروه ها با استفاده از آمار استنباطی مشخص می شود. فرض پایه این است که با احتمال قابل قبولی، تفاوت های معنی دار آماری، نمایانگر تفاوت های واقعی بین گروه ها است باید ذکر کرد که حتی در مواردی که اثر درمان کم است هم ممکن است تغییرات از لحاظ آماری معنی دار باشد.

فرض دیگر این است که ممکن است. نمونه های واقعا تصادفی از یک جمعیت داشت (که نمایانگر جمعیت مورد نظر باشند) و سعی کنیم افراد را بطور واقعا تصادفی در گروه ها تقسیم کنیم (برای اطمینان از برابری پیش از شروع پژوهش گروه ها، در متغیرهای وابسته).

فرضیه نهایی طرح های تجربی گروهی مبتنی بر انتخاب و تقسیم تصادفی افراد، این است که نتایج یک پژوهش تجربی می تواند به جمعیت بطور کلی تعمیم یابد.

## کارآزمایی کنترل شده تصادفی

قبل از توضیح طرح پژوهش کنترل شده تصادفی باید ذکر کرد که پارادایم های پژوهشی مختلف (پارادایم های کمی، کیفی، تک-سیستمی)، اهداف پژوهشی مختلفی دارند (توصیفی، آنالیز روابط، و آنالیز تفاوت ها)، و طرح های تجربی و غیرتجربی ارزشمند هستند. ضمن تایید ارزش انواع مختلف پژوهش، توافق قوی وجود دارد که نوعی از پژوهش یعنی کارآزمایی کنترل شده تصادفی یا کارآزمایی کلینیکی تصادفی (Randomized Controlled/Clinical Trial, RCT) میتواند شواهد محکمی برای یک هدف پژوهش مشخص، یعنی تعیین اثربخشی مداخلات کلینیکی فراهم کند.

طرح های کنترل شده تصادفی، تجربی و آینده نگر هستند، به این معنی که درمان (متغیر مستقل) توسط پژوهشگر و به طور کنترل شده اعمال می شود و متغیرهای وابسته (پیامدها) تحت شرایط کنترل شده جمع آوری می شوند. در این طرحها، متغیرمستقل حداقل دو سطح دارد: ۱- گروه درمان، ۲- گروه کنترل برای مقایسه. جزء "تصادفی" عبارت RCT ، به تقسیم تصادفی مشارکت کنندگان در گروه درمان و گروه کنترل اشاره دارد (شکل ۲-۱). کارآزمایی کنترل شده تصادفی شکل های مختلفی دارد، اما همه آنها قطعا ویژگی های فوق را دارا می باشند. بنابراین، طبق تعریف، همه RCT ها طرح های تجربی هستند اما همه طرح های تجربی، RCT نیستند، زیرا در آن راه های غیر تصادفی زیادی برای قرار دادن افراد در گروه ها وجود دارد. گرچه RCT ها شواهدی قوی در باره اثرات یک مداخله فراهم می کنند، اما برای پاسخ دادن به همه سوالات مناسب نیستند و همچنین برای هر موضوعی امکان انجام RCT نیست. برای مثال، اگر پژوهشگری بخواهد شواهدی قوی در باره نقش سیگار در سرطان ریه فراهم کند، باید یک RCT طراحی کند و گروه سیگاری را با یک گروه غیرسیگاری مقایسه کند. حال با توجه به شواهد قوی حاصل از کارآزمایی های غیرتصادفی، آیا اخلاقی است که گروه بزرگی از نوجوانان غیرسیگاری را به گروه های سیگاری و غیرسیگاری تقسیم کرد و آنها را در طی زندگیشان پیگیری کرد تا نرخ نسبی رخداد سرطان ریه را تعیین کرد؟ قطعا اینطور نیست. پژوهش ها در زمینه اثربخشی مداخلات باید به صورت RCT های با کیفیت انجام شوند.



شکل ۱-۲: پژوهش کنترل شده تصادفی، استاندارد طلایی در پژوهش های بالینی است.

در ساختن داروهای جدید، پژوهشگران سکانس معینی از کارآزمایی ها را در سه فاز I, II و III بر روی انسان ها انجام می دهند. فاز I بررسی دارو، معمولا در گروهی از داوطلبان سالم انجام می شود که شروع کارآزمایی برای اثبات ایمنی داروی جدید است که با دوزهای مختلف در انسان ها انجام می شود. فاز II معمولا در گروه کوچکی از بیماران که انتظار می رود دارو برای آنها مفید باشد انجام می شود تا اثربخشی دارو و عوارض کوتاه مدت و خطرات آن در بیماران مشخص، تعیین شود. بسته به نوع پیامد مورد انتظار از دارو، فاز II ممکن است غیرتصادفی (در صورت عینی بودن پیامدها) یا تصادفی (در صورت ذهنی بودن و بیشتر تحت تاثیر پلاسبو بودن پیامدها) باشد. تنها پس از تایید ایمن بودن دارو در فاز I و اثرات درمانی نویدبخش در فاز II، بررسی داروها وارد فاز III و پژوهش RCT می شوند. در این RCT ها، داروی جدید با یک درمان پلاسبو یا با یک درمان استاندارد پذیرفته شده مقایسه می شود.

در حیطه های غیردارویی از جمله پژوهش های توانبخشی، این روند استاندارد با استفاده از طرح های پژوهشی پیشرونده قوی وجود ندارد. گرچه، پیروی از چنین الگویی برای بررسی بسیاری از مداخلات توانبخشی نیز مناسب است و چنین نمونه ای از طرح پژوهشی،

با الگو گرفتن از فازهای مختلف بررسی داروها، در استفاده از تکنیک تمرین روی تردمیل با حمایت وزن بدن (body-weight-supported treadmill training) در بیماران دچار ناتوانی در راه رفتن مشاهده شده است. پژوهشگران ابتدا نتایج بررسی مردان بدون ناتوانی در راه رفتن (مشابه با فاز I بررسی دارویی) و سپس نتایج کارآزمایی غیر تصادفی بر روی تعداد کمی از افراد مبتلا به راه رفتن اسپاستیک را منتشر نموده و اثر درمان با یا بدون حمایت وزن بدن در الگوهای راه رفتن را تعیین کردند (مشابه با فاز II بررسی داروها). بعداً مشابه با کارآزمایی های فاز III، نتایج پژوهش های RCT برای مقایسه تمرین روی تردمیل و حمایت وزن بدن، با تمرینات سنتی درمان راه رفتن منتشر گردید. در نهایت، مقالات مروری و شواهد در باره اثربخشی تمرین با تردمیل و حمایت وزن بدن منتشر شد.

پژوهشگران باید از RCT هایی با کیفیت بالا برای بررسی اثر درمان ها استفاده کنند. گرچه پژوهش هایی که به صورت RCT انجام نشده اند نیز نباید فقط به این دلیل که طرح مطالعه ضعیف تر یا متفاوتی دارند، نادیده گرفته شوند. این نوع پژوهش ها باید در پرتو مقالات منتشر شده مربوطه، از لحاظ اعتبار پژوهش یا به عنوان پایه ای برای انجام RCT ها در آینده مورد ارزیابی قرار بگیرند.

## طرح های تجربی تک عاملی

طرح های تجربی تک عاملی یک متغیر مستقل دارند و گرچه استفاده از طرح های تجربی چند عاملی که بیش از یک متغیر مستقل دارند به طور فزاینده ای رایج شده است، ولی طرح های تک عاملی بخش مهمی از پژوهشها را شامل می شوند. در این بخش، تعدادی از طرح های تک عاملی رایج در پژوهش توضیح داده می شوند.

انواع طرح های تجربی تک عاملی (Single-Factor Experimental Design) عبارتند از:

- طرح گروه-کنترل پیش آزمون-پس آزمون  
Pretest-Posttest- Control-Group design
- طرح گروه-کنترل فقط پس آزمون  
Posttest-Only Control-Group Design
- طرح یک گروهی پیش آزمون-پس آزمون  
Single-Group Pretest-Posttest Design
- طرح گروه-کنترل نابرابر  
Nonequivalent Control-Group Design
- طرح سری زمانی  
Time series design
- طرح های اندازه گیری مکرر یا درمان مکرر  
Repeated treatment design

### طرح گروه-کنترل پیش آزمون-پس آزمون

طرح "نک عاملی گروه-کنترل پیش آزمون-پس آزمون" یکی از طرح های کلاسیک RCT است. در این طرح حداقل دو گروه وجود دارد که به طور تصادفی از جمعیت مورد نظر انتخاب شده و به طور تصادفی در دو گروه تقسیم می شوند. وقتی که دو گروه وجود دارد، یک گروه درمان (متغیر مستقل) را دریافت می کند و گروه دیگر (کنترل) ممکن است درمانی دریافت نکرده و یا پلاسبو دریافت کنند. اندازه گیری ها یک بار قبل از درمان انجام می شوند (پیش آزمون) و پس از پایان درمان، برای بررسی تاثیر درمان روی متغیر (های) وابسته، اندازه گیری ها مجدداً انجام می شوند (پس آزمون).

در زیر نماد طرح گروه-کنترل پیش آزمون-پس آزمون نشان داده شده است. R بیانگر تقسیم تصادفی افراد در دو گروه، O بیانگر مشاهده (اندازه گیری)، و X بیانگر مداخله (دستکاری یا مانیپولاسیون) می باشد. جای خالی در ردیف دوم، بیانگر گروه کنترل است که درمانی دریافت نمی کنند. بنابراین، مداخله و کنترل دو سطح از متغیر مستقل را تشکیل می دهند.



اصطلاحات دیگر معادل این طرح عبارتند از طرح بین گروهی (Between-groups design) یا طرح بین افراد (Between-subjects design) زیرا تفاوت های مورد نظر، تفاوت های بین گروه ها است. این نوع طرح، همچنین ممکن است به عنوان طرح گروه موازی (Parallel group design) نامیده شود که در آن هر گروه فقط یکی از سطوح متغیر مستقل را دریافت می کند. گروه کنترل در این طرح، گروه کنترل "غیر فعال" (Passive) نامیده می شود زیرا درمانی دریافت نمی کند یا در صورت درمان، درمان شم یا پلاسبو دریافت می کنند.

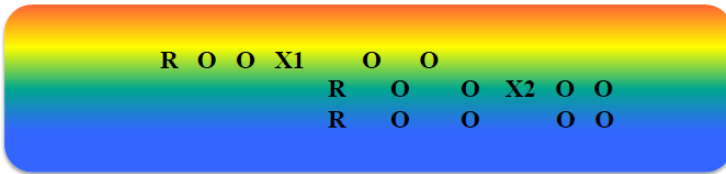
یک نمونه: در یک پژوهش، پژوهشگران اثر یونتوفورزیس کورتیزون بر درمان آرنج تنیس بازان (اپیکوندیلیت طرفی) را با استفاده از طرح گروه-کنترل پیش آزمون-پس از آزمون، بررسی کردند. گروه درمان، یونتوفورزیس کورتیزون را دریافت کردند و گروه کنترل غیرفعال، یونتوفورزیس شم دریافت کردند. اندازه گیری ها چهار بار انجام شدند (قبل از درمان، بلافاصله پس از پایان درمان، و از آنجایی که پژوهشگران مایل به بررسی ماندگاری اثرات درمان بودند، اندازه گیری ها را ۳ و ۶ ماه پس از درمان نیز تکرار کردند). سپس داده های جمع آوری شده از دو گروه در قبل از درمان و پس از درمان با استفاده از تست های آماری آنالیز گردید تا تغییرات در هر دو گروه و اختلاف تغییرات بین دو گروه روشن شود.

در پژوهش های کلینیکی، طرح گروه-کنترل پیش آزمون - پس آزمون، ممکن است به شکل دیگری بکار گرفته شود. در این شکل از RCT، پژوهشگر به دلیل مسائل اخلاقی ممکن است به گروه مقایسه (کنترل)، مداخله یا درمان استاندارد که اثر بخشی آن معلوم شده است را ارائه دهد و اثر درمان جدید را با آن مقایسه کند که در این صورت گروه کنترل، گروه "فعال" نامیده می شود. نماد این طرح در زیر نشان داده شده است:



برای مثال، پژوهشگران با استفاده از طرح گروه-کنترل پیش آزمون-پس آزمون، اثر تمرین درمانی با محدودیت (Constraint-induced therapy, CIT)، در بیماران مبتلا به سکته مغزی مزمن دچار آفازی را با یک گروه کنترل فعال مقایسه کردند. گروه درمان، ۱۰ روز تمرین درمانی با محدودیت (X1) را دریافت کردند، و گروه کنترل فعال نیز ۴ هفته درمان مرسوم را دریافت کردند. نتیجه این پژوهش بیانگر اثر CIT در بهبود آفازی بیماران در مقایسه با گروه کنترل بود.

طرح گروه-کنترل پیش آزمون-پس آزمون، می تواند بیش از دو گروه و دو اندازه گیری هم داشته باشد. در زیر نماد کلی این نوع طرح با سه گروه (یک گروه درمان، یک گروه کنترل فعال، یک گروه کنترل غیرفعال) و چهار اندازه گیری نشان داده شده است.



### طرح گروه-کنترل فقط پس آزمون

پژوهشگران وقتی از این طرح استفاده می کنند که قادر به اندازه گیری های پیش آزمون نیستند. بنابراین، قضاوت در باره اثر متغیر مستقل بر متغیر وابسته، تنها بر پایه اندازه گیری های پس آزمون می باشد. مثل طرحهای گروه-کنترل پیش آزمون-پس آزمون، این نوع طرح نیز یک RCT و طرحی بین گروهی است. نماد این طرح در زیر نشان داده شده است.



در یک پژوهش که از این مدل طرح استفاده شده بود، اثرات مکمل کربوهیدرات-پروتئین در مقایسه با پلاسبو بر روی انجام تمرین مقاومتی مردان سالم جوان بررسی شد. این پژوهش دوسوکور بود به این معنی که نه نمونه ها و نه ارزیاب ها، از نوع درمان هر فرد اطلاع نداشتند. در این پژوهش، ۳۴ مرد جوان به طور تصادفی به دو گروه تجربی (دریافت کننده مکمل کربوهیدرات-پروتئین) و پلاسبو تقسیم شدند و از آنها خواسته شد که یک ست تمرین، شامل هشت تمرین مقاومتی را انجام دهند. نتایج این پژوهش نشان داد که اثر تمرینات، بین دو گروه تفاوت معنی داری نداشت، گرچه پژوهشگران بر پایه معیارهای میوگلوبین و کراتین کیناز نتیجه گیری کردند که مکمل کربوهیدرات-پروتئین در مقایسه با پلاسبو، آسیب عضلانی را به میزان بیشتری مهار می کند.

فرض کنید ورزشکاری دچار پارگی رباط صلیبی قدامی زانو بشود. در این شرایط، امکان انجام یکسری مداخلات قبل از جراحی و اندازه گیری های پیش آزمون نیست. بنابراین پژوهشگر این امر را به پس از عمل جراحی موکول می کند. در این پژوهش فرضی، افراد بطور تصادفی در دو گروه قرار گرفته و گروهی مداخله به صورت یک روش جراحی را دریافت کرده و گروهی دیگر با روش دیگری جراحی می شوند. بنابراین، اثر مداخله (متغیر مستقل) بر پیامد (متغیر وابسته) تنها بر پایه اندازه گیری های پس آزمون می باشد. مزایای این طرح پس آزمون و بین گروهی این است که پژوهشگر می تواند اثرات مداخله را در دوره حاد بیماری نیز بسنجد.

## طرح یک گروه پیش آزمون - پس آزمون

نماد طرح یک گروه پیش آزمون-پس آزمون، در زیر نشان داده شده است.



بر خلاف طرح گروه-کنترل پیش آزمون-پس آزمون و طرح گروه-کنترل فقط پس آزمون، طرح تک گروهی، نه RCT است و نه طرح بین گروهی. چون فقط یک گروه وجود دارد، فرصتی برای تقسیم تصادفی افراد به گروه ها و همچنین امکان انجام مقایسه بین گروهی وجود ندارد. بلکه، مقایسات در یک گروه تجربی منفرد انجام می شود، و بنابراین طرح "یک گروه پیش آزمون-پس آزمون" یک طرح درون گروهی (Within-group) است.

در یک پژوهش کلینیکی برای بررسی اثر امواج ماوراء صوت پیوسته بر رینوسینوزیت مزمن، یک گروه از بیماران شامل ۳۰ نفر تحت درمان برای مدت ۱۰ جلسه، ۳ روز در هفته قرار گرفتند و قبل از درمان، بعد از درمان، و یک ماه بعد از اتمام درمان معاینه شدند (شکل ۱-۳). در این پژوهش که متغیر وابسته (معیار پیامد) "درصد بهبودی" بود، از طرح یک گروه پیش آزمون-پس آزمون با سه اندازه گیری استفاده شد (یک اندازه گیری قبل از درمان و دو اندازه گیری پس از درمان) که نماد طرح در زیر نشان داده شده است:



نتایج این پژوهش نشان داد که پس از درمان، امواج ماوراء صوت باعث بهبودی معنی دار در علائم بیماران شد (میانگین ۷۴/۴٪) و اثرات مفید درمان از لحاظ عدم رخداد بیماری تا یک ماه پس از درمان، در ۷۲٪ بیماران باقی ماند.



شکل ۱-۳: طرح یک گروه پیش آزمون-پس آزمون برای اثر امواج ماوراء صوت

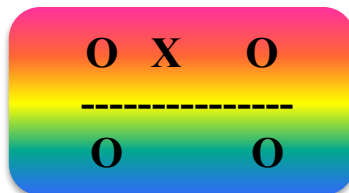
پیوسته بر رینوسینوزیت مزمن

در پژوهش دیگری، اثر کوتاه مدت امواج ماوراء صوت منقطع بر رینوسینوزیت مزمن مورد بررسی قرار گرفت. یک گروه از بیماران شامل ۲۲ نفر تحت درمان برای ۶ جلسه، ۲ تا ۳ روز در هفته قرار گرفتند و معاینات قبل از درمان، قبل از جلسه ۴، و بعد از پایان جلسه ۶ انجام شد. از آنجایی که تنها یک گروه در این پژوهش وجود داشت، بنابراین، از طرح یک گروه پیش آزمون-پس آزمون استفاده شده است.

در پژوهش دیگری، اثرات کارددرمانی در منزل بر روی یک گروه شامل ۲۰ کودک دچار فلج مغزی همی پلژی اسپاستیک بررسی شد. اندازه گیری ها قبل و بعد از درمان انجام شد. نتایج این پژوهش نشان دهنده تغییرات معنی دار در متغیرهای وابسته پس از درمان بود. طرح مورد استفاده در این پژوهش، طرح یک گروه پیش آزمون-پس آزمون (Single-group pretest-posttest design study) بود.

## طرح گروه-کنترل نابرابر

طرح گروه-کنترل نابرابر، وقتی استفاده می شود که یک گروه کنترل غیرتصادفی برای مقایسه وجود دارد. نماد طرح در زیر نشان داده شده است:



در نماد بالا، خط منقطع در بین گروه ها بیانگر این است که افراد بطور تصادفی در گروه ها تقسیم نشده اند. بنابراین، این طرح یک طرح کنترل شده تصادفی نیست. اما یک طرح بین گروهی است (Between-group design) زیرا مقایسه بین دو گروهی است که بطور غیرتصادفی انتخاب شدند انجام می شود.

این طرح از لحاظ نوع پژوهش از نوع آر-سی-تی محسوب نمی شود. پژوهشگر در این طرح دو گروه دارد ولی گروهی که به عنوان گروه کنترل شرکت داده می شوند بطور تصادفی انتخاب نشده اند. ممکن است یک گروه کنترل در دسترس باشد که پژوهشگر بتواند گروه اصلی را با آن مقایسه کند و یا بعداً گروه کنترلی را وارد پژوهش نماید. در واقع طرح بین گروهی مقایسه بین گروه هایی هستند که بطور غیرتصادفی انتخاب شده اند.

در یک پژوهش برای بررسی اثر مداخله (شامل آموزش بیماران، بخاطر آوردن و یادآوری) برای فشار خون، با هدف افزایش پایبندی به مصرف داروهای تجویز شده، گروهی از بیماران دچار فشار خون بالا با سن بالاتر از ۶۵ سال و کمتر از ۸۵ سال در دو شهر مداخله و کنترل کشور کره وارد پژوهش شدند. در این پژوهش که از طرح گروه-کنترل نابرابر (Non-equivalent control group design) استفاده شد، پژوهشگران نشان دادند که اعمال مداخله، سالمندان دچار فشار خون بالا را تشویق کرد که مراقبت پیوسته دریافت کنند.

در پژوهش دیگری برای بررسی اثر کاردرمانی (Occupational therapy) بر دست خط کودکان مدرسه، از طرح گروه-کنترل نابرابر استفاده شد. در این پژوهش، پژوهشگران از میان افرادی که برای بهبود مهارت های خوش نویسی به کاردرمانی ارجاع شده بودند، گروهی را شناسایی کردند. سپس در همان نواحی، گروهی از محصلین بدخط که به کاردرمانی ارجاع نشده بودند را به عنوان گروه مقایسه انتخاب کردند. بین دو گروه تفاوت هایی وجود داشت که احتمالاً از تقسیم غیرتصادفی افراد در گروه ها ناشی می شد (چندین کودک در گروه کاردرمانی، گفتاردرمانی و فیزیوتراپی هم دریافت می کردند، ولی گروه کنترل فقط کاردرمانی دریافت می کردند. همچنین نسبت کودکان دچار ناتوانی در یادگیری و اختلال رشدی در گروه کاردرمانی نسبت به گروه کنترل بیشتر بود). رویکرد دیگر برای طرح گروه-کنترل نابرابر، استفاده از کوهورت ها برای گروه های تجربی است. اصطلاح کوهورت عموماً برای هر گروهی استفاده می شود، اما کاربرد

اختصاصی تر کوهورت برای گروه هایی از افراد است که دسترسی به آنها در زمان های متوالی امکان پذیر است (مثل دو گروه دانشجویی در دو ترم متوالی). رویکرد کوهورت اغلب در پژوهش های آموزشی استفاده می شود که به علت رخداد طبیعی کوهورت هایی است که در زمان های متوالی اتفاق می افتند. نمونه ای از این طرح در پژوهشی است که روش های سخنرانی سنتی مورد استفاده در ترم پاییز برای گروهی از دانشجویان با روش های یادگیری فعال مورد استفاده در ترم زمستان در گروه دیگری از دانشجویان برای تدریس فیزیولوژی دانشجویان مقایسه شدند.

## طرح سری زمانی

طرح سری زمانی وقتی مفید است که پژوهشگر قصد دارد اثرات درمان را در طی یک دوره زمانی طولانی بسنجد. پژوهشگر درمان را ادامه می دهد و اثرات آن را چندین بار در طی دوره پژوهش و تجربه اندازه گیری می کند. در طرح سری زمانی، متغیر وابسته چندین بار قبل و بعد از درمان اندازه گیری می شود یعنی پژوهشگر یک سری از مشاهدات متغیر وابسته را قبل و بعد از درمان ثبت می کند. شواهد تاثیر درمان زمانی دیده می شوند که تغییرات معنی دار در داده های سری زمانی در زمان درمان وجود داشته باشد. طرح های سری زمانی برای مشاهده ثبات وضعیت پایه اندازه گیری ها قبل از شروع درمان، در یک گروه یا یک فرد است. وقتی که گروهی برای مقایسه وجود ندارد، باید قبل از اعمال درمان، ثبات یک معیار یا میزان تغییرات آن در طی زمان را ثابت کنیم. نماد طرح سری زمانی در زیر نشان داده شده است:



تعداد اندازه گیری های انجام شده قبل، بعد، یا حتی در حین مداخله، بسته به نوع پژوهش فرق می کند. اگر ارزیابی تغییرات در فواصل زمانی متعدد اهمیت داشته باشد، هر یک از

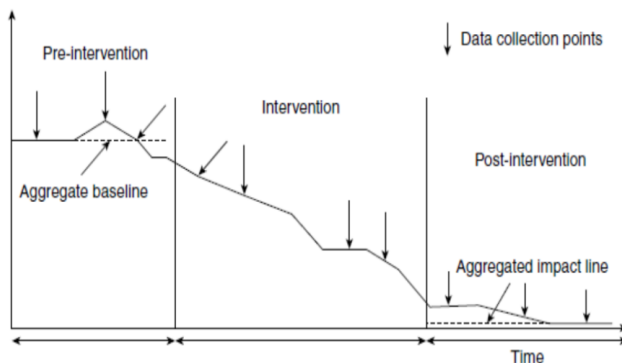
طرح ها می تواند با دو یا چند گروه و با دفعات اندازه گیری بیشتر در قبل و پس از درمان طراحی گردد.

پژوهشگر ممکن است در باره تعداد اندازه گیری های قبل از مداخله و دفعات تکرار آنها درفاصله های زمانی معین بعد از مداخله، تصمیم بگیرد. یکی از دلایل اصلی انجام این نوع پژوهش، آن است که پژوهشگر می خواهد نشان بدهد وضعیت بیماران قبل از مداخله ثابت بوده و اگر مداخله ای انجام می دهد، بتواند تغییرات پس از درمان را به مداخله نسبت دهد و نشان دهد که تغییرات به وجود آمده در آزمودنی ها بر اثر گذشت زمان و خودبخودی نیست. این روش در پژوهش های گروهی نیز قابل انجام است، ولی بیشتر در پژوهش های تک موردی (Single –subject) بکار می رود و در آن حداقل به سه اندازه گیری قبل از مداخله و سه اندازه گیری بعد از مداخله یا درمان نیاز است، فاصله اندازه گیری ها نیز بسته به نوع و هدف پژوهش و شرایط و ویژگی آزمودنی ها متفاوت خواهد بود (شکل ۱-۴). چنانچه وضعیت بیمار قبل از دریافت درمان ثابت نبوده و با نوساناتی همراه باشد، پژوهشگر نمی تواند تغییرات پس از درمان را به مداخله نسبت دهد. در یک پژوهش، از طرح سری زمانی برای بررسی اثر یک برنامه ارتقاء سلامتی و پیشگیری از آسیب، در یک مرکز رشد کودک استفاده کردند. داده ها یک بار در هفته و برای ۴ هفته، قبل و بعد از مداخله جمع آوری شد. انجام چهار اندازه گیری قبل و بعد از مداخله، در مقایسه با یک اندازه گیری قبل و بعد از مداخله، تصویر کامل تری از وضعیت سلامتی کودکان فراهم می کند (شکل ۱-۵).

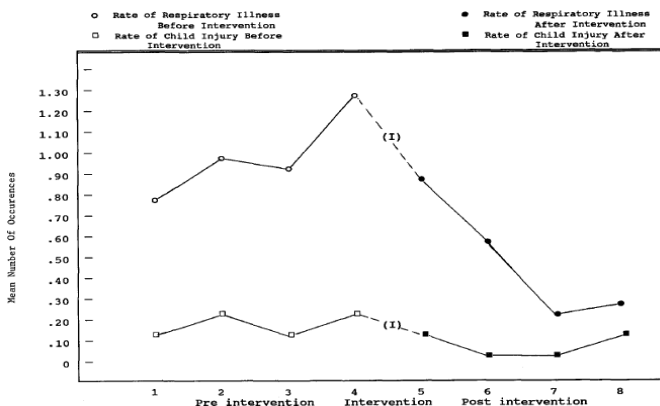
## خلاصه ویژگی های طرح سری های زمانی

- ثبات مقادیر یک یا چند متغیر، قبل از شروع درمان در یک گروه یا یک فرد.
- در نبود گروه کنترل، مستند کردن میزان تغییرات به صورت تابعی از زمان.

- تعداد اندازه گیری ها (قبل، بعد، حتی در حین مداخله) بستگی به نوع پژوهش دارد.
- در رویکردهای  $N=1$  رایج است.



شکل ۱-۴: طرح تک موردی با اندازه گیری های متعدد قبل از درمان، حین درمان، و پس از درمان



شکل ۱-۵: طرح سری زمانی با ۴ اندازه گیری قبل و بعد از درمان

## طرح اندازه گیری های مکرر

طرح های اندازه گیری مکرر بطور گسترده ای در پژوهش های علم سلامت مورد استفاده قرار می گیرد. اصطلاح اندازه گیری مکرر به این معنی است که افراد همه سطوح متغیر مستقل را دریافت کرده و برای بررسی اثر هریک، اندازه گیری می شوند. بنابراین، همه طرح های پیش آزمون-پس آزمون می توانند یک طرح اندازه گیری مکرر در نظر گرفته شوند زیرا هر فرد در هر دو سطح (قبل و بعد از درمان) یک متغیر مستقل اندازه گیری میشوند. معمولاً در طرح های اندازه گیری مکرر، یک گروه وجود دارد و نتایج طی زمان بررسی می شوند. تعداد اندازه گیری ها معمولاً بیش از دو اندازه گیری است. این طرحها همچنین طرح های درون گروهی (Within-group یا Within subjects) نامیده می شوند، زیرا اثر متغیر مستقل در افراد یک گروه بررسی می شود و نه بین گروهها. در طرح اندازه گیری مکرر (از جمله طرح های پیش آزمون-پس آزمون)، اندازه گیری پیامدها برای هر فرد، در همه سطوح متغیر مستقل انجام می شود. در طرح اندازه گیری مکرر معمولی، یک گروه وجود دارد و نتایج در طی زمان اندازه گیری می شود (برای مثال، ۳ اندازه گیری در قبل، بلافاصله بعد، و ۴ هفته بعد، یا ۴ اندازه گیری در قبل، ۱، ۴، و ۸ هفته بعد). در یک پژوهش کلینیکی تک گروهی، پیش آزمون-پس آزمون برای بررسی اثرات امواج ماوراء صوت بر رینوسینوزیت مزمن، از طرح اندازه گیری مکرر استفاده شد (با ۳ اندازه گیری: قبل، بلافاصله بعد از ۱۰ جلسه درمان، و ۱ ماه بعد از اتمام درمان).

طرح های اندازه گیری مکرر که که طرح های درون فردی (Within-subjects) یا گروهی نیز نامیده می شوند برای هر حالت پژوهش، از یک گروه (یعنی درمان) یا کنترل، استفاده می شود. زمانی که پژوهشگر احساس می کند زمان کمی دارد و یا تعداد افراد کمی را می تواند وارد پژوهش کند می تواند از طرح درمان مکرر (Repeated treatment design) استفاده کند و یک گروه را با دو یا سه مداخله مورد بررسی قرار

دهد. اما باید از بعضی عوامل که کارایی اندازه گیری های مکرر را محدود می کنند آگاه بود، بخصوص وقتی که بیش از دو درمان یا متغیر مستقل بیش از دو سطح دارد. در این موارد طرح ها در معرض "اثرات ترتیبی" (Order effect) و "انتقال اثر" (Carry-over) هستند. اثر ترتیبی عبارت است از تفاوت در پاسخ افراد پژوهش به دلیل ترتیب خاص ارائه درمان به آنها (اول، دوم، سوم). اثر ترتیبی می تواند در هر پژوهشی رخ دهد. برای مثال، در پژوهش های پرسشنامه ای افراد ممکن است بسته به ترتیبی که سوالات پرسیده می شود بطور متفاوتی پاسخ دهند. اثرات ترتیبی، به خصوص در طرحهای درون گروهی اهمیت دارند. یعنی وقتی که همان افراد در همه حالات درمان وجود دارند و پژوهشگر می خواهد اثرات بین درمان ها را مقایسه کند. ترتیبی که درمان ها ارائه می شوند ممکن است بر پیامد اثر بگذارد. بنابراین، ترتیب انجام تست ها و درمان هائی تواند بر نتایج اثر گذار باشد.

اثرات ترتیبی ممکن است به دلایل مختلف رخ دهد. برای مثال، افراد معمولاً در نتیجه تمرین، سریع تر پاسخ می دهند، یا اثرات تمرین وقتی که افراد خود را گرم می کنند بروز می کند، و یا در طی زمان عملکرد فرد بهبود می یابد. افراد همچنین ممکن است در پایان آزمایش بطور متفاوتی عمل کنند. اثر خستگی افراد نیز می تواند نتایج را تغییر دهد. مانند وقتی که مدت پژوهش طولانی است و کارها تکراری و بنابراین جذاب نیستند. اثرات تمرین را می توان با تمرین گرم کردن قبل از شروع آزمایشات کاهش داد. اثرات خستگی را نیز می توان با کوتاه کردن زمان تست ها و یا استراحت دادن و جالب تر کردن کارها کاهش داد. پژوهشگران می توانند اثرات ترتیبی را با تغییر سیستماتیک ترتیب ارائه تستها یا درمان ها کاهش دهند (روش متوازن کننده). ارائه تصادفی درمان ها به هر فرد موجب می شود که هر سطح یا درمان حداقل یک بار در همه وضعیت های ممکن پدیدار شود. بعلاوه، پژوهشگر باید مطمئن گردد که هر درمان به تعداد مساوی از افراد ارائه شده است. برای مثال، وقتی دو نوع دارو وجود دارد، برای بررسی اثرات آنها، نیمی از افراد بطور

تصادفی ابتدا داروی "الف" و سپس داروی "ب" را دریافت می کنند، نیمی دیگر از افراد نیز اول داروی "ب" و سپس "الف" را دریافت می کنند.

"انتقال اثر" مانند "اثر ترتیبی" است و وقتی رخ می دهد که اثر یک حالت تجربی (دارو یا هر درمان دیگری) انتقال پیدا می کند و بر عملکرد دارو یا درمان بعدی اثر می گذارد. در انتقال اثر، ممکن است یک درمان و یا سطحی از یک متغیر، اثر اختصاصی بر درمان بعد از آن داشته باشد (اثر فزاینده یا کاهنده). این اثرات وقتی رخ می دهند که درمان یا داروها به سرعت به دنبال یکدیگر ارائه می شوند. این اثرات همچنین به سکانس مشخصی از حالات بستگی دارند. برای مثال: تخمین قد بازیکنان ورزش های دیگر ممکن است پس از دیدن بازیکنان بسکتبال تحت تاثیر قرار بگیرد و کوتاه تر تخمین زده بشوند. بنابراین یک درمان (یا سطحی از متغیر مستقل) ممکن است اثر خاصی بر درمان بعدی داشته باشد، یعنی اثر درمان بعدی را افزایش داده یا کاهش دهد. مجدداً، پژوهشگر باید سکانس ارائه درمان ها را تغییر دهد تا امکان این اثر را به حداقل برساند. اثر انتقال میتواند با افزایش زمان بین ارائه داروها و درمان ها کاهش یابد.

برای انتخاب ترتیب درمان ها در طرح اندازه گیری مکرر با چند سطح از متغیر مستقل، دو استراتژی وجود دارد (شکل ۱-۶):

۱- انتخاب تصادفی ترتیب درمان ها و سپس چرخش وضعیت شروع (شروع تصادفی با چرخش)

۲- استفاده از تکنیک مربع لاتین (Latin square). این استراتژی، نسبت به شروع تصادفی با چرخش، سطح بالاتری از تصادفی کردن را فراهم می کند، زیرا اغلب هر حالت بطور برابر با دیگری ظاهر می شود و ظهور هر حالت پیش و پس از حالت دیگر، اغلب بطور برابر رخ می دهد.

| A Random order with rotation |     |     |     |
|------------------------------|-----|-----|-----|
| Presentation position        |     |     |     |
| 1st                          | 2nd | 3rd | 4th |
| SC                           | SA  | MA  | DR  |
| SA                           | MA  | DR  | SC  |
| MA                           | DR  | SC  | SA  |
| DR                           | SC  | SA  | MA  |
| B Latin square               |     |     |     |
| Presentation position        |     |     |     |
| 1st                          | 2nd | 3rd | 4th |
| SC                           | SA  | MA  | DR  |
| SA                           | DR  | SC  | MA  |
| DR                           | MA  | SA  | SC  |
| MA                           | SC  | DR  | SA  |

شکل ۱-۶: انتخاب ترتیب درمان ها برای چهار سطح متغیر مستقل.

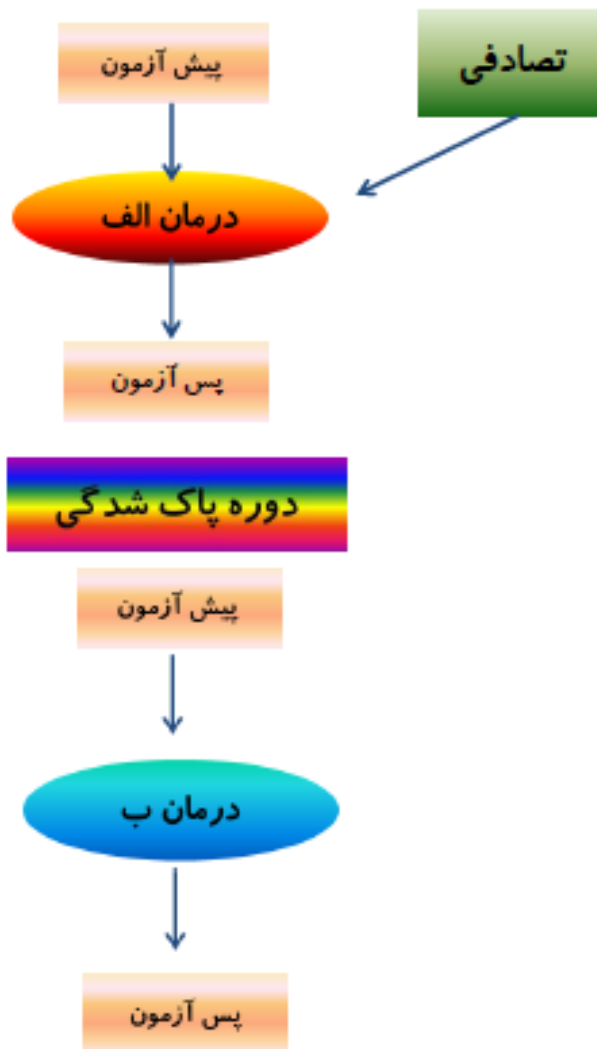
الف: ترتیب تصافی با چرخش، انتخاب تصادفی درمان ها در ردیف اول، و سپس چرخش ترتیب ها برای ردیف های باقیمانده. حالت پررنگ شده درون بیضی نشان دهنده قرارگیری هر سطح در هر وضعیت از طریق چرخش به طریق سیستماتیک می باشد. یک چهارم افراد بطور تصادفی ترتیب درمان های ردیف اول را دریافت می کنند، یک چهارم ترتیب درمان ها در ردیف دوم، و غیره.

ب: انتخاب چهار ترتیب از طریق مربع لاتین. حالت پررنگ شده درون بیضی نشان دهنده قرارگیری هر سطح در هر وضعیت بطور تصادفی است، و هر حالت فقط یک بار قبل و پس از حالت دیگر قرار می گیرد.

در یک پژوهش، اثر چهار ارتز مچ دست بر دامنه حرکتی مچ دست در طی پرتاب بسکتبال بررسی شد. متغیر مستقل یعنی "ارتز" پنج سطح داشت: چهار سطح با ارتز و یک سطح بدون استفاده از ارتز. چهل نفر تحت هر یک از این حالات (یعنی استفاده از هر یک از ارتزها و بدون استفاده از ارتز)، سه پرتاب انجام دادند. بین پنج حالت تست، استراحت مختصری داده شد. گرچه طرح درمان مکرر این اطمینان را ایجاد می کند که ویژگی های

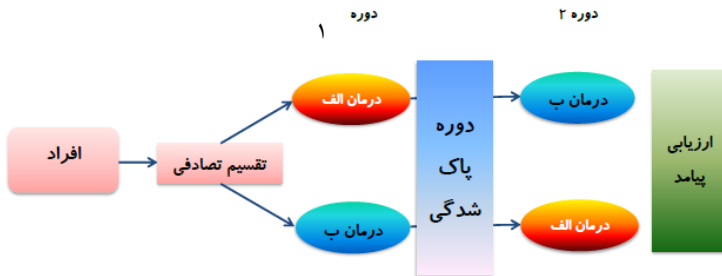
افراد مثل سن، قد، و وزن در سطوح مختلف متغیر مستقل، ثابت باقی بمانند، اما مجموعه جدیدی از عوامل مخدوش کننده که مربوط به آشنایی با روش های انجام آزمایش هستند، ایجاد می شود. برای کنترل این عوامل مخدوش کننده، پژوهشگران درمان ها را بطور تصادفی ارائه دادند.

در طرح های کنترل شده تصادفی یا آر-سی - تی معمول، دو گروه به طور تصادفی انتخاب و مداخلات را نیز به صورت تصادفی دریافت می کنند. اما در "طرح درمان مکرر" که نوعی طرح اندازه گیری مکرر است، یک گروه وجود دارد و هر فرد بیش از یک درمان را دریافت می کند. در این طرح برای کنترل اثر ترتیبی، مداخلات با ترتیبی تصادفی به افراد ارائه می شوند. در این طرح ها، برای این که اثرات مداخله قبلی از بین برود، مرحله ای به نام دوره پاک شدگی (washout) در نظر گرفته می شود برای مثال در یک طرح اندازه گیری مکرر برای بررسی اثر کافئین بر عملکرد شناختی، افراد یک فنجان قهوه کافئین دار و پلاسبو را با ترتیبی تصادفی دریافت می کنند. همه ی افراد، قبل و بعد از مصرف قهوه کافئین دار (درمان الف)، و همچنین قبل و بعد از پلاسبو (درمان ب) تست می شوند (شکل ۱-۷).



شکل ۱-۷: طرح درمان مکرر با یک گروه و دوره پاک شدگی

وقتی که ترتیب دریافت مداخلات توسط افراد دو گروه تصادفی باشد، این پژوهش یک RCT با طرح متقاطع (Cross-over) است. در کارآزمایی کنترل شده تصادفی با طرح متقاطع دو گروه افراد وجود دارند و همه درمان ها بطور تصادفی توسط بیماران دریافت می شود و اندازه گیری نتایج در طی زمان انجام می شود. به این ترتیب "دو دوره و دو درمان" وجود دارد (شکل ۱-۸). در این طرح پس از دوره پاک شدگی (Whash-out)، افراد دریافت کننده درمان الف باید درمان ب را دریافت کنند و بالعکس.



شکل ۱-۸: طرح کارآزمایی کنترل شده تصادفی با طرح متقاطع

## مزایای طرح اندازه گیری مکرر

- تعداد بیماران کمتری لازم است وارد طرح شوند. به خصوص وقتی که تعداد افراد کمی در دسترس هستند. به این ترتیب می توان یک پژوهش موثرتر با توان آماری بالاتر انجام داد و تغییرات افراد را در طی زمان بررسی کرد.
- کنترل تغییرپذیری ناشی از تفاوت های فردی، به طوری که حفظ تغییرپذیری در حد کم، پژوهش را موثرتر می کند.
- چون مداخلات در یک گروه انجام می شود، تا حدودی اثرات متغیر مخدوش کننده کاهش پیدا کرده و کنترل می شود.

## معایب طرح اندازه گیری مکرر

- دریافت همه حالات درمان برای همه افراد امکان پذیر نیست (به علت محدودیت زمانی، محل پژوهش، و ...)
- احتمال از دست دادن داده ها، ریزش افراد، وجود نقاط زمانی متفاوت بین افراد، نابرابری واریانس، و عواملی که می تواند به سوگیری در نمونه گیری (Sampling bias) منجر شده و سطح خطای نوع I را بالا ببرد.
- تهدید اعتبار داخلی طرح به دلیل "تهدید رگرسیون" (وقتی افراد چندین بار تست می شوند، نمرات آنها به سمت میانگین تمایل پیدا می کنند).
- تهدید "بلوغ" (Maturation) (افراد ممکن است در طی دوره پژوهش تغییر کنند).
- تهدید "سابقه" (History) (وقایع خارج از پژوهش که ممکن است باعث تغییر پاسخ افراد در اندازه گیری های مکرر گردد).
- اثر ترتیب اعمال مداخله (Order effect)
- انتقال اثر مداخله مرحله اول (Carry-over)

## طرح پژوهش چندعاملی

هنگام بررسی سیستم های بیولوژیکی پیچیده، ممکن است فاکتورهای تجربی متعددی (چند عاملی) در تاثیر بر پاسخ سیستم تعامل داشته باشند. برای مثال، در بررسی اثرات دو دارو که می توانند به طور همزمان مصرف شوند، مشاهده همه جفت ترکیبات سطوح دو دارو در یک تحقیق منفرد به جای تغییر سطوح یک دارو نسبت به سطح ثابت داروی دیگر، روشن کننده تر است. اگر ما اثر داروها را به طور مستقل بررسی کنیم، ممکن است اطلاعات مربوط به سینرژی ها یا آنتی سینرژیها و حساسیت در آشکار کردن اثرات داروها را از دست بدهیم. در واقع وقتی که چندین فاکتور بر یک سیستم اثر می کنند، تعامل فاکتورها می تواند حساسیت را افزایش دهد. طرح های پژوهش چند عاملی، اغلب در

پژوهش های مختلف حیطه ی سلامت استفاده می شوند. به علت استفاده گسترده از این طرح ها، لازم است پژوهشگران با زبان و مفاهیم مربوط به طرح های چند عاملی آشنا باشند.

از طرح های عاملی یا فاکتوریال، زمانی استفاده می شود که هدف پژوهش، بررسی اثر دو یا چند متغیر مستقل (یعنی مداخله) است. در طرح های عاملی، منظور از فاکتور، متغیر مستقل است. در پژوهشی با دو متغیر مستقل که هر یک دو سطح دارند، طرح عاملی  $2 \times 2$  با چهار گروه خواهیم داشت که هر یک از آنها ترکیب متفاوتی از سطوح فاکتورها را خواهند داشت. نماد یک طرح عاملی  $2 \times 2$  بصورت زیر خواهد بود:

|           |          |                        |          |
|-----------|----------|------------------------|----------|
| <b>G1</b> | <b>O</b> | <b>X<sub>1,1</sub></b> | <b>O</b> |
| <b>G2</b> | <b>O</b> | <b>X<sub>1,2</sub></b> | <b>O</b> |
| <b>G3</b> | <b>O</b> | <b>X<sub>2,1</sub></b> | <b>O</b> |
| <b>G4</b> | <b>O</b> | <b>X<sub>2,2</sub></b> | <b>O</b> |

در یک طرح عاملی  $2 \times 2$  فقط پس آزمون، نماد ترکیبات مختلف سطوح فاکتورها بصورت زیر خواهد بود:

|           |                        |          |
|-----------|------------------------|----------|
| <b>G1</b> | <b>X<sub>1,1</sub></b> | <b>O</b> |
| <b>G2</b> | <b>X<sub>1,2</sub></b> | <b>O</b> |
| <b>G3</b> | <b>X<sub>2,1</sub></b> | <b>O</b> |
| <b>G4</b> | <b>X<sub>2,2</sub></b> | <b>O</b> |

در طرح های عاملی، هر متغیر مستقل می تواند بیش از دو سطح داشته باشد. برای مثال، یک طرح عاملی  $3 \times 4$  نیازمند ۱۲ گروه درمان می باشد. بنابراین، با افزایش تعداد سطوح و فاکتورها، تعداد گروه ها افزایش می یابد.

طرح های عاملی وقتی مناسب هستند که بخواهیم اثر چندین مداخله فعال را بررسی کنیم. برای مثال، اگر پژوهشگری بخواهد اثر قرص رژیمی A را در مقایسه با پلاسبو، رژیم غذایی B را در مقایسه با رژیم غذایی معمول، و یک برنامه تمرین اختصاصی C را در مقایسه با تمرین معمول بر کاهش وزن بدن در مردان بزرگسال دچار دیابت نوع ۲ بررسی کند، استفاده از طرح عاملی در مقایسه با بررسی هر یک از آنها بطور جداگانه کارآیی بیشتری داشته و اطلاعات مفیدتری فراهم خواهد کرد. در طرح عاملی هر فرد بطور تصادفی در گروهی با ترکیب خاصی از مداخلات ABC قرار می گیرند و عموماً همه ترکیبات مورد بررسی قرار می گیرند. طرح عاملی اجازه می دهد علاوه بر شناسایی اثر اصلی هر یک از مداخلات A، B، و C، جفت تعاملات بین A و B، A و C، و B و C، و تعامل سه جانبه بین این سه نوع درمان، مورد بررسی قرار بگیرد.

مثال دیگری از یک طرح عاملی، بررسی اثرات نور، کود، و درجه حرارت بر رشد گیاه است. به جای بررسی جداگانه اثر هر یک از این متغیرها می توان از طرح عاملی برای بررسی همه ترکیبات برای کسب الگوی رشد بهینه استفاده کرد که ممکن است از سطوح بهینه هر یک از سه متغیر بررسی شده به طور جداگانه، متفاوت باشد.

پژوهشگران معمولاً از تجربیات چندعاملی استفاده می کنند زیرا آنها علاوه بر اثرات منفرد فاکتورهای متعدد بر متغیر وابسته، به اثرات تعامل بین فاکتورهای متعدد بر متغیر وابسته علاقه مند هستند. برای مثال، فرض کنید می خواهیم اثر برنامه های درمانی مختلف را بر عمل بلع مبتلایان به سگته مغزی بررسی کنیم. برای این منظور ۶۰ بیمار واجد شرایط ورود به پژوهش را انتخاب کرده و آنها را در سه گروه تقسیم می کنیم (برنامه P، برنامه OM، و برنامه FC). بنابراین، متغیر مستقل اول نوع درمان یا گروه خواهد بود. همچنین فرض کنید که دو درمانگر مختلف (A و B) بیماران را درمان خواهند کرد. ما می خواهیم بدانیم صرفنظر از نوع درمان، آیا یکی از درمانگرها ممکن است نسبت به درمانگر دیگر نتایج بهتری بدست بیاورد. پس متغیر مستقل دوم، درمانگر است. اگر درمانگر به عنوان متغیر

مستقل دوم اضافه شود، سوال سوم عبارت خواهد بود از این که آیا یک درمانگر در استفاده از یک نوع برنامه درمانی موثرتر خواهد بود و درمانگر دیگر با نوع دیگر درمان؟ به عبارت دیگر، آیا بین نوع درمان و درمانگر تعامل وجود دارد؟

بررسی تعامل بین متغیرها، طرح تجربی چند عاملی (Multiple Factor Experimental Design) را از یک طرح تک عاملی (Single Factor) متمایز می کند. در واقع، طرح چند عاملی یک طرح فاکتوریال با بیش از یک متغیر مستقل است که در آن اثرات هر یک از متغیرها یا فاکتورها بر متغیر وابسته به صورت جداگانه و همچنین اثرات همزمان چندین متغیر مستقل بر متغیر وابسته مورد نظر، بررسی و اندازه گیری می شود.

اثر تعاملی، حاصل ترکیب دو یا چند متغیر مستقل است که نسبت به هر یک از متغیرهای مستقل به تنهایی، نتیجه متفاوتی را تولید می کند. تعاملات فقط در طرح های فاکتوریال که شامل دو یا چند متغیر مستقل است، یافت می شوند. هنگام مرور نتایج یک پژوهش فاکتوریال، ما با تعیین وجود هر گونه تعاملات معنی دار شروع می کنیم. اگر تعاملات معنی داری بدست بیاید، ما دیگر نمی توانیم اثرات ساده هر یک از متغیرهای مستقل (یعنی تفاوت های بین گروهی برای هریک از متغیرهای مستقل به تنهایی) را تفسیر کنیم، زیرا نتایج حاصله، در اثر تعامل با سطوح مختلف یک یا چند متغیر مستقل دیگر تغییر می کنند. برای مثال، دوز یک مداخله مشخص، با جنس افراد و با نرخ موفقیت آنها تعامل دارد. به این ترتیب ما نمی توانیم اثرات ساده جنس یا دوز مداخله بر روی نرخ موفقیت افراد را تفسیر کنیم چون آنها بصورت تابعی از یکدیگر فرق می کنند. ما فقط می توانیم تعاملات را تفسیر کنیم. یعنی مثلاً مردان با دوزهای پایین تر و زنان با دوزهای بالاتر موفق تر هستند.

انواع طرح های پژوهش چند عاملی عبارتند از:

طرح های فاکتوریال و آشیانه ای (Factorial and nested designs)

طرح های کاملاً تصادفی و بلوک تصادفی

طرح های بین گروهی

طرح های درون گروهی

طرح های مرکب

## طرح های فاکتوریال و آشیانه ای

ویژگی طرح های چند عاملی، استفاده از چندین متغیر مستقل و مشخص کردن اثر آنها بر متغیر وابسته است. در این نوع طرح، پژوهشگر به دنبال آن است که تاثیر چندین متغیر مستقل را بر متغیر وابسته بسنجد. در طرح نوع درمان و درمانگر بر بلع مبتلایان به سکنه مغزی، طرح نوع درمان  $\times$  درمانگر یک طرح فاکتوریال است. در یک طرح فاکتوریال، فاکتورها متقاطع هستند، به این معنی که هر سطح از یک فاکتور با سطوح مختلف فاکتور دیگر ترکیب شده است. یعنی هر گروه درمان اعضایی دارد که به هر درمانگر اکسپوز شده است، و هر درمانگر به اعضای هر گروه درمان اکسپوز شده است (شکل ۱-۹). "گروه" یک متغیر است با سه سطح: P، OM، و FC. "درمانگر" متغیر دوم است با دو سطح: A و B. در این طرح، شش سلول وجود دارد که به وسیله تقاطع دو فاکتور بوجود آمده اند.

در این مثال، فرض کنید که ۶۰ بیمار به طور تصادفی و مساوی در گروه ها قرار بگیرند (۲۰ نفر در هر سلول). عمل بلع در شروع درمان اندازه گیری شده، و سپس درمان مناسب به هر بیمار ارائه می شود (هر درمان، یک بار توسط درمانگر A و یک بار توسط درمانگر B انجام می گردد)، و عمل بلع مجدداً در پایان درمان اندازه گیری می شود. متغیر وابسته، تغییر در نمره عمل بلع از پیش آزمون به پس آزمون است.

|                   |    | متغیر مستقل: درمانگر                          |                                               |
|-------------------|----|-----------------------------------------------|-----------------------------------------------|
|                   |    | A                                             | B                                             |
| متغیر مستقل: گروه | P  | میانگین تغییر نمره بلع<br>PA<br>-7/25 میانگین | میانگین تغییر نمره بلع<br>PB<br>-7/20 میانگین |
|                   | OM | میانگین<br>OMA<br>-7/35 میانگین               | میانگین<br>OMB<br>-7/30 میانگین               |
|                   | FC | میانگین<br>FCA<br>-7/45 میانگین               | میانگین<br>FCB<br>-7/40 میانگین               |
|                   |    | A میانگین -7/40                               | B میانگین -7/30                               |

ترکیب دو  
متغیر مستقل

شکل ۱-۹ الف: طرح فاکتوریال  $2 \times 3$ .

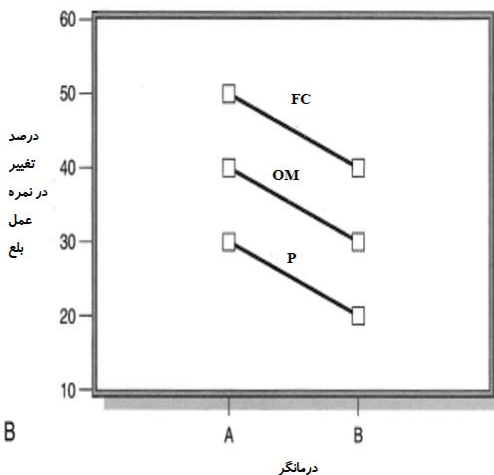
ستون سمت چپ شکل، اثر هر روش درمان را صرفنظر از نوع درمانگر نشان میدهد. حروف هر سلول در ستونهای سمت راست، ترکیب دو متغیر مستقل را نشان می دهد. برای مثال، PB بیانگر این است که افراد این سلول درمان P را از درمانگر B دریافت کرده اند. نتایج بیانگر این است که صرفنظر از درمانگر، درمان FC موثرتر است، و درمانگر A، صرفنظر از نوع درمان، بهتر است. میانگین داده ها برای هر ترکیب و میانگین کل برای ستون ها و ردیف ها ارائه شده است.

نماد طرح فوق به صورت زیر می باشد:

یک طرح فاکتوریال گروه-کنترل پیش آزمون-پس آزمون (متغیرهای مستقل: درمان و درمانگر، متغیر وابسته: تغییر در عملکرد بلع).

|   |   |    |   |      |
|---|---|----|---|------|
| R | O | X1 | O | n=20 |
| R | O | X2 | O | n=20 |
| R | O | X3 | O | n=20 |

شکل ۹-۱ الف نشان می دهد که صرفنظر از نوع درمان، درمانگر A در کل نسبت به درمانگر B بهتر است (میانگین کل بهبود بلع بیماران برای درمانگر A به میزان ۴۰٪ در مقابل ۳۰٪ بهبودی برای بیماران درمانگر B می باشد). برای نوع درمان، صرفنظر از درمانگر ارائه کننده درمان ها، در کل برنامه FC (با بهبود کل ۴۵٪) نسبت به دو نوع درمان دیگر (با بهبودی کل به ترتیب ۲۵ و ۳۵ درصد)، بهتر بوده است. در این مثال، هیچگونه تعاملی بین درمان و درمانگر وجود ندارد. روش FC صرفنظر از درمانگر، درمان برتر می باشد، و درمانگر A، صرفنظر از نوع درمان، نتایج بهتری بدست می آورد. این عدم تعامل در شکل ۹-۱ ب بوسیله خطوط موازی بین درمانگرها برای سه نوع درمان نشان داده شده است. این نتایج را می توان به صورت زیر خلاصه کرد: "اثر اصلی (Main effect) برای "نوع درمان" و "درمانگر" وجود دارد، اما بین نوع درمان و درمانگر "تعامل" (Interaction) وجود ندارد.



شکل ۹-۱ ب: خطوط موازی بیانگر عدم تعامل بین درمانگر و نوع درمان است.

در شکل ۱۰-۱ الف برای متغیرها داده های دیگری نشان داده شده است. در شکل الف، در کل هیچگونه تفاوتی در درمان ها وجود ندارد (میانگین بهبودی برای هر نوع درمان،

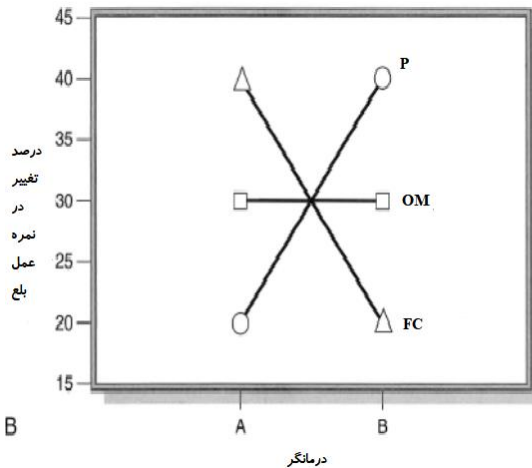
صرفنظر از درمانگر، ۳۰ درصد است) و در کل هیچگونه تفاوتی در نتایج درمانگران وجود ندارد (میانگین بهبودی برای هر درمانگر، صرفنظر از نوع درمان، ۳۰ درصد است). گرچه، تعامل معنی داری بین نوع درمان و درمانگر وجود دارد. در شکل ۱-۱۰ ب، درمانگر A نتایج بهتری با استفاده از درمان FC بدست آورده است و درمانگر B نتایج بهتری با استفاده از درمان P بدست آورده است. هر دو درمانگر نتایج متوسط با استفاده از رویکرد OM بدست آورده اند. یک راه بیان این نتایج بطور خلاصه این است: "هیچگونه اثر اصلی برای نوع درمان یا درمانگر وجود ندارد، اما بین نوع درمان و درمانگر تعامل وجود دارد." اگر ما فقط انواع درمان های مختلف را بدون توجه به فاکتور دوم (درمانگر) بررسی میکردیم، نتیجه گیری می کردیم که هیچگونه تفاوتی بین این درمان ها وجود نداشت. پژوهش های دو عاملی به ما اجازه می دهد که نتیجه گیری پیچیده تری کنیم: اگرچه تفاوت کلی بین درمان ها و درمانگرهای مورد بررسی وجود نداشت، درمانگرهای معین هنگام استفاده از درمان های معین بطور قابل ملاحظه ای موثرتر هستند.

|                   |    | متغیر مستقل: درمانگر |               |
|-------------------|----|----------------------|---------------|
|                   |    | A                    | B             |
| متغیر مستقل: گروه | P  | 20- میانگین          | 40- میانگین   |
|                   | OM | 30- میانگین          | 30- میانگین   |
|                   | FC | 40- میانگین          | 20- میانگین   |
|                   |    | 30- میانگین A        | 30- میانگین B |

شکل ۱-۱۰ الف: نمونه طرح فاکتوریل چند عاملی با تعامل بین متغیرها.

میانگین داده ها برای افراد برای هر ترکیب، و میانگین کل برای ستون ها و ردیف ها ارائه شده است. میانگین کل ستونها (۳۰ درصد) و ردیفها (۳۰ درصد)، عدم تفاوت در نوع

درمان، صرفنظر از درمانگر، و عدم تفاوت در نتایج بین درمانگرها، صرفنظر از درمان را نشان می دهد.



شکل ۱-۱۰ ب: خطوط غیرموازی بیانگر تعامل بین درمانگر و نوع درمان است.

## طرح آشیانه ای

در طرح های آشیانه ای (Nested design) همه فاکتورها یکدیگر را قطع نمی کنند. برای مثال، در اندازه گیری کنترل کیفیت پلاسمای خون اهدایی، پلاسمای بگ های A در مجموعه ای از مراکز خون B پردازش شده اند یعنی پلاسمای بگ های A در مراکز B آشیانه کرده اند. در این شرایط، انتخاب نمونه پلاسمای بگ ها و همچنین انتخاب مراکز برای بررسی بطور تصادفی خواهد بود. در اینصورت هنگام آنالیز داده ها، لازم است در نظر داشته باشیم که پلاسمای بگ ها از کدام مراکز هستند.

ما می توانیم مثال سه روش درمانی مختلف و دو درمانگر را با اضافه کردن درمانگر و یک کلینیک اضافه پیچیده تر کنیم. در این مثال، ۳ درمان (متغیر مستقل با ۳ سطح)، ۲ بیمارستان (متغیر مستقل با ۲ سطح عمومی و تخصصی توانبخشی اعصاب) و ۳ درمانگر

در هر بیمارستان (متغیر مستقل با ۳ سطح) خواهیم داشت. یک سوال پژوهشی ممکن است این باشد که آیا پیامدهای بیمار در دو بیمارستان فرق می کند. سوال پژوهشی دیگر این است که آیا درمانگرهای مختلف نتایج مختلف بدست می آورند. این سوال با استفاده از سه درمانگر مختلف در هر بیمارستان بررسی می شود. چون در هر بیمارستان درمانگرهای مختلف وجود دارد، پس متغیر درمانگر در متغیر بیمارستان آشیانه کرده است. نمای شماتیک این طرح فرضی در شکل ۱-۱۱ نشان داده شده است که در آن درمان و بیمارستان ها تقاطع کرده اند و درمانگر در بیمارستان آشیانه کرده است.

|                   |   | متغیر مستقل: بیمارستان |     |     |         |     |     |
|-------------------|---|------------------------|-----|-----|---------|-----|-----|
|                   |   | S تخصصی                |     |     | G عمومی |     |     |
| متغیر مستقل: گروه | P | PSA                    | PSB | PSC | PGD     | PGE | PGF |
|                   | O | OSA                    | OSB | OSC | OGD     | OGE | OGF |
|                   | F | FSA                    | FSB | FSC | FGD     | FGE | FGF |
|                   |   | A                      | B   | C   | D       | E   | F   |
|                   |   | متغیر مستقل: درمانگر   |     |     |         |     |     |

شکل ۱-۱۱: دیاگرام شماتیک طرح سه فاکتوری آشیانه ای.

سه روش درمان (P, O, F) و دو بیمارستان عمومی (G) و تخصصی (S) فاکتورهای تقاطع یافته هستند. درمانگر (A-F) در بیمارستان آشیانه کرده است.

## طرح های کاملاً تصادفی و بلوک تصادفی

مثال درمان  $\times$  درمانگر که در بالا توضیح داده شد، نمونه یک طرح کاملاً تصادفی است چون "نوع درمان" و "درمانگر" هر دو متغیرهای مستقل هستند و افراد بطور تصادفی در گروه های درمان و درمانگر تقسیم شدند.

گاهی اوقات "طرح بلوک تصادفی" (Randomized-block design) استفاده می شود که در آن یکی از فاکتورها قابل دستکاری نیست. یعنی یکی از متغیرهای مستقل فعال است (قابل دستکاری است)، و فاکتور دیگر یک متغیر کیفی زمینه ای (Attribute) مثل سن یا بیماری است. اگر، برای مثال، بخواهیم به فاکتورهای جنس بیمار و درمان نگاه کنیم، جنس یک متغیر زمینه ای (Attribute) یا طبقه ای (Classification) است. افراد بر اساس جنس در بلوک ها قرار می گیرند و سپس بطور تصادفی به گروه های درمانی تقسیم می شوند. این نوع آرایش که "طرح مرکب" (Mixed design) نیز نامیده میشود در پژوهش های حوزه سلامت بسیار رایج است. اغلب اوقات، می توان فاکتورهای درون گروهی (Within-subjects) را نیز به عنوان یک متغیر مستقل در نظر گرفت، و همه افراد همه سطوح درمان را نیز دریافت کنند. فاکتور درون گروهی معمولاً گروههای پژوهش هستند که بیشتر اوقات بر اساس یک متغیر زمینه ای (برای مثال کودکانی که رشد طبیعی دارند در مقابل کودکان با اختلالات شنوایی) طبقه بندی شده اند.

طرح بلوک تصادفی یک طرح تجربی است که در آن واحدهای تجربی در گروه هایی به نام "بلوک ها" هستند. درمان ها بطور تصادفی به واحدهای تجربی درون هر بلوک اختصاص می یابند. وقتی که همه درمان ها حداقل یک بار در هر بلوک ظاهر شدند، طرح بلوک تصادفی کامل است. در غیر این صورت، طرح بلوک تصادفی ناکامل خواهد بود. این نوع طرح برای به حداقل رساندن اثرات خطای سیستماتیک استفاده می شود. اگر پژوهشگر بر تفاوت بین دو درمان متمرکز شود، اثرات ناشی از واریاسیون های بین بلوک ها باید حذف شوند. به عبارت دیگر، در طرح بلوک تصادفی، پژوهشگر افراد را به زیر

گروه هایی به نام "بلوک" تقسیم می کند بطوری که تغییرپذیری در هر بلوک کمتر از تغییرپذیری بین بلوک ها است. سپس، افراد درون هر بلوک بطور تصادفی به گروه های درمان تقسیم می شوند. در مقایسه با طرح کاملاً تصادفی، این طرح تغییرپذیری درون گروه های درمان و متغیرهای بالقوه مخدوش کننده را کاهش می دهد و به این ترتیب تخمین بهتری از اثرات درمان می دهد.

مثالی از طرح بلوک تصادفی، بررسی اثر واکسن در مقایسه با پلاسبو است، افراد بر اساس جنس در دو بلوک "مرد" و "زن" قرار می گیرند و سپس بطور تصادفی در گروه درمان واکسن یا پلاسبو تقسیم می شوند (شکل ۱-۱۲). جنسیت فاکتوری است که نمی توان آن را تغییر داد و دستکاری کرد. چنانچه پژوهشگر بخواهد اثر آن را کنترل کند، فاکتور جنسیت را بلوک کرده و افراد از هر دو جنس را بطور تصادفی در گروه ها تقسیم می کند. در این مثال، هر بلوک شامل ۵۰۰ نفر است که بطور مساوی در دو گروه تقسیم شده اند و به این ترتیب در گروه واکسن و پلاسبو به تعداد مساوی مرد و زن وجود دارد. برای این طرح، ۲۵۰ مرد پلاسبو و ۲۵۰ مرد واکسن دریافت می کنند. همچنین ۲۵۰ زن پلاسبو و ۲۵۰ زن واکسن دریافت می کنند. چنین طرحی، این اطمینان را ایجاد می کند که هر نوع درمان (واکسن و پلاسبو) برای نسبت برابری از زنان و مردان انجام می شود. در نتیجه، تفاوت در اثرات انواع درمان را نمی توان به جنس نسبت داد زیرا گرچه مردان و زنان از لحاظ فیزیولوژیک متفاوت هستند و به داروها واکنش متفاوتی نشان می دهند، ولی به دلیل تساوی تعداد آنها در دو گروه، عملاً اثر جنس یکسان بوده و تفاوت های حاصله ناشی از نوع درمان خواهند بود. در این طرح، اثر متغیر "جنس" به عنوان یک منبع بالقوه تغییرپذیر و یک متغیر بالقوه مخدوش کننده حذف می شود.

| جنس | درمان |        |
|-----|-------|--------|
|     | واکسن | پلاسبو |
| مرد | ۲۵۰   | ۲۵۰    |
| زن  | ۲۵۰   | ۲۵۰    |

شکل ۱-۱۲: طرح بلوک تصادفی. اعداد، تعداد افراد زن و مرد را در گروه ها نشان می

دهد.

## طرح های بین گروهی، درون گروهی و مرکب

بطور خلاصه، در طرح های تجربی چندعاملی، ما می توانیم طرح های بین گروهی (Between-Groups)، درون گروهی (Within-Group)، و مرکب (Mixed) داشته باشیم.

در طرح بین گروهی، گروه های مختلف سطوح مختلفی از درمان ها را دریافت می کنند. در طرح درون گروهی، همه افراد گروه همه درمان های تجربی را دریافت می کنند. در طرح مرکب، حداقل یک فاکتور بین گروهی و یک فاکتور درون گروهی وجود دارد. بیشتر اوقات، یکی از متغیرهای مستقل قابل دستکاری است (متغیر فعال، عموماً فاکتور درون گروهی Within-subjects) ولی دیگری نیست [متغیر زمینه ای یا کیفی (همچنین ویژگی یا صفت Assigned یا Attribute)، عموماً فاکتور بین گروهی].

### طرح چند عاملی بین گروهی

#### دریافت سطوح مختلفی از درمان ها توسط گروه های مختلف

مثال درمان  $\times$  درمانگر، همچنین یک طرح بین گروهی برای هر دو فاکتور است. در هر یک از شش سلول طرح، افراد متفاوتی قرار گرفته اند (سه گروه  $\times$  دو درمانگر). سوالات پژوهش این طرح عبارتند از:

بین گروه هایی که درمان های مختلف دریافت کرده اند، چه اختلافی وجود دارد؟

بین گروه هایی که درمانگرهای مختلف داشته اند، چه اختلافی وجود دارد؟-

آیا بین اثر درمان و درمانگر تعامل وجود دارد؟

در یک طرح بین گروهی دو عاملی  $2 \times 2$ ، پژوهشگران اثر تمرینات فیزیکی و غذای غنی شده بر بهبود عملکرد سالمندان را مورد بررسی قرار دادند. یک فاکتور "تمرین" بود با دو سطح: تمرین و عدم تمرین. یک فاکتور نیز رژیم غذایی بود با دو سطح: غذای غنی شده و فقدان غذای غنی شده. در این طرح، با تقاطع سطوح متغیرهای موجود، چهار سلول ایجاد می شود که هر یک افراد مختلفی را در برمی گیرند:

۳۹ نفر فقط تمرین کردند .

۳۹ نفر فقط غذای غنی شده خوردند.

۴۲ نفر، هم تمرین کردند و هم غذاهای غنی شده خوردند

۳۷ نفر گروه کنترل بودند (نه تمرین کردند و نه غذاهای غنی شده خوردند)

نتیجه این پژوهش نشان داد که بین اثرات تمرین کردن و خوردن غذاهای غنی شده تعامل وجود ندارد، گروه تمرین کننده در مقایسه با گروه تمرین نکرده، بهبودی معنی داری در انجام فعالیت های عملکردی داشتند، و بین گروهی که غذاهای غنی شده خوردند و گروهی که نخوردند تفاوت معنی دار وجود نداشت. بر اساس نتایج بدست آمده، عملکرد و تناسب فیزیکی جمعیت سالمند، با تمرین افزایش یافت، ولی مصرف روزانه غذاهای غنی شده با ریز مغذی ها هیچگونه فوائد عملکردی نداشت.

در طرح درون گروهی، یک گروه از افراد، همه سطوح متغیر مستقل را دریافت می کنند. در یک طرح کاملاً درون گروهی، پژوهشگران اثرات بیوفیزیکی امواج ماوراء صوت بر سرعت هدایت عصب مدیان و به طور مشخص زمان تاخیری دیستال را بررسی کردند. در این طرح، دو متغیر مستقل وجود داشت که شامل: "درمان" (پنج سطح شامل انواع امواج ماوراء صوت با فرکانس ۲ و ۳ مگاهرتز، و امواج ماوراء صوت پلاسبو) و "زمان" (اندازه گیری ها قبل از درمان، ۲، ۴، و ۶ دقیقه از درمان، و پس از درمان) بود. همه ۱۵ نفر شرکت کننده در پژوهش، همه انواع امواج ماوراء صوت را بطور تصادفی دریافت کردند، و برای هر درمان، پیامدها در ۵ زمان اندازه گیری شدند. بنابراین هر دو متغیر مستقل در

این پژوهش، فاکتورها با اندازه گیری های مکرر، در یک طرح کاملاً درون گروهی هستند.

### طرح چند عاملی درون گروهی

#### یک گروه از افراد همه سطوح متغیر مستقل را دریافت می کنند

در یک طرح درون گروهی پیش آزمون-پس آزمون، ۲۹ بیمار پس از سکتۀ مغزی تحت درمان با سوزن زدن خشک برای بهبود اسپاستیسته عضلات خم کننده مچ دست سمت مبتلا قرار گرفتند. در این تحقیق، اندازه گیری ها قبل، بعد، و یک ساعت پس از پایان درمان انجام شد. نتیجه این پژوهش، اثر مثبت سوزن زدن خشک در کاهش اسپاستیسته بیماران را نشان داد.

طرح مرکب (Split-plot) ترکیبی از فاکتورهای بین گروهی (Between-subjects) و درون گروهی (Within-subjects) است.

### طرح مرکب

#### ترکیبی از فاکتورهای بین گروهی و درون گروهی

#### حداقل یک فاکتور بین گروهی و یک فاکتور درون گروهی

در یک پژوهش با طرح مرکب، پژوهشگران اثرات توانایی استنباط بزرگسالان مبتلا به آسیب نیمکره راست مغز را بررسی کردند. در این پژوهش، ۱۴ بیمار و ۱۴ فرد سالم بدون آسیب مغزی شرکت داشتند. بنابراین، دو گروه بر مبنای یک فاکتور بین گروهی، که یک متغیر کیفی (Assigned) بود (بیمار و سالم)، تشکیل شدند. هر فرد باید داستان هایی را می خواند که نتیجه آنها قابل پیش بینی بود و با استنباط خود، نتیجه ی آنها را پیش بینی می کرد. معیار پیامد این مطالعه، توانایی استنباط افراد بود. در واقع، "قابلیت پیش بینی

نتیجه داستان، "متغیر مستقل این طرح بود که دو سطح بالا یا کم داشت و چون هر فرد هر دو سطح متغیر مستقل را دریافت می کرد، پس فاکتور درون گروهی را تشکیل میداد. همانگونه که انتظار می رفت، بزرگسالان سالم بدون آسیب مغزی استنباطات مثبتی در هردو حالت قابلیت پیش بینی داشتند، و افراد گروه دچار آسیب مغزی فقط وقتی که قابلیت پیش بینی بالا بود استنباطات بیشتری داشتند. بنابراین، بین گروه ها و قابلیت پیش بینی تعامل وجود داشت.

مثال دیگری برای یک پژوهش با طرح مرکب، بررسی اثر دارو بر افسردگی است. "درمان" یک متغیر مستقل است با دو سطح (دریافت دارو یا عدم دریافت دارو) و به این ترتیب یک فاکتور بین گروهی است. "زمان" متغیر مستقل است با دو اندازه گیری، قبل و بعد از درمان. متغیر وابسته نیز نمره افسردگی است. طرح پژوهش یک طرح مرکب  $2 \times 2$  (زمان: قبل-بعد)  $\times$  (درمان: دارو-عدم دارو) است با اندازه گیری مکرر فاکتور اول یعنی زمان. "اندازه گیری مکرر" اشاره به ویژگی "درون گروهی" طرح دارد و روشن می کند کدام فاکتور، درون گروهی است (جدول ۱-۱). اضافه کردن یک فاکتور بین گروهی و تبدیل طرح به یک طرح مرکب، بسیاری از تهدیدهای اعتبار داخلی در طرح های درون گروهی را کاهش می دهد.

جدول ۱-۱: داده های فرضی برای طرح مرکب بررسی اثر دارو بر افسردگی، با در نظر گرفتن "زمان" و "دارو" به عنوان متغیرهای مستقل و نمره افسردگی به عنوان متغیر وابسته.

| متغیر وابسته:<br>نمره افسردگی |       | زمان |      |
|-------------------------------|-------|------|------|
|                               |       | قبل  | بعد  |
| دارو                          | درمان | ۴۷/۳ | ۲۸/۶ |

|      |      |          |  |
|------|------|----------|--|
| ۴۶/۵ | ۴۸/۱ | عدم دارو |  |
|------|------|----------|--|

## خلاصه

پژوهش تجربی، با دستکاری کنترل شده متغیرها توسط پژوهشگر مشخص می شود. گرچه کارآزمایی های کنترل شده تصادفی (RCT)، شواهد خوبی برای تعیین اثر مداخلات کلینیکی فراهم می کنند، اما بعضی سوالات مربوط به مداخلات را نمی توان با RCT ها پاسخ داد و بعضی موضوعات نیز برای تست با RCT ها مناسب نیستند. به علاوه، RCT ها به علت مشکلات انتخاب نمونه ی واقعا تصادفی، در تعمیم نتایج یا اعتبار خارجی (External validity) مشکل دارند. در طرح های پژوهش تک عاملی فقط یک متغیر دستکاری می شود. طرح های پژوهش تک عاملی شامل طرح گروه-کنترل پیش آزمون-پس آزمون، طرح تک گروهی پیش آزمون-پس آزمون، طرح گروه-کنترل نابرابر، طرح سری زمانی، و طرح درمان مکرر است.

در طرح های پژوهشی چند عاملی، پژوهشگر اثر بیش از یک متغیر مستقل را بررسی میکند و به تعاملات بین متغیرها علاقه مند است. انواع رایج طرح های چند عاملی شامل طرح های فاکتوریال و آشیانه ای، طرح های کاملاً تصادفی و بلوک تصادفی، و طرحهای بین گروهی، درون گروهی، و مرکب هستند.



## فصل دوم

### طرح های پژوهش غیر تجربی

بخش های فصل

پژوهش توصیفی

پژوهش توصیفی گذشته نگر، پژوهش توصیفی آینده نگر

مشاهده، معاینه، مصاحبه، پرسشنامه

آنالیز روابط

آنالیز گذشته نگر روابط، آنالیز آینده نگر روابط

آنالیز تفاوت ها

آنالیز گذشته نگر تفاوت ها، آنالیز آینده نگر تفاوت ها

خلاصه

در فصول قبل، پژوهش های تجربی یعنی انواع پژوهش طراحی شده برای پاسخ دادن به سوالات علت و معلولی را توضیح دادیم. اما پژوهشگران در پژوهش های کلینیکی از انواع روش ها استفاده می کنند. بسیاری از این روش ها، مثل مشاهده، مصاحبه، و پرسشنامه ها، رایج هستند. با در نظر گرفتن اصول پژوهش، فعالیتهای کلینیکی را همچنین می توان با پژوهش های غیرتجربی بررسی کرد، و به این دلیل پژوهش های غیرتجربی نیز برای دانشمندان اهمیت دارد. پژوهش غیرتجربی همچنین پژوهش کیفی یا همبستگی نامیده می شود.

امروزه پژوهش های غیرتجربی بطور فزاینده ای اهمیت پیدا کرده و جایگاه خود را پیدا کرده اند. بر خلاف پژوهش های تجربی، در پژوهش های غیرتجربی متغیرها دستکاری نمی شوند. در پژوهش غیرتجربی، کنترل متغیرهای مستقل ممکن نیست و پژوهشگر تاریخچه پدیده های گذشته را بررسی می کند، پدیده های موجود را مستند می کند، یا پدیده های جدید را مشاهده می کند. پژوهش های غیرتجربی، طبقه ای از متودولوژی های پژوهشی هستند که هدف آنها بررسی چگونگی ارتباط متغیرها است. یعنی ارتباط متغیرها با یکدیگر مشاهده شده و سپس یافته ها تفسیر گردیده و توضیح داده می شوند. در شرایط زیر، عموماً طرح های غیرتجربی مناسب هستند:

- ۱- وقتی که دانش کمی در باره موضوع داریم.
  - ۲- وقتی که سنجش پیامدها آسان نیست.
  - ۳- وقتی که مسائل اخلاقی اجازه انجام طرح های تجربی را نمی دهند.
  - ۴- وقتی که هدف پژوهش چیزی غیر از بررسی مؤثر بودن درمان است.
- گاهی ممکن است طیفی از سوالات پژوهشی را در نظر بگیریم که نیازمند شروع با پژوهش های غیرتجربی باشند و سپس در ادامه به سمت پژوهش های تجربی پیش بروند. بنابراین، نمی توان گفت که پژوهش های تجربی فقط به این دلیل که سوالات متفاوتی را نسبت به پژوهش های غیرتجربی پاسخ می دهند، با ارزش تر و مهم تر هستند.

وقتی کلینیسین ها و دانشمندان با پدیده ای جدید یا کمتر بررسی شده مواجه هستند، ابتدا لازم است همه اطلاعات ممکن در باره آن را گردآوری کنند. موضوع اتیسم نمونه خوبی در این حیطه است. برای پژوهش در باره اتیسم، ابتدا لازم است که ویژگی های اتیسم را مطالعه کنیم. بدانیم ابعاد ناتوانی این افراد از لحاظ فیزیکی، ارتباطی، و هیجانی چه هستند؟ و چه ویژگی های مشترکی در تاریخچه افراد اتیسم وجود دارد؟ این سوالات بیشتر برای توصیف یک پدیده مناسب هستند و سوالات علت و معلولی مناسب برای پژوهش تجربی نیستند. پس از این که اطلاعات کافی درباره اتیسم به دست آوردیم، میتوانیم بعضی سوالات، "چه-اگر" (What-if)، را با هدف تعیین موثر بودن روش های درمانی، مطرح کنیم. مثلاً، یک درمان مشخص، چه اثری بر روی رفتارهای معین افراد اتیسم دارد؟ این پرسش، سوالی برای یک پژوهش تجربی است.

یک پژوهش تجربی نیازمند دستکاری یا خلق متغیر مستقل، تقسیم تصادفی به گروه ها یا سطح متغیر مستقل اصلی، و کنترل متغیرهای مخدوش کننده است. اگر پژوهشی در باره رابطه بین دو فاکتور طبیعی (مثل قد و جنسیت) باشد، تجربی نیست. همچنین، پژوهش هایی که گروه ها را بر پایه ویژگی های ذاتی مثل جنس یا دست غالب مقایسه می کنند تجربی نیستند، چون متغیرهایی مثل قد، اندازه پا، جنسیت، یا دست غالب، متغیرهای مستقلی هستند که توسط پژوهشگر دستکاری و تعیین نمی شوند (شکل ۱-۲). پژوهشهای آرشیوی از جمله آنهایی که بر اطلاعات یک پایگاه داده موجود متکی هستند، ماهیت تجربی ندارند، چون افراد به طور تصادفی در گروه ها تقسیم نگردیده، و متغیرهای مخدوش کننده کنترل نشده اند. همچنین اندازه گیری ها به ندرت توسط افراد بی اطلاع (Blind) از شرایط بیماران بدست آمده اند، و بیماران نیز به ندرت از ابزار درمان بی اطلاع هستند (Blind). به علاوه، ممکن است کنترل بسیار کمی بر سایر بیماری ها که متغیر وابسته در آن حالات اندازه گیری شده اند، وجود داشته باشد.

تنوع طرح های غیرتجربی بیشتر از طرح های تجربی است، چون پژوهش غیرتجربی منطبق با تعریف تجربی برای دستکاری کنترل شده متغیرهای مستقل نیست. در حقیقت در ماتریکس طرح ها، متناسب با هر نوع پژوهش، طرح های پژوهشی غیرتجربی وجود دارند (شکل ۲-۲). در این فصل، نمونه هایی از پژوهش های غیرتجربی منطبق با شش سلول "ماتریکس انواع پژوهش" ارائه خواهد شد. یکی از عواملی که بحث پژوهش غیرتجربی را پیچیده می کند، وجود اصطلاحات مختلف برای توصیف انواع پژوهش های غیرتجربی است. در جدول ۲-۱، تعریف اشکال مختلف پژوهش به طور خلاصه ارائه شده است. بعضی از اشکال پژوهش غیرتجربی؛ مثل گزارش مورد ها (Case report)، پژوهشهای کیفی (Qualitative research)، پژوهش های اپیدمیولوژیک، پژوهش پرسشنامه ای (Survey)، و پژوهش متودولوژیک، اهمیت بسیاری دارند. این نوع پژوهش ها، با هزینه ای نسبتاً پایین، اطلاعات ارزشمندی را فراهم می کنند. توجه به این نکته بسیار مهم است که از پژوهش های غیرتجربی نمی توان روابط علت و معلولی بین متغیرهای مستقل و وابسته را استنباط کرد.



شکل ۲-۱: طرح پژوهش تجربی و طرح پژوهش غیرتجربی

جدول ۱-۲: انواع پژوهش های غیرتجربی

---

- گزارش مورد (Case report): مستند کردن سیستماتیک یک واحد معین. معمولاً توصیفی از درمان یک فرد، و گاهی اوقات توصیف یک واحد اداری، آموزشی، و غیره.
- طرح کیس-کنترل (Case-control): یک طرح پژوهشی اپیدمیولوژیک که در آن گروه های افراد با و بدون یک بیماری یا ویژگی معین ("اثر") مقایسه می شوند تا تعیین گردد که آیا آنها به طور متفاوتی در معرض علل فرضی بیماری یا ویژگی بوده اند یا خیر؟ یعنی کمک می کند که تعیین کنیم، آیا اکسپوزر یا مواجهه (علت)، با یک پیامد (بیماری) مرتبط است؟
- طرح کوهورت (Cohort): یک طرح پژوهشی اپیدمیولوژیک که رو به جلو و از علت به اثر کار می کند. یعنی گروه هایی از افراد که تصور می شود ریسک های مختلفی برای توسعه یک بیماری یا ویژگی دارند را شناسایی کرده و آنها را در طی زمان مشاهده می کند تا تعیین نماید که کدام گروه از افراد، با احتمال بیشتری بیماری یا ویژگی مورد نظر را بروز می دهند.
- پژوهش همبستگی (Correlational research): انجام پژوهش برای تعیین روابط بین متغیرها.
- پژوهش تکاملی (Developmental research): مشاهده ی یک پدیده در طی زمان انجام می شود تا تاریخچه طبیعی پدیده مورد نظر مستند شود.
- پژوهش اپیدمیولوژیک (Epidemiological research): پژوهشی که بروز (incidence) یک بیماری یا ضایعه را مستند می کند، علل بیماری یا ضایعه را تعیین می کند، یا مکانیزم هایی را برای کنترل بیماری یا ضایعه ارائه می دهد.

- پژوهش ارزیابی (Evaluation research): انجام پژوهش برای تعیین اثر یک برنامه یا سیاست.
- پژوهش تاریخی (Historical research): پژوهشی که در آن وقایع گذشته مستند می‌شوند. زیرا در عین حالی که مورد توجه هستند، دیدگاهی فراهم می‌کنند که می‌تواند برای تصمیم‌گیری در زمان حال مورد استفاده قرار گیرد.
- متا-آنالیز (Meta-analysis): فرآیند پژوهشی که بوسیله آن نتایج پژوهشهای مختلف قبلی، به روشی کمی سنتز می‌شوند.
- پژوهش متودولوژیک (Methodological research): انجام پژوهش برای تعیین پایایی و اعتبار اندازه‌گیری‌های کلینیکی و پژوهشی.
- پژوهش نرماتیو (Normative research): پژوهشی که از نمونه‌های بزرگ برای فراهم کردن مقادیر طبیعی معیارهای مورد نظر استفاده می‌کند.
- پژوهش سیاست‌گذاری (Policy research): انجام پژوهش برای استفاده در سیاست‌گذاری و اجرا.
- پژوهش کیفی (Qualitative research): انجام پژوهش برای درک عمیق پدیده‌ای مشخص، که معمولاً در آن از مصاحبه و مشاهده استفاده می‌شود.
- آنالیز ثانویه (Secondary analysis): پژوهشی که در آن داده‌های جمع‌آوری شده، برای پاسخ دادن به سوالات پژوهشی جدید، مجدداً آنالیز میشوند.
- پژوهش پیمایشی، زمینه‌یابی، یا پرسشنامه‌ای (Survey research): پژوهشی که داده‌های آن با تکمیل پرسشنامه، یا پاسخ دادن به سوالات مصاحبه جمع‌آوری می‌شود.

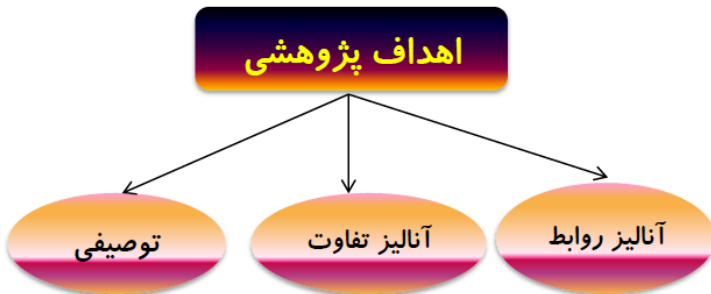
|           |                 | زمانبندی جمع آوری داده ها |           |
|-----------|-----------------|---------------------------|-----------|
|           |                 | گذشته نگر                 | آینده نگر |
| هدف پژوهش | توصیفی          | غیر تجربی                 | غیر تجربی |
|           | آنالیز روابط    | غیر تجربی                 | غیر تجربی |
|           | آنالیز تفاوت ها | غیر تجربی                 | غیر تجربی |
|           |                 | تجربی                     | تجربی     |
|           |                 | تجربی یا غیر تجربی        |           |
|           |                 | مانیپولاسیون              |           |

شکل ۲-۲: ماتریکس انواع پژوهش که سه بعد طرح پژوهش را نشان می دهد: هدف پژوهش (ردیف ها)، زمان جمع آوری داده ها (ستون ها)، و مانیپولاسیون (سلول ها).

همانطور که در ماتریکس انواع پژوهش (شکل ۲-۲) نشان داده شده است، پژوهش با اهداف توصیفی، بررسی روابط، و بررسی تفاوت ها انجام می گردد (شکل ۲-۳).

## پژوهش توصیفی

هدف پژوهش توصیفی مستند کردن ویژگی های یک پدیده از طریق جمع آوری سیستماتیک داده ها است. در پژوهش توصیفی، تصویری اجمالی از یک نمونه منفرد، با یک بار اندازه گیری (پژوهش مقطعی Cross-sectional)، یا چندین بار اندازه گیری و توصیف نمونه در طی یک دوره زمانی (پژوهش طولی Longitudinal) ارائه می شود.



شکل ۲-۳: اهداف پژوهش

در پژوهش های توصیفی معمول، متغیرهای مختلف بسیاری ثبت می شوند، ولی در این پژوهش، پیش فرض علت (Cause) و معلولی (Effect) وجود ندارد. بنابراین، در گزارش پژوهش های توصیفی، معمولاً بین متغیرهای مستقل و وابسته تمایزی انجام نمی شود، بلکه تمایز می تواند بین طرح های پژوهش توصیفی آینده نگر و گذشته نگر انجام شود.

### پژوهش توصیفی گذشته نگر

هدف پژوهش توصیفی گذشته نگر (Retrospective descriptive research، مستند کردن گذشته است (شکل ۲-۱). توصیف گذشته ممکن است برای ارزیابی زمان حال نسبت به گذشته و تصمیم گیری در حال حاضر بر پایه اطلاعات گذشته، استفاده شود. نوع رایج پژوهش های توصیفی گذشته نگر، اتکا بر داده های آرشیوی است. آرشیوها "اسناد ثبت شده برای بازگویی فعالیت افراد، موسسات، حکومت ها، و سایر گروه ها" میباشند. داده های آرشیوی معمولاً در پرونده های پزشکی، روزنامه ها و مجلات، راهنماهای تلفن، صورت جلسات، برنامه های خبری تلویزیون، و مجموعه ای از سایر منابع یافت می شوند. اطلاعات یافت شده در منابع آرشیوی باید از لحاظ مربوط بودنشان به موضوع پژوهش، به طور سیستماتیک توسط پژوهشگر آنالیز شوند. آنالیز محتوا یک

فرآیند پرزحمت است که شامل تعاریف عملی و قوانین تصمیم گیری برای رکوردها است، تا داده های مورد نظر استخراج گردد.

نمونه ای از تصمیم ها در پژوهشی دیده شد که پژوهشگران از داده های مستندات توانبخشی برای تعیین پیامد توانبخشی پس از ضایعه تروماتیک مغز استفاده کردند. ضابطه ی ورود افراد در پژوهش این بود که ضایعه تروماتیک مغزی "متوسط تا شدید" داشته باشند. تعریف عملی ضایعه مغزی عبارت بود از "آسیب اکستاسی مغز(غیر دژنراتیو و غیر مادرزادی)، که ناشی از یک نیروی فیزیکی خارجی بوده، باعث کاهش یا تغییر هوشیاری گردد، و توانایی های شناختی یا عملکرد فیزیکی فرد را مختل کند" قوانین تصمیم گیری نیز باید تعریف شوند تا پژوهشگران بتوانند تعیین کنند که کدام یک چارت های مرور شده، با همه بخش های این تعریف منطبق هستند. از آنجایی که از پیش تعیین کردن همه واریاسیون های احتمالی ممکن در پرونده های پزشکی مشکل است، تعاریف عملی معمولاً با کمک یک پروژه پایلوت استخراج داده ها تعیین می شود. بر پایه ی پژوهش پایلوت، قوانین تصمیم گیری قطعی و تعاریف عملی، قبل از شروع پژوهش اصلی تعیین می گردند.

علاوه بر استفاده ی گذشته نگر از داده های پرونده های پزشکی، پژوهشگران ممکن است داده های جمع آوری شده برای هدف دیگری را مجدداً آنالیز کنند (آنالیز ثانویه)، و برای این منظور باید مقدار قابل توجهی از داده ها را فیلتر کنند تا صحت (Accuracy) و مفید بودن داده ها را به حداکثر برسانند.

## پژوهش برای ارزیابی (Evaluation research)

پژوهش برای ارزیابی، نوع دیگری از پژوهش غیرتجربی توصیفی است که معمولاً به طور گذشته نگر انجام می شود. گاهی اوقات در این نوع پژوهش، ارزیابی پیامدهای یک برنامه در یک مرکز منفرد بررسی می شود. مثالی از این نوع پژوهش، ارزیابی برنامه

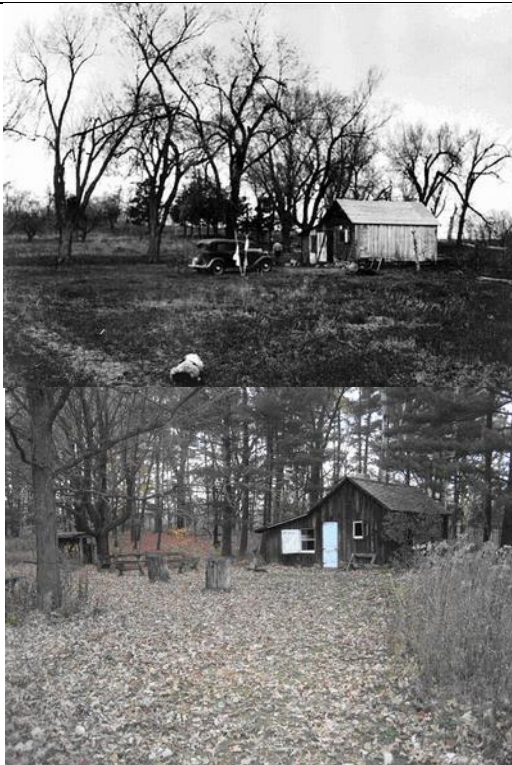
توانبخشی در مرحله حاد پس از تعویض کامل مفصل زانو در یک مرکز بیمارستانی است. گاهی ممکن است از این نوع پژوهش برای بررسی فرآیندهای درگیر در ارائه خدمت به گروه های بزرگی از بیماران در مراکز مختلف استفاده شود.

## پژوهش اپیدمیولوژیک

اپیدمیولوژی، بررسی بیماری، آسیب، و سلامتی در یک جمعیت می باشد. پژوهش اپیدمیولوژیک شامل روش های مختلفی (که همه آنها گذشته نگر نیستند) است که از آنها برای بررسی مسائل سلامتی مبتنی بر جمعیت استفاده می شود. در یک پژوهش، برای بررسی درمان افراد مسن دچار شکستگی ران، از روش های گذشته نگر استفاده شد. در این پژوهش، داده های حجم نمونه ی بسیار بزرگی از افراد مسن مبتلا به شکستگی ران (۲۶۴۳۴ نفر) برای تعیین ویژگی بیماران، روش جراحی به کار رفته برای ترمیم شکستگی، و نرخ زنده ماندن تا ۳ سال پس از شکستگی، بررسی شد. حجم نمونه بزرگ، که احتمالاً نماینده جمعیت بزرگی از افراد سالمند دچار شکستگی ران است، ویژگی مهمی است که پژوهش اپیدمیولوژیک را از پژوهش های کوچک تر (که از نمونه های در دسترس از یک یا چند بیمارستان استفاده می کنند)، متمایز می کند.

## پژوهش تاریخی (Historical research)

رویکرد پژوهشی دیگری که از روش های غیرتجربی توصیفی گذشته نگر استفاده میکند، پژوهش تاریخی است. هدف پژوهش تاریخی، مستند کردن وقایع گذشته است و بینشی فراهم می کند که می تواند برای هدایت تصمیم گیری های حال حاضر استفاده شود.



برای مثال، یک پژوهش تاریخی می تواند دستاوردهای افراد برجسته در یک حرفه را مستند کند (برای مثال ساخت حنجره مصنوعی). شکل دیگر پژوهش تاریخی، مستند کردن روند مراقبت های کلینیکی در طی زمان است، برای مثال، مرور گذشته نگر تقریباً ۲۰۰۰ بیمار درمان شده در برنامه های توانبخشی قلبی در طی یک دوره ۱۰ ساله. آنالیز تاریخیچه یا اثر سیاست های عمومی، شکل دیگری از پژوهش توصیفی گذشته نگر، به نام سیاست پژوهی است (Policy research). برای مثال بررسی تاریخچه قانونگذاری برای افراد دچار ناتوانی.

شکل دیگری از پژوهش مبتنی بر توصیف گذشته نگر، عبارت است از "گزارش مورد" (Case report) و "پژوهش موردی" (Case study). گرچه این اصطلاحات گاهی

اوقات بجای یکدیگر استفاده می شوند، اما آنها را باید از هم متمایز کرد. "گزارش مورد" با کنترل کمتر و نظام مندی (سیستماتیک) کمتری همراه است. مثلاً در یک گزارش مورد، اثر امواج ماوراء صوت منقطع در یک بیمار ۱۶ ساله مبتلا به سینوزیت مزمن که کاندید جراحی بود بررسی شد. بیمار به مدت ۲۰ جلسه تحت درمان با امواج ماوراء صوت قرار گرفت و نتیجه سی تی اسکن و بررسی علائم نشان داد که سینوزیت مزمن و علائم بیمار کاملاً بهبود یافته است.

## پژوهش توصیفی آینده نگر (Prospective descriptive research)

پژوهش توصیفی، پژوهشگر را قادر می سازد تا داده هایی که با هدف توصیف یک پدیده جمع آوری کرده است را کنترل کند. هدف پژوهش توصیفی، مستند کردن ماهیت یک پدیده از طریق جمع آوری سیستماتیک داده ها است. آینده نگر بودن جمع آوری داده ها باعث می شود که نتایج چنین پژوهشی نسبت به نتایج پژوهش های گذشته نگر قابل باورتر باشد (شکل ۱-۲). چهار روش پایه برای جمع آوری داده در پژوهشهای توصیفی آینده نگر عبارتند از:

- مشاهده (Observation)
- معاینه (Examination)
- مصاحبه (Interview)
- پرسشنامه (Questionnaire)
- مشاهده

مشاهده انواع مختلفی دارد. مشاهده می تواند کاملاً ساختارمند یا کمتر ساختارمند باشد. در یک پژوهش مشاهده ای، تکامل حرکات لگد زدن در نوزادان نارس و فول ترم با استفاده از فیلمبرداری بررسی شد. فرکانس لگد زدن و حرکات نوزادان آنالیز گردید و

داده ها در هفته ۱۲، ۱۸ و ۲۴ هفتگی جمع آوری شد. "اجزای توصیفی" این پژوهش، شامل توصیف الگوهای حرکتی در زمان های مختلف برای هر گروه بود. مقایسه الگوهای حرکتی در طی زمان و بین گروه ها "اجزای تحلیلی پژوهش" را تشکیل می دهند. پژوهش تکاملی (Developmental) اصطلاح دیگر برای این نوع پژوهش ها است. پژوهش تکاملی به این معنی است که نوزادان در بیش از یک نقطه از زمان توصیف می شوند تا اثرات گذشت زمان ثبت گردد.



## معاینه

نمونه ای از پژوهش های توصیفی آینده نگر، معاینه بیماران است. در یک پژوهش، پژوهشگران بازسازی رباط متقاطع قدامی (Anterior cruciate ligament, ACL) را در بیمارانی که حداقل ۴۰ سال داشتند بررسی کردند. در این پژوهش، بیمارانی که بازسازی ACL داشتند، به طور میانگین ۵۵ ماه پس از جراحی، معاینه شدند. نمونه های این پژوهش، از طریق مرور پرونده های پزشکی شناسایی شدند، اما داده های این پژوهش، به طور آینده نگر جمع آوری گردیدند تا به سوالات مربوط به وضعیت زانوها در پیگیری (Follow-up) پاسخ بدهند.

نمونه ای از پژوهش های توصیفی آینده نگر، در پژوهش های "متودولوژیک" مشاهده می شود. هدف پژوهش های متودولوژیک، مستند کردن و بهبود پایایی (Reliability) و اعتبار (Validity) تست های کلینیکی و پژوهشی است. در یک پژوهش، توانایی درمانگران لمس درمانی برای آشکارسازی میدان انرژی انسان، که ادعا داشتند قادر به آشکارسازی و دستکاری "میدان انرژی انسانها" با استفاده از دست های خود هستند) را بررسی کردند. برای بررسی توانایی درمانگران برای آشکار کردن میدان انرژی، ۲۱ نفر با سابقه ۱ تا ۲۷ سال، در پژوهش شرکت کردند. سوال این بود که آیا آنها واقعا می توانند میدان انرژی انسان را درک کنند؟ هر یک از افراد شرکت کننده در پژوهش، در حالی که پرده ای بین ایشان و معاینه کننده وجود داشت، کف دست هایش را رو به بالا نگه میداشت (شکل ۲-۴). معاینه کننده برای هر فرد، ۱۰ بار تست را تکرار می کرد. او هر بار یکی از دست هایش را روی یکی از دست های درمانگر نگه می داشت و از درمانگر خواسته می شد که معین کند کدام یک از دست هایش به دست های معاینه کننده نزدیک است. برای اعتبار تئوری لمس، درمانگران باید قادر می بودند که ۱۰۰٪ پاسخ صحیح بدهند. نتیجه نشان داد که اکثر درمانگران لمس درمانی قادر به تشخیص دست صحیح بطور پایا نبودند.



شکل ۲-۴: معاینه کننده دستش را روی یکی از دست های فرد نگه می دارد.

## مصاحبه

مصاحبه عبارت از یک گفتگوی ساختارمند بین پژوهشگر و بیمار/فرد می باشد که هدف آن جمع آوری اطلاعات است. در مصاحبه، هم پژوهشگر و هم بیمار/فرد، نقش هایی دارند که باید انجام بدهند. پژوهشگر کسب اطلاعات می کند و بیمار/فرد ارائه اطلاعات می کند. پژوهشگر می تواند برای کمک به فرد در ارائه اطلاعات، از تکنیک ها یا روش های مصاحبه مختلف استفاده کند. بعضی روش های مصاحبه موثرتر از بقیه است. در یک پژوهش بر روی افراد سالمند، داده ها به روش مصاحبه جمع آوری شد و از آنها در باره ماندن در خانه، فاکتورهایی که آنها را قادر می سازد مستقل باقی بمانند، موقعیت هایی که ممکن است استقلال آنها را به خطر بیاندازد، و منابعی که به آنها کمک خواهد کرد مستقل باقی بمانند، سوال شد. در پژوهش بر روی افراد سالمند، استفاده از روش مصاحبه نسبت به پرسشنامه ارجح است. زیرا سالمندان ممکن است مشکلات بینایی داشته و خواندن پرسشنامه برای آنها مشکل باشد. بعضی از سالمندان نیز سطح تحصیلات کمی دارند که خواندن پرسشنامه و نوشتن پاسخ ها را برای آنها مشکل می کند. از طرفی، بعضی سوالات ممکن است حاوی مفاهیم پیچیده ای باشند که پاسخ دادن به آنها در قالب پرسشنامه، احتمالاً فاقد عمق و وسعت لازم که در یک مصاحبه وجود دارد خواهد بود.

## پرسشنامه

در یک پرسشنامه (Survey)، سوالات و انواع پاسخ های ممکن به سوالات، از پیش تعیین شده اند. پژوهشگر مطابق با سوالات یا اظهارات از پیش تعیین شده در پرسشنامه عمل می کند، و معمولاً از پاسخ دهنده ها نیز خواسته می شود که از شیوه پاسخ از پیش تعیین شده پیروی کنند. برای مثال، سوالات یک پرسشنامه ممکن است از پاسخ دهنده ها بخواهد میزان موافقت یا مخالفت شان را با چند عبارت ذکر کنند یا ممکن است خواسته شود درجه اهمیت چندین موضوع را بر اساس یک مقیاس عددی درجه بندی کنند. پرسشنامه ها اغلب به صورت کتبی (یا امروزه به صورت الکترونیک) ارائه می شوند.

همچنین پژوهش پرسشنامه ای "زنده" به صورت تماس های تلفنی نیز امکان پذیر است. لازم به ذکر است که برای پاسخ دادن به سوالات پرسشنامه سطح کافی سواد خواندن و نوشتن باید وجود داشته باشد.

روش پرسشنامه ای در پژوهشی بر روی بزرگسالان مبتلا به فلج مغزی استفاده شد. البته از آنجایی که در پژوهش های پرسشنامه ای معمولاً پاسخ های افراد به صورت نوشتاری است، باید نمونه های پژوهش به بیماران بزرگسال مبتلا به فلج مغزی که ناتوانی های ذهنی مشخصی نداشتند محدود گردید. در نوع خاصی از پژوهش پرسشنامه ای، از تکنیک دلفی (Delphi) برای تعیین عقاید مدیران برنامه آموزشی فیزیوتراپی در باره فعالیتهای اساسی دانشجویان فیزیوتراپی استفاده شد. در تکنیک دلفی، مجموعه ای از پرسشنامه ها برای رسیدن به موارد مورد توافق گروهی از صاحب نظران در زمینه ای خاص، استفاده می شود. در روش دلفی چند دور پرسشنامه تهیه می شود. هر دور پرسشنامه ها بر پایه اطلاعات جمع آوری شده در دور قبلی ایجاد می گردند و عموماً از افراد خواسته می شود سطح توافق یا مخالفت شان را با عقاید گروه دور قبلی بیان کنند. پاسخ دهندگان گروه تحصیل کرده بودند و سطح مهارت خواندن و نوشتن آنها برای پاسخ دادن به پرسشنامه ها کافی بود. به علاوه، از آنجایی که نمونه ها در سطح ملی در پژوهش شرکت می کردند، بنابراین پست کردن پرسشنامه بسیار موثرتر و کم هزینه تر از انجام مصاحبه با افراد بود. مزیت دیگر پست کردن پرسشنامه این است که محرمانه بودن و بی نام بودن پاسخ ها می تواند لحاظ شود. در صورت بی نام نبودن پرسشنامه، افراد ممکن است پاسخ هایی بدهند که گمان می کنند از لحاظ اجتماعی قابل قبول تر هستند.

شیوه سوالات باز که بعداً لازم است آنالیز محتوی شوند تا تم های آنها استخراج شوند، با تعاریف پژوهش کیفی منطبق است.

## آنالیز روابط

دومین گروه اصلی از پژوهش های غیرتجربی شامل طرح هایی است که هدف اولیه آنها آنالیز روابط بین متغیرها است. در فرمت کلی این نوع پژوهش، یک گروه از افراد از لحاظ چندین متغیر مختلف تست می شوند و روابط ریاضی بین متغیرها بررسی می شود. این نوع پژوهش را پژوهش همبستگی می نامند (Correlational research). اصطلاح همبستگی برای یک تکنیک آماری خاص نیز استفاده می شود، ولی پژوهش همبستگی فقط شامل پژوهش هایی نیست که در آن از تکنیک آماری همبستگی استفاده شده است. آنالیز روابط نیز شامل پژوهش هایی است که فراتر از فقط استفاده از تکنیک های آماری همبستگی است. برای مثال، پژوهشگران اپیدمیولوژیک از مجموعه ای از نسبت ها (Ratio) برای بیان ارتباط بین ویژگی های فردی و وجود یا عدم وجود بیماری استفاده می کنند. بنابراین، اصطلاح همبستگی گرچه برای آنالیز آماری مربوط استفاده می شود، اما آنالیز روابط اصطلاح وسیع تری است که بطور دقیق تر برای توصیف یک نوع کلی از پژوهش بکار می رود. بنابراین هدف اولیه، آنالیز روابط بین متغیرها است.

## دلایل بررسی روابط

دلایل مختلفی برای شناسایی روابط بین متغیرها وجود دارد. این دلایل عبارتند از:

- روشن شدن حیطه های مفید برای انجام پژوهش های تجربی در آینده
- پیش بینی یک متغیر بر پایه نمرات متغیر دیگر
- تعیین پایایی ابزارهای اندازه گیری
- تعیین اعتبار یک معیار

- دلیل اول برای اثبات روابط بین متغیرها بدون اعمال مداخله و دستکاری، روشن کردن حیطه های مفید برای انجام پژوهش های تجربی در آینده است. این نوع پژوهش را

پژوهش اکتشافی نیز می گویند (Heuristic research)، زیرا هدف پژوهش، کشف یا آشکار کردن روابطی است که ممکن است به روشن شدن بیشتر موضوع منجر شود. ارزش پژوهش های اکتشافی نه لزوماً در نتایج سریع آن، بلکه در تعیین جهت حرکت پژوهشگر است.

- هدف اختصاصی دوم برای آنالیز روابط بین متغیرها، پیشینی یک متغیر بر پایه نمرات متغیر دیگر است یعنی اجازه می دهد نمرات یک متغیر را بر اساس نمرات متغیر دیگر پیش بینی کرد. در کلینیک، در گروهی از بیماران که اخیراً دوره درمان خود را کامل کرده اند، ارتباط مشخص و قوی بین یک سری ویژگی های معین از پذیرش تا ترخیص، ممکن است به پیش بینی بهتر وضعیت ترخیص بیماران در آینده کمک کند.

- هدف اختصاصی سوم در آنالیز روابط، تعیین پایایی ابزارهای اندازه گیری است. منظور از پایایی، میزان تکرارپذیری اندازه گیری ها است. در کلینیک، مداخلات ما بر پایه اندازه گیری ها یا مشاهدات معین است. برای مثال، اگر یک متخصص ارتز و پروتز نتواند تراز بودن سطح لگن را در بیماران دچار قطع اندام دریافت کننده پروتز جدید، به طور پایا تعیین کند، ممکن است یک روز پروتز را کوتاه کند و روز دیگر طویل، و سرانجام در روز دیگر تغییری در طول پروتز ندهد. یا اگر ترازو وزن افراد را بطور پایا مشخص نکند، یعنی امروز وزن را کم نشان دهد، در روز دیگر بالا، و در روز دیگر تغییری در وزن نشان ندهد، پزشک را در تشخیص و تصمیم گیری برای درمان دچار مشکل خواهد کرد. بنابراین، طراحی درمان و مداخلات در کلینیک، بر اساس اندازه گیری ها و مشاهدات تکرار پذیر است و تعیین پایایی اندازه گیری ها با استفاده از روشهای آماری، شاخصی از میزان اعتماد به معیارها است.

- دلیل چهارم برای آنالیز روابط بین متغیرها، تعیین اعتبار (Validity) یک تست یا معیار است.

اعتبار عبارت است از مناسب بودن، معنی دار بودن، و مفید بودن استنباطات خاص، از نمرات یک تست. در تعیین اعتبار یک تست جدید، نمرات تست جدید با نمرات تستی

استاندارد که قبلاً اثبات شده، مقایسه می شود تا میزان توافق نتایج تست ها به عنوان شاخصی از اعتبار تست جدید تعیین گردد.

## آنالیز گذشته نگر روابط

روابط را می توان به صورت گذشته نگر و از طریق بررسی پرونده های پزشکی یا آنالیز ثانویه داده های جمع آوری شده برای مقاصد دیگر، آنالیز کرد (شکل ۱-۲). در نمونه ای از آنالیز گذشته نگر روابط، پرونده ۱۸۲۶ نفر که از ۱۹۸۳ تا ۱۹۸۵ معاینه شده بودند، بررسی شد تا نسبت شانس وابستگی (Odds ratio) انجام کارهای عملکردی افرادی با یک بیماری خاص، نسبت به شانس وابستگی در آن کار در میان کل افراد تعیین گردد. نتایج این پژوهش نشان داد که احتمال وابستگی در بالا رفتن از پله، در مقایسه با کل افراد، برای افراد مبتلا به سکتة مغزی ۲/۸۲ برابر بیشتر و در مبتلایان به آرتروز زانو ۱/۹۸ برابر بیشتر بود.

## آنالیز آینده نگر روابط

آنالیز روابط اغلب به طور آینده نگر و با کنترل همزمان انتخاب افراد و روش استفاده از ابزارهای اندازه گیری، انجام می شود. نمونه ای از این پژوهش در بررسی رابطه نیروی عضله و ویژگی های راه رفتن در ۲۲ فرد دچار قطع اندام از زیر زانو به دنبال یک بیماری عروقی استفاده شد. نتایج این پژوهش رابطه بین قدرت عضلات بازکننده ران در سمت اندام قطع شده با سرعت راه رفتن، رابطه بین قدرت عضلات دورکننده ران در سمت سالم با تعداد قدم ها در دقیقه، و رابطه بین قدرت عضلات بازکننده زانو در هر دو سمت با طول گام را نشان داد. گرچه تعیین وسعت رابطه ی بین فاکتورها، مورد توجه درمانگران افراد مبتلا به قطع اندام از زیر زانو (ترانس تیال) است، ولی شناسایی روابط به این معنی نیست که دستکاری یک متغیر بر سایر متغیرها موثر خواهد بود، و بنابراین نمی توان استنباط علت

و معلولی از این روابط داشت. اما از نتایج این پژوهش می توان نتیجه گیری کرد که تقویت این گروه عضلات (بازکننده هاو دورکننده های ران) باعث بهبود توانایی راه رفتن (سرعت، تعداد قدم ها در دقیقه، و طول گامها) خواهد شد. از آنجایی که در این پژوهش، متغیرها بصورت کنترل شده دستکاری نشده اند، نمی توان استنباط علت و معلولی از این روابط داشت.

نمونه ای از پژوهش های آنالیز روابط عبارتند از:

- بررسی ارتباط چاقی و سطح ویتامین D با بیماری آسم کودکان
- بررسی ارتباط چاقی با بیماری آرتریت روماتوئید
- بررسی ارتباط سطح ویتامین D با بزرگ شدن لوزه در کودکان
- ارتباط نوع گروه های خونی و ابتلا به هیپاتیت

آنالیز آینده نگر روابط، برای اثبات اعتبار یک تست جدید نیز استفاده می شود. در یک پژوهش، برای بررسی ارزش معاینه فیزیکی در آشکار کردن پارگی لیگامان های طرفی مچ پا، در مقایسه با آرتروگرافی، پژوهشگران ۱۶۰ بیمار مراجعه کننده به بخش اورژانس با سابقه ضایعه حاد قسمت طرفی مچ پا را بررسی کردند. آرتروگرافی به عنوان معیار ضابطه یا "استاندارد طلایی" در نظر گرفته شده بود که نتایج معاینه فیزیکی با آن مقایسه شد. پژوهشگران رابطه بین معاینه فیزیکی و آرتروگرافی را با مفاهیم سنتی اپیدمیولوژیک حساسیت (Sensitivity)، ویژگی (Specificity)، ارزش پیش بینی مثبت (Positive predictive value)، و ارزش پیش بینی منفی (Negative predictive value) ارزیابی کردند. با مقایسه ی نتایج معاینه فیزیکی با آرتروگرافی، نتیجه گیری شد که نتایج معاینه فیزیکی مانند آرتروگرافی است و با ایجاد ناراحتی کمتر در بیمار، انجام می شود. این نتیجه، اعتبار معاینه فیزیکی را تایید می کند.

پژوهش نرماتیو (Normative research) شکل دیگری از پژوهش توصیفی است. در پژوهش های نرماتیو، گروه های بزرگی از افراد که نماینده جمعیت هستند معاینه می شوند تا مقادیر طبیعی برای متغیر مورد بررسی تعیین شود. در یک پژوهش توصیفی و

"نرماتیو"، پژوهشگران نیروی عضلات اندام فوقانی و تحتانی را با دینامومتر دستی در ۱۵۶ فرد سالم ۵۰ تا ۷۹ سال (۳ گروه سنی، حداقل ۲۵ مرد و ۲۵ زن در گروه سنی) برای تعیین مقدار نیروی نرمال اندازه گیری کردند (شکل ۲-۵). بیست و پنج نفر در هر یک از زیرگروه ها برای یک پژوهش نرماتیو کم است، اما با توجه به کمبود داده در این زمینه، این اطلاعات می تواند "نرماتیو" باشد، چون در حال حاضر فقط این داده ها موجود هستند.



شکل ۲-۵: پژوهش توصیفی نرماتیو برای تعیین نیروی عضلات اندام،

استفاده از دینامومتر دستی برای اندازه گیری نیروی عضله لترال روتاتور شانه اما باید توجه داشت که یک پژوهش ممکن است ترکیبی از طرح های پژوهشی را به کار بگیرد تا به سوالات پژوهش پاسخ دهد. در مثال فوق، گرچه هدف اصلی پژوهش فراهم کردن داده های نرماتیو برای نیروی طبیعی عضلات اندام ها است اما با توجه به بررسی تفاوت نیروی عضلات مورد بررسی (۱۳ گروه عضله) و مقایسه نیروی عضلات سمت غالب و مغلوب، یک نوع پژوهش تحلیلی با اندازه گیری های مکرر نیز هست.

همچنین از آنجایی که رابطه بین سن، جنس، وزن، و قد با نیروی عضلات دو سمت نیز مورد بررسی قرار گرفته است (تا با استفاده از آنالیز رگرسیون، معادلات پیش بینی نیروی عضله برای سمت غالب و مغلوب محاسبه گردد)، یک پژوهش با آنالیز روابط نیز هست. نتایج این پژوهش نشان داد که نیروی عضلات دو سمت غالب و مغلوب و نیروی بین گروه های عضلانی، تفاوت معنی دار داشته، و بین سمت و عمل عضله، تعامل معنی دار وجود دارد.

## آنالیز تفاوت

هدف کلی این نوع از پژوهش، تعیین تفاوت بین گروه ها و درمان ها به روشی پایا با پژوهش های تجربی و همچنین غیرتجربی است (شکل ۱-۲). آنالیز غیرتجربی تفاوت بین گروه ها یا درمان ها، پژوهش "پس از واقعیت" یا "پس رویدادی" ex post facto, (after the fact)

یا پژوهش سببی-مقایسه ای (Causal-comparative) نامیده می شود. در این نوع از پژوهش، متغیر مستقل دستکاری نشده، اما فرض می شود که علت تفاوت ها در متغیر وابسته است. در طرح پس رویدادی، تقسیم افراد به گروه ها تحت کنترل پژوهشگر نیست و بر اساس ویژگی های موجود افراد شرکت کننده در پژوهش تعیین می شود. به عبارتی، این متغیرها قابل دستکاری نیستند. باید توجه داشت که طرح پس رویدادی به این معنی نیست که سوالات بعد از جمع آوری داده ها ایجاد شده اند. جمع آوری داده ها در این طرح ها می تواند گذشته نگر یا آینده نگر باشد.

نمونه های بسیاری از آنالیز غیرتجربی تفاوت ها وجود دارد که در آنها متغیر مستقل ذاتا غیرقابل دستکاری است. برای مثال، بعضی از سوالات پژوهشی در باره تفاوت بین گروه هایی از افراد با ویژگی های کیفی (Attributes) مختلف هستند، مثلا مقایسه ورزشکارانی که ناتوانی های مادرزادی دارند با ورزشکارانی که ناتوانی های اکتسابی

دارند. ویژگی های دیگر مورد استفاده برای گروه بندی افراد به طور غیرتجربی، عبارتند از سن، جنس، و وجود یا عدم وجود بیماری.

در طرح پس از واقعیت، توضیح رابطه بین چیزی که رخ داده است و متغیرهای همبسته کنترل شده از طریق همسان سازی، برای نشان دادن علت و اثر (پیامد) است. گاهی شرایطی وجود دارد که در آن انجام آزمایش یا تجربه برای تعیین علت و اثر، غیرعملی، غیراخلاقی، یا غیرقانونی است.

چگونه می توان بدون آزمایش، شواهدی از رابطه علت و معلولی، فراهم کرد؟ در یک تجربه یا آزمایش معکوس "پس از واقعیت"، افراد با همسان سازی جفتی بر پایه متغیرهای مستقل، به گروه های "تجربی" و "کنترل" تقسیم می شوند. فرض کنید پژوهشگری می خواهد اثر سیگار کشیدن بر ریت اختلالات نوزاد را در میان زنان باردار ۲۰-۳۰ سال بررسی کند. در چنین حالتی، پژوهشگر نمی تواند و نیازی هم ندارد که از بعضی زنان باردار بخواهد سیگار بکشند و از بعضی دیگر بخواهد سیگار نکشند. در شرایط عادی، بعضی زنان در طی بارداری سیگار می کشند و بعضی سیگار نمی کشند. ما متغیر مستقل (سیگار کشیدن) را دستکاری نمی کنیم، اما می توانیم اثرات آن را "پس از واقعیت" بررسی کنیم. یعنی متغیرهای مستقلی که احتمالاً بر متغیر وابسته ما اثر دارند، شناسایی میشوند. بنابراین یک نمونه از زنانی که اخیراً زایمان داشته اند شناسایی شده و سپس بر اساس سیگار کشیدن و سیگار نکشیدن به دو گروه تقسیم می شوند. و در نهایت، کودکان دچار اختلالات در هر دو گروه شمارش، توصیف، و مقایسه می گردند.

## آنالیز گذشته نگر تفاوت ها

پرونده های پزشکی منبع گسترده ای از اطلاعات در باره بیماران و پیامدها هستند. وقتی بیمارانی که تحت دوره هایی از درمان ها قرار داشته اند یا ویژگی های مشترکی دارند،

بتوانند از پرونده های پزشکی شناسایی شوند بررسی رابطه بین درمان یا ویژگی معین آنها با یک پیامد، به صورت گذشته نگر ممکن خواهد شد.

در زیر نمونه طرح هایی که به طرق مختلف از طرح "پس از واقعیت" گذشته نگراستفاده کرده اند، توضیح داده می شوند.

در یک پژوهش گذشته نگر بزرگ در باره توانبخشی پس از جراحی زانو، پرونده های پزشکی بیش از ۵۰۰۰ نفر در یک سال مرور شد. چهار گروه از بیماران، درمان های زیر را دریافت کرده بودند: بدون ورزش، ورزش در خانه، ورزش ایزوتونیک، و ورزش ایزوکینتیک. این گروه ها، شامل بیماران درمان شده در یک بازه زمانی یکسان، اما با پروتوکل های توانبخشی مختلف بودند. این نوع پژوهش گذشته نگر با گروهی بزرگ از بیماران و با هدف تعیین پیامد های کلی درمان را پژوهش پیامد (Outcomes research) می گویند.

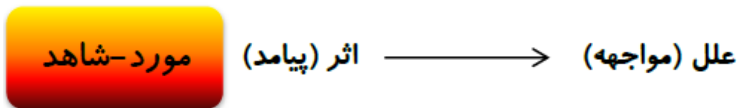
یک عدم مزیت طرح های گذشته نگر این است که به علت تقسیم غیر تصادفی افراد در گروه ها، تعیین دلیل قرار گرفتن یک بیمار در گروه درمانی خاص غیرممکن است. برای مثال، آیا آنهایی که می خواستند به رقابت های ورزشی برگردند در گروه ایزوکینتیک قرار گرفتند، و بنابراین سوگیری برای بیماران با انگیزه وجود داشته است؟ اگر در پرونده های پزشکی، اطلاعاتی در باره اهداف و انگیزه بیمار نباشد، هیچ راه دیگری وجود ندارد که تعیین کنیم آیا این سوگیری وجود داشته است یا خیر.

نمونه دوم طرح گذشته نگر پس رویدادی، بررسی اثر تغییر در سیاست استفاده از محدود کننده های فیزیکی برای سالمندان بستری بر تعداد افتادن ها و شدت ضایعات ناشی از افتادن می باشد. برای این منظور، حوادث یک مرکز مراقبت سالمندان، قبل و بعد از سیاست بدون محدودیت در گروه هایی متوالی از افراد سالمند بررسی شد. نتیجه این پژوهش نشان داد که قبل و بعد از تغییر سیاست، تعداد افتادن ها یکسان بود اما شدت ضایعات با سیاست بدون محدودیت کاهش یافت. تعداد افتادن ها با ضایعات کمتری همراه بود و تعداد افتادن های بدون ضایعه بالاتر بود. این یافته ها بیانگر آن بود که افتادن

در سالمندان به دلیل تغییرات ناشی از فرآیند سالمندی ادامه خواهد داشت. اما رفع موانع و محدودیت های فیزیکی از سالمندان و اجازه آزادی حرکت، شدت ضایعات ناشی از افتادن را کاهش می دهد.

تفاوت بین این پژوهش و پژوهش قبلی در باره توانبخشی پس از عمل زانو این است که پاسخ به سوال این پژوهش نیازمند نمونه گیری بیماران از دو نقطه در زمان بود. نقطه قوت پژوهش قبلی در بررسی توانبخشی زانو، استفاده از بیماران درمان شده در یک چهارچوب زمانی منفرد بود.

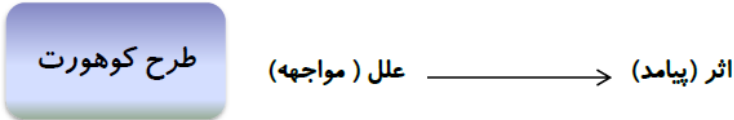
نمونه سوم طرح گذشته نگر پس رویدادی، از مدل "مورد-شاهد" (کیس-کنترل) پیروی می کند. در این طرح، گروهی از بیماران با «اثر» مشخص شناسایی گردیده، و سپس یک گروه «بدون اثر» شناسایی می شوند. آنگاه «علل فرضی» برای اثرات بررسی می شوند، و نسبت بیماران با آن علل در دو گروه مقایسه می شوند (معلوم بودن اثر، در جستجوی علل فرضی) (شکل ۲-۶).



شکل ۲-۶: طرح مورد-شاهد (کیس-کنترل)

نمونه یک طرح مورد-شاهد، پژوهش در باره بررسی اثر فیزیوتراپی تنفسی و ضایعه مغزی در نوزادان بسیار نارس می باشد. در این پژوهش، نوزادان بسیار کم وزن دچار ضایعه مغزی (مورد یا کیس) با نوزادان همسان بدون ضایعه مغزی (شاهد یا کنترل) مقایسه شدند. سپس پژوهشگران با برگشت به عقب، تعداد جلسات فیزیوتراپی تنفسی هر کودک در طی ماه اول زندگی را مستند کردند. آنها دریافتند که در گروه نوزادان کیس با ضایعه مغزی، تعداد جلسات فیزیوتراپی تنفسی در مقایسه با نوزادان بدون ضایعه مغزی دو تا سه برابر بالاتر بود. اعتماد به پژوهش های مورد-شاهد به میزان زیادی به ضوابط مورد استفاده برای تعریف گروه های مورد و شاهد بستگی دارد.

نمونه چهارم گروه بندی افراد در طرح های گذشته نگر پس رویدادی، از طرح "کوهورت" (Cohort) پیروی می کند. در طرح کوهورت که گاهی "متضاد" طرح مورد-شاهد در نظر گرفته می شود، گروهی از افراد در طی زمان پیگیری می شوند، این طرح ها با «علل فرضی» شروع شده و برای تعیین «اثرات» جستجو می شود (شکل ۲-۷).



شکل ۲-۷: طرح کوهورت

در یک طرح کوهورت گذشته نگر، در زمان طرح سوال پژوهش، همه داده ها جمع آوری شده اند: محقق داده های مناسب را استخراج می کند، بیماران بر اساس متغیر یا متغیرهای مورد نظر گروه بندی می شوند، و تفاوت در مجموعه ای از متغیرهای پیش بینی کننده بررسی می شود. این رویکرد در پژوهش، برای بررسی پذیرش مجدد بیماران در یک مرکز توانبخشی کودکان بستری استفاده شد. مرور گذشته نگر پرونده های کودکانی که به اکسیژن و ونتیلاتور نیاز داشتند انجام شد و سپس کودکان بر طبق پذیرش مجدد بعدی گروه بندی شدند. کودکانی که مجددا پذیرش شدند و کودکان پذیرش نشده، مجددا مقایسه گردیدند. نتیجه این پژوهش نشان داد کودکانی که مجددا پذیرش شدند نیاز به ونتیلاتور بیشتر، مراقبت پرستاری، و ترانسفرهای حاد مراقبتی داشتند.

نمونه نهایی آنالیز گذشته نگر تفاوت بین گروه ها، تکنیک پژوهشی تخصصی به نام متاآنالیز است. این تکنیک یکی از سه رویکرد عمومی سنتز نتایج پژوهشها است:

- مقاله مروری روایتی (Narrative reviews)

- مرور سیستماتیک

- متاآنالیز

متاآنالیز یک راه کمی برای سنتز نتایج پژوهش های مختلف در باره یک موضوع معین است. منطق متاآنالیز این است که اندازه تفاوت بین گروه های درمان (اندازه اثر)، به طور

ریاضی استاندارد می شود. یعنی نتایج پژوهش های قبلی با متغیرهای وابسته مختلف اما مرتبط می توانند مقایسه شوند. دو گروه پژوهشی مستقل، برای سنتز شواهد موثر بودن کاردرمانی و فیزیوتراپی برای درمان افراد مبتلا به پارکینسون، از متاآنالیز استفاده کردند. نوزده مقاله در باره کاردرمانی بیماری پارکینسون مرور شد و اثرات مثبت برای پیامدهای مرتبط با ظرفیت ها و توانایی ها، همچنین عملکرد بیمار در فعالیت ها و کارهای مختلف، مشخص شدند. دوازده مقاله نیز در باره فیزیوتراپی پارکینسون مرور شد و اثرات مثبت برای فعالیت های زندگی روزمره، سرعت راه رفتن، و طول گام مشخص شد. در گفتار درمانی، متاآنالیز ۲۷ پژوهش از ۵۷ کارآزمایی کنترل شده تصادفی در مبتلایان به سکته مغزی دچار آفازی، شواهدی در باره اثربخشی مداخلات گفتاردرمانی در بهبود ارتباط عملکردی، خواندن، نوشتن، و زبان بیانی در مقایسه با عدم انجام گفتاردرمانی فراهم کرد. متاآنالیز ۸۱ پژوهش، یک رابطه معکوس بین فعالیت فیزیکی و خطر دیابت نوع ۲ نشان داد. در نمونه ای دیگر، متاآنالیز ۱۸ پژوهش با ۳۸۹۹۱ بیمار، اثرات فعالیت فیزیکی بر پیشرفت نوروپاتی دیابتی را بررسی کرد. نتیجه این متاآنالیز نشان داد که فعالیت فیزیکی برای بهبود نوروپاتی دیابتی و آهسته کردن پیشرفت آن موثر است.

در مقالات مروری روایتی (Narrative reviews)، پژوهشگر بر اساس منابع علمی، یک موضوع و مسئله مشخص را توصیف می کند. پژوهش های مربوط به یک موضوع مشخص شناسایی گردیده و مسئله مورد نظر توصیف می شود. در این نوع مقالات، سوال پژوهشی از پیش تعیین شده یا استراتژی جستجوی مشخص وجود ندارد، سیستماتیک نیستند و از یک پروتکل مشخص پیروی نمی کنند. برای مرور مقالات هیچ استاندارد یا راهنمایی وجود ندارد. در یک مقاله مروری روایتی، نویسندگان استراتژی های تشخیصی و ارزیابی اختلال بلغ (دیس فاژی) در بیماران مبتلا به اسکروز متعدد را توصیف و تفسیر کردند. مرور سیستماتیک نیازمند استراتژی های جستجوی مستند شده و ضوابط ورود و خروج مشخص است و انواع سوگیری را کاهش می دهد.

## آنالیز آینده نگر تفاوت ها

آنالیز آینده نگر تفاوت ها آخرین سلول ماتریکس شش سلولی انواع پژوهش است. این سلول تنها سلولی است که مشترک بین طرح های تجربی و غیرتجربی است. طبق تعریف، طرح های تجربی باید آینده نگر باشند، و هدف آنها تعیین اثرات مداخلات بر یک متغیر وابسته بوسیله آنالیز تفاوت بین گروهی یعنی تفاوت بین گروه هایی که مداخله ای دریافت کرده اند و گروه هایی که مداخله ای دریافت نکرده اند یا تفاوت ها در یک گروه که بیش از یک درمان تجربی دریافت کرده اند، می باشد. تفاوت های بین گروه ها یا درون یک گروه، همچنین می تواند بدون دستکاری کنترل شده انجام گردد.

نمونه آنالیز تفاوت ها در یک طرح غیرتجربی، مقایسه تحرک دینامیک سریع با بی حرکتی بعد از ترمیم تاندون خم کننده دست است. برای انجام این مقایسه، پژوهشگران افراد را بطور گذشته نگر با مرور پرونده های پزشکی آنها انتخاب کردند تا تعیین کنند چه کسانی بعد از عمل جراحی بی حرکت شدند و چه کسانی آتل دینامیک را دریافت کردند. اگرچه تقسیم افراد به گروه های درمان قبل از طرح سوال انجام شده بود، اما پژوهشگران داده ها را 4/5 ماه پس از جراحی جمع آوری کردند. مانند مقایسه های دیگر ما بین پژوهش های گذشته نگر و آینده نگر، ما می توانیم اعتماد بیشتری به این پژوهش داشته باشیم زیرا آنالیز داده های آینده نگر به پژوهشگران اجازه می دهد معیارهای استفاده شده در پژوهش را استاندارد کنند. اگر پژوهش کاملاً گذشته نگر بود، یعنی تعیین گروه ها و متغیرهای وابسته از مدارک پزشکی استخراج شده بودند، یکسانی معیارها مورد سوال بودند. اگر پژوهش کاملاً آینده نگر بود، یعنی تقسیم افراد به گروه های درمان تصادفی بود، و سپس اندازه گیری ها انجام می شد، پژوهش دیگر غیرتجربی نبود زیرا دستکاری مداخلات و ارائه درمان تحت کنترل پژوهشگران بوده است.

مقایسات آینده نگر افراد با یا بدون بیماری مشخص تحت عنوان طرح مورد-شاهد (کیس-کنترل) مقطعی نامیده می شود. این طرح مشابه طرح کیس-کنترل کلاسیک است

که در آن پژوهشگر کیس ها را شناسایی می کند و سپس برای افراد، کنترل همسان از لحاظ ویژگی های مهم مثل سن را جستجو می کند. اما با طرح کیس-کنترل کلاسیک متفاوت است و در آن جستجو رو به عقب برای علل احتمالی انجام نمی شود. بلکه پژوهشگران داده ها را بطور مقطعی و آینده نگر جمع آوری می کنند تا تفاوت دو گروه دارای بیماری و بدون بیماری را در متغیرهای دیگر تعیین کنند. در مقایسه وضعیت قدرت اندام تحتانی بین کودکان فلج مغزی اسپاستیک و کودکان سالم بدون فلج مغزی همسان شده از لحاظ سن، از طرح کیس-کنترل مقطعی استفاده شد.

همچنین، برای طرح کوهورت گذشته نگر یک همتای آینده نگر وجود دارد. در یک طرح کوهورت آینده نگر، پیامدهای بیمار محور پس از جراحی کمر بررسی شد. جراحان بیماران را قبل از جراحی وارد پژوهش کردند و پژوهشگران ویژگی های زمینه ای اجتماعی، علائم و نشانه های قبل از جراحی، تشخیص، و رویکردهای جراحی را ثبت کردند. سپس داده های پیامد شامل عملکرد، کیفیت زندگی، و رضایت کلی از درمان را ۱ سال پس از جراحی جمع آوری کردند. تفاوت بین گروه های بهتر شده و بدتر شده آنالیز شد و دریافتند که سن بالاتر و جراحی قبلی کمر، با پیامدهای بدتر همراه بودند. آینده نگر بودن این پژوهش کوهورت به این معنی است که جمع آوری داده های پیش بینی کننده و پیامد، هر دو تحت کنترل دقیق پژوهشگران بودند.

## خلاصه

بر خلاف پژوهش تجربی، پژوهش غیرتجربی نیازمند دستکاری کنترل شده متغیرها نیست. بر طبق این تعریف، مجموعه زیادی از طرح های پژوهشی غیرتجربی وجود دارد. در پژوهش های توصیفی، برای توصیف پدیده مورد نظر، از جمع آوری داده ها بصورت گذشته نگر یا آینده نگر استفاده می شود. در پژوهش آنالیز روابط، پژوهشگران از جمع

آوری داده ها بصورت آینده نگر یا گذشته نگر ، برای سنجش متغیرها و تعیین عوامل پیش بینی کننده، نسبت شانس (Odd)، و پایایی یا اعتبار معیارها استفاده می کنند.

## فصل سوم

طرح های اپیدمیولوژیک

بخش های فصل

طرح های اپیدمیولوژیک غیر تجربی

پژوهش های مقطعی

پژوهش های مورد-شاهد (کیس-کنترل)

پژوهش های کوهورت

خلاصه

## طرح های اپیدمیولوژیک

اپیدمیولوژی عبارت است از بررسی میزان گسترش و عوامل تعیین کننده حالات سلامتی و بیماری در جمعیت های انسانی. بر طبق این تعریف، پژوهشهای اپیدمیولوژیک نقش توصیفی (بررسی توزیع سلامتی و بیماری) و همچنین تحلیلی (بررسی عوامل موثر در سلامتی و بیماری) دارند. پژوهش های سلامت، شامل هر گونه پژوهش مهم برای حرفه های مختلف می باشند. در این تعریف، عنصر تشخیص (Diagnosis) و غربالگری (Screening) نهفته است که هر دو برای تمایز بین افراد سالم و بیمار ضروری هستند. عنصر دیگر در اپیدمیولوژی، تاکید بر خدمات پزشکی، جراحی، و سایر خدمات سلامت برای درمان و پیشگیری از بیماری ها در سطح جامعه است. همچنین بر نقش متخصصین حرفه های مختلف از جمله توانبخشی (فیزیوتراپی، کاردرمانی، گفتاردرمانی، و سایر) در پیشگیری از ناتوانی (Disability) نیز تاکید دارد. جزء مهم دیگر در این تعریف، تاکید بر جمعیت ها است یعنی پژوهش های اپیدمیولوژیک شامل گروه های بزرگی از مردم

است. گرچه بعضی از جمله روان درمانگر ها (Psychotherapist) تمایل دارند مشکلات را در سطح فرد بررسی کرده و درمان کنند، اما برای پیشگیری بر "مدل سلامت همگانی" (Public Health Model) یعنی جمعیت بزرگی از مردم تاکید می شود. بنابراین روشن است که اپیدمیولوژی کاربردهای مهمی در سلامت عموم و جامعه دارد.

در پژوهش های اپیدمیولوژیک، ضمن بررسی گروه های بزرگ از مردم، ویژگی های آنها با آنالیز روابط میان فاکتورهای دموگرافیک و سلامتی مختلف که در زیرگروه های معین یا در طی زمان آشکار می شوند، توصیف می شود. هر دو این ویژگی ها در پژوهش های اپیدمیولوژیک ماهیت غیرتجربی دارند، زیرا وضعیت سلامتی و بیماری موجود را مستند می کنند، و برای تغییر این حالات (مثلا با اعمال یک درمان کنترل شده) تلاشی نمی کنند. علیرغم ماهیت غالبا غیرتجربی، بعضی پژوهش های اپیدمیولوژیکی نیز ممکن است ماهیت تجربی داشته باشند.

در این فصل، طرح های اپیدمیولوژیکی غیرتجربی مورد استفاده برای مقایسه سلامتی و بیماری (که به طور طبیعی در گروه های مختلف رخ می دهند) ارائه و توضیح داده میشوند.

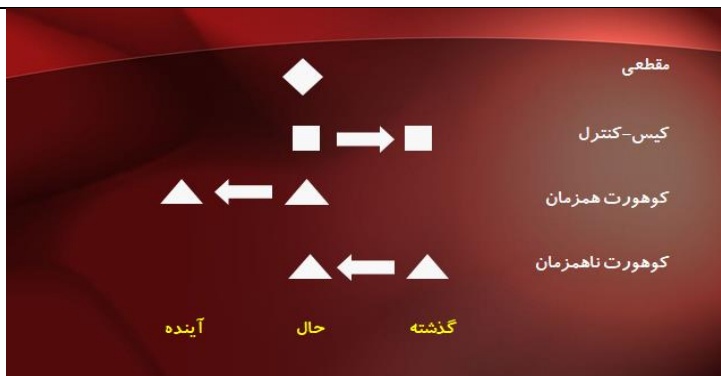
## طرح های اپیدمیولوژیک

سه طرح رایج غیرتجربی که در پژوهش های اپیدمیولوژیک استفاده می شوند عبارتند از:

۱- مقطعی (Cross-sectional)

۲- مورد-شاهد (Case-control)

۳- کوهورت (Cohort)



شکل ۳-۱: طرح های اپیدمیولوژیک با زمان اندازه گیری ها و همچنین جهت پیشرفت طرح مشخص می شوند.

برای پژوهش مقطعی، داده ها در یک نقطه از زمان اندازه گیری می شوند. در طرح های کیس-کنترل، جهت پژوهش از اثرات به علل، و در طرح های کوهورت، جهت پژوهش از علل به اثرات است.

طرح های مقطعی، کیس-کنترل، و کوهورت را تحت عنوان طرح های "تحلیلی مشاهده ای" نیز طبقه بندی می کنند. شکل ۳-۱ ارتباط هر یک از طرح ها را با زمان عناصر مختلف طرح نشان می دهد. قبل از توضیح هر یک از طرح ها، باید ترمینولوژی مورد استفاده در پژوهش های اپیدمیولوژیکی توضیح داده شود.

دو اصطلاح رایج که در اپیدمیولوژی استفاده می شوند عبارتند از: مواجهه (Exposure) و پیامد (Outcome). عبارتهای معادل مواجهه، "علت" و "ریسک فاکتور" یا متغیر مستقل است. عبارتهای معادل پیامد نیز، "بیماری" و "اثر" یا متغیر وابسته است (شکل ۳-۲).



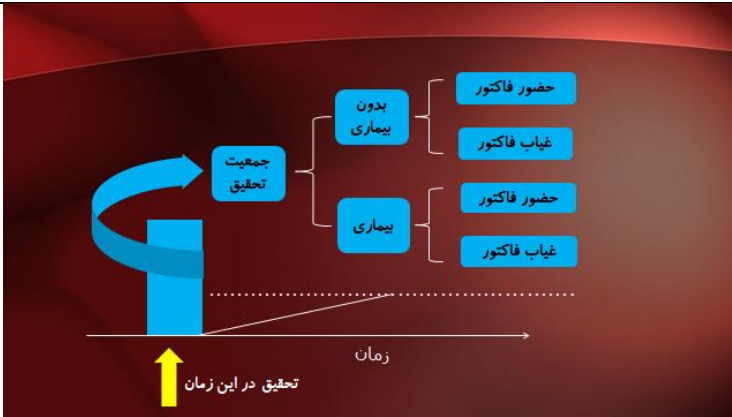
شکل ۳-۲: ترمینولوژی مورد استفاده در پژوهش های اپیدمیولوژیکی

## پژوهش های مقطعی

پژوهش های مقطعی برای مستند کردن وضعیت یک گروه، در یک نقطه مشخص از زمان استفاده می شوند. به عبارتی، پژوهش های مقطعی پژوهشی مشاهده ای هستند که متغیرهای مختلف و داده های جمعیت مورد نظر را در مقطعی از زمان اندازه گیری میکنند تا وضعیت سلامتی هر یک از افراد آزمودنی را در آن مقطع زمانی مشخص نمایند. پژوهش های مقطعی، تصویری از «علل» و «پیامدها» را در یک جمعیت و در یک مقطع زمانی معین، فراهم میکند و مدل اصلی پژوهش های اپیدمیولوژیکی، توصیفی است. در پژوهش های مقطعی، الگوی رخداد بر اساس زمان، مکان، و فرد، بررسی گردیده و تخمین فراوانی بیماری یا شیوع (Prevalence) و روند گسترش زمانی آن توصیف می شود. پژوهش های مقطعی برای برنامه ریزی، توزیع منابع، و تولید فرضیات، مفید هستند. در پژوهش مقطعی توصیفی، شیوع یک یا چند پیامد سلامت در یک جمعیت معین مشخص می شود. در پژوهش مقطعی تحلیلی، داده های مربوط به «علل» و «پیامدها»، در یک زمان

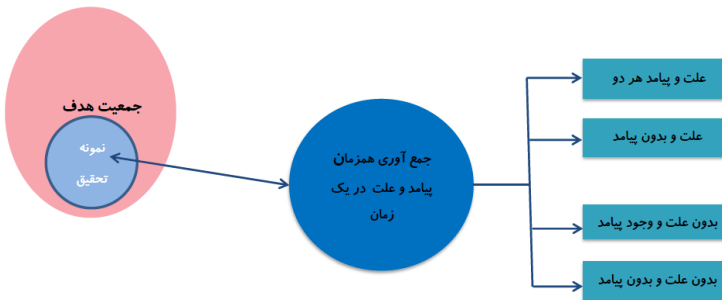
مشخص و با هدف مقایسه تفاوت پیامدها در افراد دارای «علت» و افراد بدون «علت»، جمع آوری می شود. "شیوع" عبارت است از نسبت افراد یک جمعیت که در یک زمان معین، یک بیماری یا ویژگی مشخص را دارند (صرف نظر از این که آنها از چه زمانی بیماری را برای اولین بار نشان دادند). و "بروز" (Incidence) عبارت است از تعداد مورد های جدید یک ویژگی یا بیماری، در یک دوره زمانی معین. در یک پژوهش مقطعی مکرر، برای بررسی روند شیوع بیماری ریوی انسدادی مزمن در اسپانیا، دو پژوهش اپیدمیولوژیک در سال های ۱۹۹۷ (تعداد ۴۰۳۰) و ۲۰۰۷ (تعداد ۳۸۰۲) انجام شد. نتایج، کاهش شیوع از ۹/۱٪ در سال ۱۹۹۷ به ۴/۵٪ (۴/۵٪/۵۰٪ کمتر) در سال ۲۰۰۷ و تغییر توزیع شدت بیماری ریوی انسدادی مزمن در اسپانیا را نشان داد.

در پژوهش های مقطعی، تصویری از یک گروه از افراد یا وضعیت موجود ارائه می شود و در آن «پیامدها» و «علل»، در افراد پژوهش بطور همزمان اندازه گیری می شوند (عدم پیگیری آینده نگر یا گذشته نگر) (شکل ۳-۳). در پژوهش های مقطعی، شیوع یک بیماری در پژوهش کلینیکی، توزیع متغیرها در یک جمعیت، و رابطه یک "علت" با یک "پیامد" در جمعیت بررسی می شود. اما رابطه علت و معلولی استنباط نمی شود. در یک پژوهش مقطعی توصیفی در باره شیوع اختلال عملکرد تنفسی در بیماران درمان شده مبتلا به آسم، از ۴۳۸۱ نفر در دامنه سنی ۱۷ تا ۶۵ سال، ۳۰۷ نفر ضوابط ورود را داشتند. معیار پیامد، تکمیل پرسشنامه توسط بیماران بود. بر طبق نتایج، ۲۲۷ نفر پرسشنامه را برگرداندند (نرخ پاسخ، ۷۴٪). حدود یک سوم زنان و یک پنجم مردان اختلال در عملکرد تنفسی داشتند.



شکل ۳-۳: طرح پژوهش مقطعی

انتخاب افراد در پژوهش های مقطعی، از جمعیت در دسترس انجام می شود. به مجرد انتخاب نمونه افراد از جمعیت، پژوهشگران داده ها را جمع آوری و رابطه بین «پیامدها» و «علل» را ارزیابی می کنند (شکل ۳-۴).



شکل ۳-۴: پژوهش مقطعی برای بررسی رابطه بین پیامد و علل

در پژوهش مقطعی تحلیلی برای بررسی رابطه بین سیگار کشیدن غیرفعال و علائم تنفسی، داده های ۷۸۸۲ فرد بزرگسال که هرگز سیگار نکشیده بودند، از ۳۶ مرکز در ۱۶ کشور اروپایی جمع آوری و آنالیز شدند. داده ها از طریق مصاحبه ساختارمند جمع آوری شد. معیارهای پیامد عبارت بودند از اسپیرومتری، تست چالش متاکولین، و IgE. از تست

آماري آناليز رگرسيون براي بررسي رابطه بين علائم تنفسي و عملکرد ريه استفاده شد. نتايج نشان داد كه شيوع سيگار كشيدن غير فعال در محل كار، از ۲/۵٪ تا ۵۳/۸٪ متغير بود. ارتباط معني دار بين سيگار كشيدن غير فعال با سفتي شبانه سینه، تنگي نفس شبانه، تنگي نفس بعد از فعاليت، و افزايش پاسخ برونش وجود داشت.

از پژوهش های مقطعی برای مستند کردن وضعیت سلامتی هر نمونه پژوهش، در یک نقطه منفرد از زمان استفاده می شود. در بعضی موارد، ممکن است نقطه‌ی زمانی، دامنه‌ی از تاریخ های تقویمی باشد (پژوهش بین فروردین و اسفند یک سال)، یا ممکن است یک واقعه مشخص باشد (ورود به دوره سالمندی در ۶۵ سالگی). در اغلب پژوهش هایی که در آنها وضعیت سلامتی مستند می شود، مقطع زمان، نقطه ورود به سیستم سلامت است. برای مثال، در پژوهش در باره سندروم پیچ خوردگی حاد مچ پا، ۶۰۰ بیمار مراجعه کننده به درمانگاه اورژانس بیمارستان در یک دوره ۳۳ ماه مورد ارزیابی قرار گرفتند. اگرچه پژوهش طی یک دوره تقریباً ۳۳ ماه انجام شد، هر یک از بیماران فقط یک بار ارزیابی شدند. "نقطه‌ی از زمان" که این پژوهش را مقطعی می کند "وقوع" مراجعه به درمانگاه اورژانس بیمارستان با شکایت از پیچ خوردگی مچ پا می باشد.

بسیاری از گزارش های مربوط به پژوهش های کیس-کنترل و کوهورت، شامل یک جزء مقطعی می باشند که گروه های بیمار را در یک نقطه منفرد از زمان توصیف می کنند. برای مثال، در یک پژوهش در باره سکنه مغزی، ناتوانی، و دمانس افراد سالمند روستاهای ایتالیا که با هدف اصلی مستند کردن رابطه بین سکنه مغزی و دمانس (طرح کوهورت ناهمزمان) انجام شد، پژوهش مقطعی بخشی از طرح بود و نتایج شامل سن و نرخ شیوع اختصاصی جنسیت برای سکنه مغزی در جمعیت مورد بررسی گزارش شد.

## طرح پژوهش مقطعی مکرر یا سریال

در این نوع پژوهش، جمع آوری داده ها از یک جمعیت، در زمان های مختلف انجام میشود. در هر زمان، نمونه ها یا افراد مختلفی از جمعیت هدف در پژوهش شرکت دارند. برای مثال، در یک پژوهش پرسشنامه ای سالانه، شرکت کنندگان در سال جدید متفاوت از افرادی هستند که سال قبل در پژوهش شرکت کردند. چنین داده ای می تواند یا بصورت مقطعی و با بررسی داده ها در یک سال پژوهشی آنالیز شود (آنالیز مقطعی مکرر داده ها) یا برای آنالیز در طی زمان، داده ها ترکیب شوند. از آنجایی که داده های مقطعی مکرر، مربوط به نمونه های مختلفی از جمعیت در طی زمان می باشد، می توان از آنها برای آنالیز تغییرات یک جمعیت یا گروه در طی زمان، استفاده کرد که "مجموع تغییرات در طی زمان" نیز نامیده می شود. نمی توان از این نوع پژوهش برای بررسی تغییرات فردی (مثل پژوهش کوهورت) استفاده کرد. با تغییر داده مجموع، ما به تغییرات گروه های مختلف اشاره می کنیم. اگر نمونه های نماینده جمعیت، در سال های متوالی یک پژوهش حاضر باشند، ما می توانیم تغییرات رفتار یا شرایط گروه های مختلف را مقایسه کنیم. برای مثال، ما می توانیم نتیجه گیری کنیم که چگونه سطوح سیگار کشیدن مردان و زنان در طی زمان تغییر کرده است. اما نمی توانیم استنباط کنیم که چگونه رفتار سیگار کشیدن یک جمعیت معین در طی زمان تغییر کرده است، زیرا در سال های مختلف، افراد مختلفی نمونه ها را تشکیل می دهند. بررسی شیوع بیماری ریوی انسدادی مزمن در اسپانیا یک پژوهش مقطعی مکرر است که در آن نتایج داده های دو گروه در سالهای ۱۹۹۷ و ۲۰۰۷ جمع آوری و مقایسه شدند و نتایج آن موید کاهش شیوع بیماری و تغییر توزیع شدت بیماری بود.

با گزارش کامل و شفاف اطلاعات مربوط به جنبه های مختلف یک پژوهش مقطعی، می توان مفید بودن یافته ها و خطر سوگیری را ارزیابی کرد. برای گزارش نتایج پژوهش های مقطعی، از راهنمای STROBE (Strengthening the Reporting of

Observational studies in Epidemiology) استفاده می شود که شامل چک لیستی با ۲۲ آیت است و قدم های لازم برای هدایت این نوع پژوهش ها را فراهم می کند. راهنمای استروب، پیشنهاداتی برای مقدمه، روش، سوگیری (بایاز)، حجم نمونه، متغیرها، روش های آماری، نتایج، بحث، و سایر اطلاعات از جمله حامی مالی ارائه می دهد.

## حجم نمونه

محاسبه حجم نمونه در طراحی پژوهش مقطعی اهمیت دارد. محاسبه حجم نمونه برای طرحهای مقطعی توصیفی و تحلیلی متفاوت است. در پژوهش مقطعی توصیفی، هدف تخمین شیوع یک پیامد مشخص است و به توان آماری بستگی ندارد. ولی برای محاسبه حجم نمونه در پژوهش های مقطعی تحلیلی، لازم است پژوهشگر تخمینی از شیوع، واریانس شیوع، تفاوت معنی دار بین افراد مواجهه یافته و مواجهه نیافته با عامل مورد نظر، سطح معنی داری، و توان دلخواه داشته باشد. امروزه محاسبات حجم نمونه با وجود نرم افزارها (برای نمونه، GPower) آسان شده و حتی بطور آنلاین امکان پذیر است (برای نمونه، <http://riskcalc.org:3838/samplesize/>).

## خلاصه ی ویژگی های پژوهش های مقطعی

- ارزان هستند و انجام آنها آسان و سریع است.
- مشکلات اخلاقی ندارند.
- جمع آوری داده ها برای همه متغیرها در یک زمان معین است.
- افراد در طی زمان پیگیری نمی شوند.
- برای فراهم کردن شواهد مقدماتی جهت طرح ریزی پژوهش در آینده مناسبند.
- برای تولید فرضیه مناسبند.
- اطلاعاتی در باره شیوع «پیامدها» و «علل» فراهم می کنند.

- اندازه گیری و بررسی «علل و پیامدها» به طور همزمان انجام می شود و از آنها رابطه علت و معلولی استنباط نمی شود.
- بسیاری از پژوهش های مقطعی به وسیله پرسشنامه ها و مصاحبه ها انجام می شوند.
- بهترین راه برای تعیین شیوع و بررسی ارتباطات «علل» و «پیامدها» هستند.
- در یک پژوهش مقطعی تحلیلی، تعیین تقدم و تأخر «علت» نسبت به «پیامد» مشکل است.
- برای انجام یک پژوهش عمیق، یافته های زیادی استفاده می شود.

## عدم مزایای پژوهش های مقطعی

- ناتوانی در ارزیابی بروز (Incidence)
- مناسب نبودن برای بررسی بیماری های نادر
- عدم استنباط رابطه علت و معلولی
- احتمال مشکل بودن تفسیر ارتباطات
- عدم بررسی رابطه زمانی بین «پیامدها» و «علل»
- نیاز به انتخاب نمونه افراد از یک جمعیت پژوهش بزرگ و ناهمگون
- احتمال سوگیری «عدم پاسخ»، «به خاطر آوردن»، و «نمونه گیری»

## پژوهش های کیس-کنترل

همانطور که پیش تر ذکر شد، اصطلاحات "آینده نگر" و "گذشته نگر" بر حسب زمان جمع آوری داده ها نسبت به ارائه مسئله پژوهش تعریف می شوند. همچنین گفته شد که گاهی اوقات این دو اصطلاح برای توضیح "جهتی" که پژوهش پیش می رود استفاده می شوند. بر اساس این تعریف جایگزین، جهت یک پژوهش آینده نگر، همانند طرح های کوهورت، از "علت" به "پیامد" است. جهت یک پژوهش گذشته نگر، همانند

طرح های کیس-کنترل، از "پیامد" به "علت" است. در اینجا، این اصطلاحات اشاره به زمان جمع آوری داده ها (نسبت به زمان ارائه فرضیه ها) دارد. هنگام ارجاع به جهت پیشبرد پژوهش، اصطلاحات "کوهورت" و "کیس-کنترل" استفاده می شود.

در پژوهش های کیس-کنترل یا مورد-شاهدی که گذشته نگر و مشاهده ای هستند، پژوهشگران افراد دارای بیماری مورد نظر (کیس) و افراد بدون بیماری مورد نظر (کنترل) را شناسایی می کنند. آنها سپس برای وجود یا عدم وجود ریسک فاکتورها که ممکن است وجود یا عدم وجود بیماری را توضیح دهند، به گذشته نگاه می کنند (شروع با «پیامدها» و جستجو برای «علل»). در حقیقت، در پژوهش های کیس-کنترل، پژوهشگران با اثر (پیامد، بیماری) شروع می کنند و در جستجوی علت هستند. به عبارت دیگر، وجود متغیر مستقل (علت، ریسک فاکتور) در دو گروه کیس و کنترل بررسی می شود. بنابراین، ویژگی پژوهش کیس-کنترل این است که پژوهشگران با "پیامد ها" شروع می کنند و در جستجوی "علل" هستند.

### مراحل پژوهش مورد-شاهدی (شکل ۳-۵)

- اول، افراد با «پیامد» مورد نظر (کیس) شناسایی می شوند.
- دوم، افراد همسان بدون «پیامد» مورد نظر (کنترل) شناسایی می شوند.
- سوم، پژوهشگران همه افراد (کیس ها و کنترل) را برای تعیین وجود یا عدم وجود فاکتورهای مختلف به عنوان «علت» فرضی پیامد مورد نظر (مرور پرونده ها، جمع آوری اطلاعات جدید، بررسی ویژگی های قبل از رخداد پیامد) ارزیابی می کنند.
- چهارم، آنالیز آماری انجام می شود (مقایسه فراوانی «علت» ها در دو گروه، و تعیین این که آیا افراد با علل فرضی، خطر نسبی بالاتری برای ابتلا به بیماری دارند).

یک مثال بسیار رایج، پژوهش هایی هستند که افراد دچار سرطان ریه (پیامد) را معاینه میکنند و تعیین می کنند چه تعداد از آنها سیگاری بوده اند و چه تعداد سیگاری نبوده اند (شکل ۳-۶). مثال رایج دیگر، معاینه افراد مبتلا به دیابت و تعیین افراد چاق در مقایسه با افرادی که چاق نیستند می باشد. فراوانی بالاتر سیگاری ها و افراد چاق به این معنی است که به ترتیب سیگار و چاقی «علل» «پیامد» ها هستند. (شکل ۳-۷).



شکل ۳-۵: طرح پژوهشی کیس-کنترل

در یک پژوهش با طرح کیس-کنترل، برای تعیین اثر بیماری های قبلی بر مرگ و میر بیماران تروما، پیامد مرگ ناشی از تروما، کیس های فوت شده، ۳۰۷۴ نفر از بین تقریباً ۲۰۰/۰۰۰ بیمار ترومایی بودند، و هر کیس فوت شده، با چهار کیس ترومایی که زنده مانده بودند (گروه کنترل، مرخص شده از همان بیمارستان، در گروه سنی مشابه، و با شدت ضایعه یکسان) همسان سازی گردید. در این پژوهش، ۱۱ بیماری (پیامد) در افراد جمع آوری شد. نتایج این پژوهش با تست نسبت شانس (odds ratio) نشان داد که در افراد مبتلا به سیروز، خطر فوت کردن ۴/۷ برابر افراد بدون سیروز و در افراد مبتلا به COPD خطر فوت کردن ۱/۸ برابر افراد بدون COPD بود.

در پژوهش دیگری با طرح کیس-کنترل برای تعیین اثر فیزیوتراپی تنفسی زودرس بر رخداد ضایعه مغزی در نوزادان بسیار نارس، ۱۳ نوزاد در طی ۳ سال با مرور پرونده پزشکی شناسایی شدند (کیس). هر کیس با دو کنترل بدون ضایعه مغزی همسان سازی شدند (کنترل). برای تعیین علل، ۵۱ «فاکتور» مرتبط با رخداد ضایعه مغزی شناسایی شدند. نتایج این پژوهش نشان داد که فقط وجود بریچ، هیپوتانسیون در «هفته اول» زندگی، تعداد فیزیوتراپی تنفسی در طی «ماه اول» زندگی، بین دو گروه تفاوت داشتند. بر اساس این نتایج، فیزیوتراپی تنفسی برای نوزادان بسیار کم وزن در طی ماه اول زندگی متوقف شد، و ضایعه مغزی بیشتر در گروه نوزادان بستری در بیمارستان دیده نشد.



شکل ۳-۶: طرح پژوهش کیس-کنترل برای بررسی علت سرطان ریه

در پژوهش دیگری با طرح کیس-کنترل برای بررسی خطرات موجود در منزل و خطر افتادن و شکستگی‌های ران در افراد سالمند، گروه کیس، ۴۳ بیمار با افتادن‌های مکرر و ۵۲ بیمار با شکستگی ران بودند. گروه کنترل برای دو گروه به ترتیب ۱۵۷ و ۱۹۵ نفر بودند. پیامدها، افتادن و شکستگی ران بودند. نتایج این پژوهش نشان داد که شانس

شکستگی های ران ناشی از افتادن در خانه هایی با نرده های توالت نامناسب، نسبت به خانه های دارای نرده های ایمن، ۱۱ برابر بیشتر بود.

### مزایای طرح کیس-کنترل

- قابلیت اجرا
- سریع
- امکان ارزیابی «علل»/«اکسپوژر» متعدد
- بررسی تعداد نمونه های بیشتر
- مناسب برای بیماری های نسبتاً نادر
- نسبتاً ارزان
- بررسی مواردی که شاید پژوهش در باره آن اخلاقی نباشد.

### عدم مزایای طرح کیس-کنترل

- ارزیابی فقط یک «پیامد»/«بیماری»
- روشن نبودن روش مستند کردن یا شناسایی «علل» فرضی
- عدم ثبت درست همه «فاکتورها» برای همه کیس ها و کنترل ها
- احتمال بیشتر سوگیری به دلیل گذشته نگر بودن
- مشکلات استخراج داده های صحیح از تاریخچه افراد
- مشکل در اثبات رابطه زمانی صحیح بین «علت» و «بیماری»

| «بیماری»                           |        |                                   |
|------------------------------------|--------|-----------------------------------|
| سرطان ریه                          | سالم   |                                   |
| (a) 20                             | (b) 5  | سیگاری                            |
| (c) 80                             | (d) 95 | غیرسیگاری                         |
|                                    |        | «علت»                             |
| شانس ابتلا برای کپس ها: $a/c=0.25$ |        | شانس ابتلا برای کنترل: $b/d=0.53$ |
| $OR=ad/bc; 1900/053=4.75$          |        |                                   |

شکل ۳-۷: نسبت شانس ابتلا به سرطان ریه.

در این مثال، شانس ابتلا به سرطان ریه در افراد سیگاری حدود ۵ برابر افراد غیرسیگاری است.

## پژوهش های کوهورت

پژوهش کوهورت نوعی پژوهش مشاهده ای برای بررسی روابط بین ریسک فاکتورها و پیامدهای سلامت است. در پژوهش های کوهورت که پژوهش کوهورت همزمان (Concurrent cohort) نیز نامیده می شود، پژوهشگران افراد دارای ریسک فاکتورهای مختلف را شناسایی می کنند و به آینده نگاه می کنند تا مشاهده کنند آیا بیماری مورد نظر بروز می کند.

در نوعی از پژوهش های کوهورت که پژوهش های کوهورت ناهمزمان یا کوهورت تاریخی نامیده می شود، افراد بر پایه داده های جمع آوری شده در گذشته، در گروه های دارای ریسک فاکتور قرار داده می شوند. سپس در حال حاضر یا آینده اندازه گیری ها انجام می شود تا تعیین کنند در چه کسی پیامد مورد نظر بروز کرده است.

ویژگی پژوهش های کوهورت، پیشرفت از "علت" به "اثر" است (شکل ۳- ۸). در پژوهش کوهورت، افرادی انتخاب می شوند که بیماری یا ضایعه مورد نظر را ندارند، و می توانند از لحاظ یک ریسک فاکتور یا فاکتورهای مورد نظر طبقه بندی شوند. افراد "سالم" (حداقل با توجه به بیماری یا ضایعه مورد نظر) براساس ریسک فاکتورهای مختلف، برای یک دوره زمانی پیگیری می شوند تا خطرات نسبی آنها برای رخداد بیماری یا ضایعه مقایسه شود.



شکل ۳- ۸: طرح پژوهش کوهورت

طرح کوهورت شکل های مختلفی دارد. در شکل اول، افراد در زمان حال شناسایی شده و پیگیری می شوند تا ببینند چه کسی پیامد مورد نظر را نشان می دهد (حال-آینده، شکل ۳- ۸). این نوع از طرح کوهورت به عنوان طرح کوهورت یا طرح کوهورت همزمان (Concurrent) یا آینده نگر (Prospective cohort design) یا طولی (Longitudinal) شناخته می شود. پژوهش های کوهورت «آینده نگر» پیش تر طراحی شده و در طی یک دوره زمانی در آینده انجام می شوند. برای نمونه، پژوهش برای تعیین فاکتورهای موثر در بیماری های قلبی عروقی.

در یک طرح کوهورت همزمان، همبستگی سن در زمان ضایعه با پیامدهای روانی اجتماعی ضایعه تروماتیک مغز بررسی شد. چهارصد فرد مبتلا به ضایعه تروماتیک مغز، ۱ ماه و ۱ سال پس از ضایعه، برای تعیین عملکرد روانی و اجتماعی ارزیابی شدند. متغیر مستقل، سن بود (در ۵ گروه سنی، ۱۸ سال تا ۶۰ سال و مسن تر. متغیر وابسته، عملکرد روانی اجتماعی، موقعیت زندگی پس از ضایعه، و وضعیت شغلی پس از ضایعه بود. نتیجه این پژوهش نشان داد که در کل، افراد مسن مبتلا به ضایعه تروماتیک مغز، در مقایسه با افراد جوان تر اختلال عملکرد روانی اجتماعی بیشتری داشتند.

هدف پژوهش کوهورت آینده نگر فرامینگهام، تعیین فاکتورهای موثر در بیماری قلبی عروقی بود و ارزیابی جمعیت هر دو سال یک بار، برای پیش بینی ۱۰ ساله ی خطر بیماری قلبی عروقی در افراد بدون بیماری قلبی عروقی انجام می شد. جمعیت ۲۸۰۰۰ نفر بودند که از میان آنها ۶۵۰۰ نفر بطور تصادفی، در سن ۵۹-۳۰ سال انتخاب شدند. شرط ورود، فقط نداشتن شواهد کلینیکی بیماری عروقی و تصلب شرایین بود. در این پژوهش از پرسشنامه استفاده شد و «علت ها» شامل سیگار، مصرف الکل، چاقی، فشار خون بالا، کلسترول بالا، فعالیت فیزیکی کم بود.

در یک طرح کوهورت آینده نگر، پیامدها پس از جراحی کمر بررسی شد. اطلاعات در شروع طرح و بعد از ۱۲ ماه جمع آوری شد. با استفاده از اطلاعات حاصل از پیگیری ۱۲ ماهه، افراد گروه به کسانی که با جراحی بهبود یافتند و کسانی که بهبود نیافتند تقسیم شدند. اطلاعات به دست آمده در شروع پژوهش، برای شناسایی فاکتورهای مختلف بین دو گروه بررسی شد. نتایج این پژوهش، ارتباط سن بالا و جراحی قلبی کمر با پیامدهای بدتر را نشان داد.

پژوهشگران با هدف تعیین شیوع تاخیر در صحبت کردن کودکان و بررسی وضعیت پیش بینی کنندگی وضعیت مادر، خانواده، و متغیرهای مربوط به کودکان، از یک طرح کوهورت آینده نگر پیروی کردند. نمونه ای تصادفی شامل ۱۷۶۶ کودک انتخاب شد و

برای مدت ۸ سال پیگیری شدند. مقاله در سال ۲۰۰۷ منتشر شد. نرخ پاسخ در داده های بدست آمده در سال اول و دوم، ۸۵٪ بود. از این تعداد کودک، ۱۳/۴٪ آنها در صحبت کردن تاخیر داشتند. سابقه فامیلی، جنس پسر، و تاخیر در رشد نورویولوژیکی پیشینی کننده تاخیر در صحبت کردن کودکان بودند.

## مراحل پژوهش کوهورت

- شروع با افراد «سالم»، بدون بیماری مورد نظر
- ارزیابی «علت»، و طبقه بندی افراد بر پایه «ریسک فاکتور»
- پیگیری در یک بازه زمانی برای مشاهده بروز «بیماری»
- محاسبه بروز بیماری در گروه ها، مقایسه خطرات نسبی بروز بیماری یا ضایعه در آنها

در شکل دوم، افراد بر پایه داده های جمع آوری شده در گذشته، از لحاظ ریسک فاکتور گروه بندی می شوند. سپس در حال یا آینده ارزیابی می شوند تا تعیین گردد که چه کسی پیامد مورد نظر را نشان می دهد. این نوع از طرح کوهورت، طرح کوهورت تاریخی (Historical) یا ناهمزمان (Nonconcurrent) نامیده می شود.

در طرح کوهورت ناهمزمان، برای پاسخ به این سوال که آیا بیماران مبتلا به سکته مغزی در مقایسه با افراد بدون سکته مغزی در معرض خطر بالاتر ابتلا به دمانس هستند، بیماران به دو گروه مبتلا به سکته مغزی و بدون سکته مغزی تقسیم شدند. بنابراین لازم بود سکته مغزی قبلا رخ داده باشد. «جهت» این پژوهش، از «علت»، یعنی سکته مغزی در گذشته» به «پیامد، یعنی دمانس»، در حال حاضر بود. نتیجه، احتمال حدود ۶ برابری خطر ابتلا به دمانس در افراد مبتلا به سکته مغزی نسبت به افراد بدون سکته مغزی را نشان داد.

در شکل سوم، یا طرح کوهورت گذشته نگر (Retrospective)، همه داده ها در زمان طرح سوال وجود دارند، پژوهشگران داده های مناسب را استخراج می کنند و افراد را بر اساس متغیر یا متغیرهای مورد نظر گروه بندی می کنند. سپس برای تعیین تفاوت ها، در

مجموعه ای از متغیرهای پیش بینی کننده جستجو می کنند. در پژوهش های کوهورت گذشته نگر، داده ها در گذشته جمع آوری شده اند، اما آنالیز در حال حاضر انجام می‌گردد. برای نمونه، شناسایی گذشته نگر «علت» و اطلاعات بیماری.

در یک طرح کوهورت گذشته نگر، پذیرش مجدد کودکان بستری در بخش توانبخشی ریوی بررسی شد. پژوهشگران پرونده های پزشکی کودکان را به طور گذشته نگر مرور کردند. کودکانی که در طی بستری به اکسیژن یا ونتیلاتور نیاز داشتند شناسایی شده و بر طبق پذیرش مجدد، گروه بندی شدند. بر اساس نتیجه ی این پژوهش، کودکانی که مجددا پذیرش شدند، در مدت بستری اولیه، به ونتیلاتور و مراقبت پرستاری بیشتری نیاز داشتند.

در همه اشکال طرح کوهورت، پژوهشگر ابتدا وجود یا عدم وجود ریسک فاکتور، و سپس وجود یا عدم وجود پیامد مورد نظر را تعیین می کند. طرح کوهورت، معکوس، طرح کیس-کنترل است که در آن پیامد، قبل از شناسایی ریسک فاکتورها وجود دارد (شکل ۳-۹).



شکل ۳-۹: طرح پژوهش کوهورت در مقایسه با طرح کیس-کنترل

خلاصه انواع طرح کوهورت (شکل ۳-۱۰)

۱- طرح کوهورت آینده نگر یا همزمان

افراد در حال حاضر شناسایی شده و برای مشاهده «بیماری/پیامد» پیگیری میشوند.

۲- طرح کوهورت تاریخی یا کوهورت ناهمزمان

گروه بندی افراد از لحاظ «ریسک فاکتور/علت» بر پایه داده های جمع آوری شده در گذشته، و پیگیری برای «بیماری/پیامد» تا حال حاضر، سپس ارزیابی در حال حاضر یا آینده برای تعیین بروز «بیماری» در افراد.

۳- طرح کوهورت گذشته نگر

همه داده ها قبلاً جمع آوری شده اند. استخراج داده های مناسب، گروه بندی افراد برپایه متغیر یا متغیرهای مورد نظر، و جستجو برای تفاوت ها در مجموعه ای از متغیرهای پیشبینی کننده.



شکل ۳-۱۰: طرح پژوهش کوهورت

## خلاصه

"اپیدمیولوژی" بررسی فراوانی بیماری یا ضایعه، و عوامل بیماری زا در جمعیت ها است. عملکرد کلیدی یک پژوهش اپیدمیولوژیک این است که پژوهشگر را قادر می سازد تا سوال پژوهشی را با حداقل ابهام پاسخ دهد. در اپیدمیولوژی، پژوهشگران به اندازه گیری یا ارزیابی رابطه علت با یک بیماری (پیامد) علاقه مند هستند. آنها به عنوان قدم اول، فرضیه ای را بر پایه سوال پژوهشی تعریف می کنند، سپس تصمیم می گیرند که کدام طرح پژوهش برای پاسخ دادن به سوال مناسب است. طرح پژوهشی انتخابی توسط پژوهشگر، چگونگی انجام و هدایت پژوهش را تعیین می کند. "شیوع" نسبت موردهای یک بیماری یا ضایعه در یک جمعیت در یک نقطه مشخص از زمان است. "بروز" نرخ کیس های جدید یک بیماری یا ضایعه در یک جمعیت در خطر است. طرح های پژوهشی اپیدمیولوژیک رایج عبارتند از: پژوهش های مقطعی (اندازه گیری افراد در یک نقطه از زمان)، پژوهش های کیس-کنترل (کار در جهت رو به عقب از اثر به علت)، و پژوهش های کوهورت (کار در جهت رو به جلو از علت به اثر).



## فصل چهارم

تئوری اندازه گیری

بخش های فصل

تعریف اندازه گیری

ویژگی های متغیر

مقیاس های اندازه گیری

انواع متغیرها

اندازه گیری پایایی

اندازه گیری اعتبار

پاسخ دهی به تغییرات

خلاصه

## تئوری اندازه گیری

وقتی شما بتوانید آنچه را که در باره اش صحبت می کنید اندازه گیری کنید و آن را به صورت اعداد بیان کنید، شما چیزی در باره آن می دانید.

لورد کلونین

متخصصین حرفه های مختلف سلامت، برای کمک به تصمیم گیری در باره مشکلات بیماران، چگونگی مداخله کردن، و زمان قطع درمان ها، و پژوهشگران برای سنجش ویژگی هایی که بررسی می کنند، از اندازه گیری ها استفاده می کنند. در حقیقت، بعضی

پژوهشگران بیشتر پژوهش هایشان بر ارزیابی معیارها متمرکز است. پژوهشگران همچنین از اندازه گیری ها برای سنجش دوز درمان داده شده به بیماران استفاده می کنند. گرچه، آگاهی در باره مفید بودن اندازه گیری ها برای پژوهشگران ضروری است، ولی کلینیسین ها نیز لازم است مفاهیم مربوط به ارزیابی و کارایی معیارهایی که استفاده می کنند را بدانند.

در این فصل، چهارچوبی برای درک و ارزیابی اندازه گیری های مورد استفاده توسط متخصصین سلامت ارائه می شود و تعاریف اندازه گیری، ویژگی متغیرها، مقیاس های اندازه گیری، انواع متغیرها، و پایایی، اعتبار، و پاسخ دهی به تغییرات، مورد بحث قرار می گیرد.

## تعاریف اندازه گیری

به طور کلی، اندازه گیری وسیله ای برای تمایز است. در تعریف دقیق تر، اندازه گیری عبارت است از قوانینی برای اختصاص دادن اعداد به اشیاء، به طریقی که نمایانگر مقدار آن ویژگی ها باشند. بر اساس تعریف اول، طبقه بندی بیماران به گروه های مختلف بیماری، شکلی از اندازه گیری است، اما بر اساس تعریف دوم اینگونه نیست. در یک تعریف کلی می توان اندازه گیری را به عنوان یک پروسه سیستماتیک برای تمایز اشیاء در نظر گرفت. بنابراین، اندازه گیری یک فرآیند تصادفی نیست، بلکه فرآیندی بر طبق قوانین و راهنما ها است.

تمایز می تواند با "نام" (Name)، "شماره" (Numeral)، یا "عدد" (Number) انجام شود. برای مثال، طبقه بندی افراد به «کم وزن»، «وزن طبیعی»، و «بیش وزنی»، اختصاص دادن "نام" برای تمایز افراد بر اساس ویژگی ترکیب ایده آل بدن می باشد. اگر به این گروه ها عنوان ۱، ۲، ۳ داده شود، در این صورت به هر فرد "شماره ای" اختصاص میابد

که نمایانگر ترکیب بدن او می باشد. یک "شماره" سمبلی است که معنی مقداری ندارد، به عبارتی شکلی از نامیدن است. طبقه بندی بر اساس درصد چربی بدن (مثلاً ۱۰٪ یا ۱۴٪) اختصاص دادن یک "عدد" به این ویژگی است که مقدار چربی بدن را نشان می دهد. یک "عدد" یک "شماره" است با اختصاص مقدار به آن.

## ویژگی های متغیر

متغیر مشخصه یک فرد، شیء، و یا پدیده است که می تواند مقادیر مختلفی داشته باشد (مثل سن، وزن). وقتی ما تغییرات در یک متغیر را اندازه گیری می کنیم، در حقیقت بعضی ویژگی های آن (Attribute) را اندازه گیری می کنیم. این جنبه از متغیر است که در تعریف عملی متغیر وابسته مشخص می شود. بطور کلی، انواع ویژگی هایی که ممکن است اندازه گیری شود حداقل یکی از موارد زیر است:

فاصله (Distance)، (برای مثال، مسافتی که فرد می برد یا مسافتی که راه می رود)

مدت (Duration)، (برای مثال، مدت زمان راه رفتن)

فرکانس (Frequency)، یا فراوانی رخداد یک رفتار، "ریت" (Rate) یک ویژگی است و عبارت است از فرکانس رخداد در فواصل زمانی مختلف.

مقدار (Magnitude)، بزرگی یک پاسخ است (برای مثال، اندازه گیری مقدار صدا به دسیبل، یا مقدار فشار اندازه گیری شده به پوند، اونس، یا کیلوگرم).

توپوگرافی (Topography)، عبارت است از حالت یا راستای قسمت های بدن در یک وضعیت یا عمل حرکتی، که همچنین می تواند الگوی حرکات باشد.

زمان تاخیر (Delay)، مدت زمان تاخیر در بروز یک رفتار بعد از یک سیگنال.

الگو (Pattern)، عبارت است از سکانس رفتار مثل الگوی راه رفتن.

ویژگیهایی که قرار است اندازه گیری شوند باید انعکاسی از سازه های تئوریک یا مفاهیم متغیر مستقل باشند. برای مثال، اگر هدف پژوهش، بررسی اثر مداخله بر آسانی راه رفتن در سگته مغزی باشد، اندازه گیری مسافت طی شده در راه رفتن نامناسب است زیرا مسافت طی شده در راه رفتن اغلب معیاری برای تحمل است. در این مورد خاص، اندازه گیری زمان تحمل وزن روی هر پا در طی راه رفتن، معیار مناسب تری برای "آسانی" راه رفتن است.

## مقیاس های اندازه گیری

مقیاس ها یا سطوح کلاسیک اندازه گیری چهار تا هستند: اسمی، رتبه ای، فاصله ای، و نسبتی (شکل ۴-۱). این مقیاس ها بر مبنای داشتن یا نداشتن ویژگی های یک سیستم اعداد واقعی مشخص می شوند. یک سیستم عدد واقعی با رتبه (Order)، فاصله (Distance)، و صفر (Origin) تعریف می شود. رتبه به معنی اعداد بالاتر است که نمایانگر مقادیر بیشتر ویژگی است که اندازه گیری می شود. فاصله به معنی مقدار تفاوت بین اعداد متوالی است که یا مساوی یا لگاریتمی است (یعنی فاصله بین اعداد متوالی با مضربی برابر است). صفر به معنی عدد ۰ است و نمایانگر فقدان کیفیت اندازه گیری شده است.

## مقیاس اسمی

مقیاس های اسمی (Nominal/Categorical scale) هیچیک از ویژگی های سیستم عدد واقعی را ندارند. مقیاس اسمی بدون هرگونه ارزش گذاری روی طبقه ها، فقط طبقه بندی می کند. در مواردی که برای یک متغیر، رتبه، فاصله، یا عدد صفر برای طبقه بندی وجود ندارد، طبقه بندی می تواند با اسم (Name) یا شماره (Numeral) شناسایی شود. البته اغلب بهتر است به جای شماره، طبقه ها را اسم گذاری کنیم، چون به این ترتیب هیچگونه تفاوت کمی بین طبقه ها استنباط نمی شود.

### نمونه هایی از مقیاس اسمی

- طبقه بندی بیماران مبتلا به فلج مغزی:

- همی پلژی (فلج یک نیمه بدن)
- دیپلژی (فلج پاها)
- کوادری پلژی (فلج چهار اندام)

- طبقه بندی رضایت بیماران:

- رضایت دارم
- رضایت ندارم

سوال از افراد در باره محل زندگی:

- شهر
- حومه شهر
- روستا

جنسیت افراد:

- مرد
- زن

در مقیاس های اسمی، خود طبقه بندی بدون ارزش است، یعنی ارزش کمی ندارد.

### مقیاس رتبه ای

مقیاس های رتبه ای فقط یکی از سه ویژگی یک سیستم عدد واقعی را دارا هستند: یعنی رتبه (Order/Rank). بنابراین، یک مقیاس رتبه ای می تواند برای تعیین کم و بیش یک کیفیت معین از یک فرد یا یک شیء استفاده شود. مقیاس های رتبه ای این اطمینان را فراهم نمی کنند که بین طبقه ها یا رتبه ها فواصل مساوی وجود دارد. از آنجایی که در یک مقیاس رتبه ای فواصل یا نابرابر هستند یا معلوم نیست، محاسبات ریاضی مثل جمع، تفریق،

ضرب، یا تقسیم اعداد رتبه ای بی معنی است. بسیاری از مقیاس های عملکردی رتبه ای هستند.

## نمونه هایی از مقیاس های رتبه ای

- طبقه بندی رضایت بیمار:

- رضایت کامل
- رضایت متوسط
- رضایت کم

- مقدار کمکی که یک بیمار برای راه رفتن نیاز دارد:

- حداکثر
- متوسط
- حداقل
- مستقل

- میزان رضایت مشتریان از خدمات یک شرکت:

- خیلی ناراضی
- ناراضی
- خنثی
- راضی
- خیلی راضی

بعضی اوقات از اعداد (Numeral) برای نقاط یک مقیاس رتبه ای استفاده میشود، اما اعتبار این رویکرد مورد سوال واقع شده است چون اعداد اغلب بگونه ای در نظر گرفته می شوند که گویا مفهوم کمی دارند.

نوع دوم مقیاس رتبه ای، رتبه بندی یا رنکینگ (Ranking) است، مثل رنکینگ دانشگاه ها، تنیس بازان، و تیم های فوتبال باشگاهی و ملی. گرچه فاصله بین دو تنیس باز اول و دوم ممکن است ناچیز باشد، اما تفاوت بین تنیس باز دوم و سوم ممکن است بسیار بیشتر باشد. فواصل بین رنک ها معمولاً معلوم نیست و نمی توان فرض کرد که مساوی هستند. مقیاس رتبه ای یک قدم بالای مقیاس اسمی قرار دارد، و اگر پژوهشگر بخواهد اطلاعات بیشتری به دست بیاورد می تواند از مقیاس های رتبه ای برای اندازه گیری پیامد استفاده کنند. برای میزان رضایت، شادی، و درجات، می توان از مقیاس های رتبه ای استفاده کرد. در شکل ۴-۲، مقیاس رتبه ای تندی برای فلفل نشان داده شده است.

## مقیاس های فاصله ای

مقیاس های فاصله ای ویژگی های "رتبه" و "فاصله" سیستم عدد واقعی را دارد، اما فاقد صفر معنی دار است. یک نقطه صفر معنی دار، نمایانگر فقدان مقدار اندازه گیری شده است. بنابراین، مقیاس فاصله ای همه ویژگی های مقیاس رتبه ای را دارد، و علاوه بر آن محاسبه تفاوت بین متغیرها را ممکن می کند. ویژگی اصلی این مقیاس تفاوت یکسان در فاصله بین اشیاء است. در داده های فاصله ای، فاصله بین اعداد برابر است اما این ها به صفر واقعی مربوط نیستند، بنابراین نمایانگر مقدار مطلق نیستند.

مقیاس درجه حرارت سلسیوس و فارنهایت نمونه های کلاسیک مقیاس های فاصله ای هستند. نقاط صفر در دو مقیاس درجه حرارت قراردادی هستند. در مقیاس فارنهایت، نقطه صفر درجه حرارتی است که در آن آب نمک یخ می زند، و در مقیاس سلسیوس درجه حرارتی است که در آن آب تازه یخ می زند. هیچکدام فقدان ویژگی پایه حرارت را ندارد، و در هر دو مقیاس درجه حرارت می تواند پایین تر از صفر درجه برود. البته هر دو مقیاس فواصل منظم (اما متفاوت) دارند. به علت فواصل مساوی، جمع و تفریق با مقیاس

های فاصله ای معنی دار است. ده درجه افزایش درجه حرارت به این معنی است که درجه حرارت چه از صفر درجه به  $10^{\circ}$  یا از  $100^{\circ}$  به  $110^{\circ}$  به این میزان افزایش یافته است. حرارت  $80^{\circ}$  درجه همیشه بالاتر از  $50^{\circ}$  درجه است و تفاوت بین این دو درجه حرارت با تفاوت بین  $70^{\circ}$  درجه و  $40^{\circ}$  درجه یکسان است.

در مواردی که تفاوت بین متغیرها مورد توجه می باشد، از مقیاس های فاصله ای استفاده می شود. مقیاس فاصله ای تفاوت بین دو متغیر را می سنجد در صورتی که دو مقیاس اسمی و رتبه ای فقط قادر به ارتباط دادن مقادیر کیفی با متغیرها هستند. مقیاس های فاصله ای موثر هستند چون راهی برای انجام آنالیز آماری داده های فراهم شده باز می کنند. گرچه، ضرب و تقسیم داده های درجه حرارت فارنهایت یا سلسیوس مفید نیستند، چون در این اعمال فرض بر وجود مقادیر "صفر واقعی" است. به عبارت دیگر در مقیاس های فاصله ای درجه حرارت  $100^{\circ}$  بر اساس فارنهایت دو برابر گرمای  $50^{\circ}$  نیست و فقط  $50^{\circ}$  گرم تر است.

همه تکنیک های قابل استفاده برای آنالیز داده های اسمی و رتبه ای، برای مقیاس های فاصله ای نیز کاربرد دارند. جدا از آن تکنیک ها، چند روش آنالیز نظیر آمار توصیفی و آنالیز رگرسیون همبستگی، به طور گسترده ای برای آنالیز داده های فاصله ای استفاده می شوند. آمار توصیفی اصطلاحی است که برای آنالیز داده های عددی به منظور توصیف، نمایش، یا خلاصه کردن داده ها بصورت معنی دار استفاده می شود. در این مقیاسها، از میانگین، میانه، و نما برای محاسبه شاخص مرکزی استفاده می شود (جدول ۴-۱).

سوالات زیر جزء طبقه مقیاس های فاصله ای قرا می گیرند:

- درآمد شما چقدر است؟

- درجه حرارت در شهر شما چقدر است؟

جدول ۴-۱: تست های آماری مورد استفاده در مقیاس های اسمی، رتبه ای، فاصله ای، و

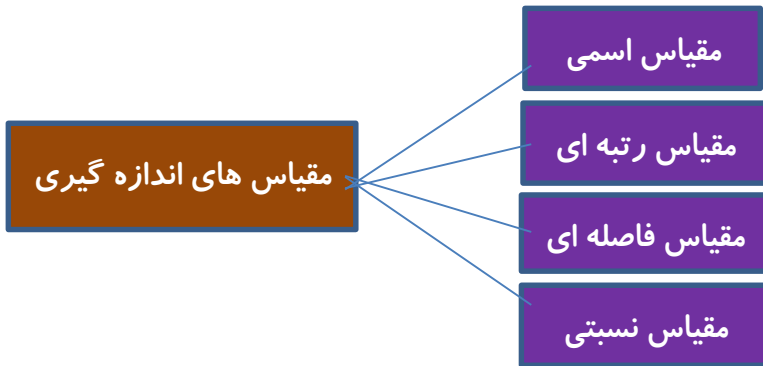
نسبتی

| نسبتی | فاصله ای | رتبه ای | اسمی |                           |
|-------|----------|---------|------|---------------------------|
| بله   | بله      | بله     | -    | ترتیب متغیرها             |
| بله   | بله      | بله     | بله  | نما                       |
| بله   | بله      | بله     | -    | میانه                     |
| بله   | بله      | -       | -    | میانگین                   |
| بله   | بله      | -       | -    | ارزیابی تفاوت بین متغیرها |
| بله   | بله      | -       | -    | جمع و تفریق متغیرها       |
| بله   | -        | -       | -    | ضرب و تقسیم متغیرها       |
| بله   | -        | -       | -    | صفر مطلق                  |

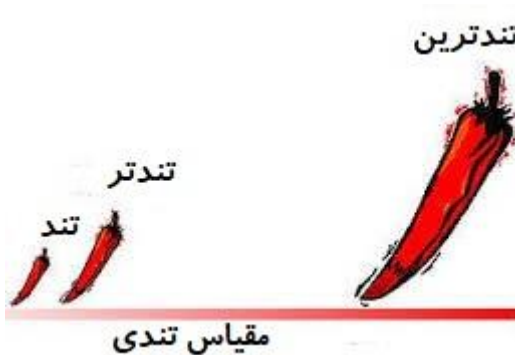
## مقیاس های نسبتی

مقیاس های نسبتی، هر سه جزء سیستم عددی واقعی، یعنی رتبه، فاصله، و صفر را دارند. اعداد در داده های نسبتی، نمایانگر واحدهایی با فاصله های برابر هستند که از صفر واقعی اندازه گیری شده اند. سن، طول، زمان، وزن، قدرت، فشار خون، دامنه حرکتی، و قد عموماً مقیاس های نسبتی هستند چون فقدان آنها نمره صفر می گیرد، و فاصله بین اعداد مساوی هستند. بنابراین، اعداد نمایانگر مقدار واقعی متغیر اندازه گیری شده هستند، و بنابراین می توان گفت قد، وزن، و ... یک فرد دو برابر دیگری است. همه اعمال ریاضی جمع، تفریق، ضرب، و تقسیم می توانند برای مقیاس های نسبتی استفاده شوند. مقیاس درجه حرارت کلوین نمونه ای از مقیاس نسبتی است چون فواصل بین درجات مساوی

هستند و نقطه صفر نمایانگر فقدان گرما است. داده های مقیاس های نسبتی، ماهیت کمی دارند و بنابراین همه تکنیک های آنالیز کمی می تواند برای محاسبه داده های نسبتی استفاده شوند.



شکل ۴-۱: مقیاس اسمی ضعیف ترین و مقیاس نسبتی قویترین مقیاس های اندازه گیری هستند.



شکل ۴-۲: مقیاس رتبه ای برای "تندی"

## تعیین مقیاس اندازه گیری

برای تعیین مقیاس یک معیار، اول پژوهشگر باید مطمئن شود که آیا نقطه صفر واقعی وجود دارد، آیا فواصل بین اعداد مساوی هستند، و آیا ترتیبی برای اعداد یا اسمی وجود دارد که تشکیل دهنده معیار رتبه ای باشند. گرچه این امر ساده به نظر می رسد، اما مواردی هست که ممکن است هنگام تعیین مقیاس یک اندازه گیری نقش داشته باشند. برای مثال، آیا طبقه بندی بیماران به صورت "کم وزن"، "طبیعی"، و "بیش وزن" یک مقیاس اسمی است یا رتبه ای؟ از آنجایی که به هریک از این طبقات می توان عددی را اختصاص داد و این طبقات دارای رتبه هستند، بنابراین یک مقیاس رتبه ای است. وقتی که مقیاس های رتبه ای مقادیر بسیاری داشته باشند (برای مثال، نمره کل از صفر تا ۱۰۰ در مقایسه با مقیاس کمک به رفتن بیماران با فقط پنج مقدار)، داده ها ممکن است ویژگی های ریاضی یک مقیاس فاصله ای را داشته باشند. از آنجایی که مقیاس اندازه گیری تعیین می کند که چه عملیات ریاضی را می توان انجام داد، بنابراین تعیین دقیق نوع مقیاس برای تعیین تست های آماری مناسب اهمیت دارد.

## انواع متغیرها

با توجه به اندازه گیری، متغیرها می توانند پیوسته (Continuous) یا گسسته (Discrete) باشند. متغیر گسسته متغیری است که مقادیر آن از هم مجزا هستند. سطوح متغیرها در یک مقیاس اسمی، طبق تعریف، گسسته هستند. طبقه بندی افراد مبتلا به فلج مغزی بر اساس اندام های درگیر (مثل فلج چهار اندام) نمونه ای از یک متغیر گسسته است چون آنها را نمی توان بین طبقات قرار داد و هر نوع به یکی از طبقات تعلق دارند. متغیرهای گسسته ای که فقط شامل دو مقدار هستند متغیرهای دوتایی (Dichotomous) نامیده می شوند، مثل جنسیت (مرد، زن) یا بیماری (وجود دارد، وجود ندارد). متغیرهایی که شامل شمارش رفتارها یا افراد هستند، متغیرهای گسسته هستند چون کسری از افراد یا رفتارها ممکن

نیست. اگر معیار مورد نظر، تعداد دفعاتی است که یک کودک در ۱۰ بار آزمایش نام شیئی را صحیح می گوید، ممکن نیست که در یک آزمایش منفرد نمره ۷/۵ به دست آورد. اما اگر سکانس ۱۰ آزمایش در ۴ روز متوالی تکرار شود با نمرات برای مثال ۸، ۶، ۹، و ۷ (جمع نمرات ۳۰) می توان میانگین نمرات را محاسبه کرد ( $30 \div 4 = 7.5$ ). بنابراین، اگر متغیرهای گسسته شامل دامنه نسبتاً وسیعی از مقادیر باشند، ویژگی های یک سیستم عددی واقعی را داشته باشند، یا میانگین نمرات آزمایشات قابل محاسبه باشد، به این ترتیب مانند متغیرهای پیوسته خواهند بود.

متغیر پیوسته متغیری است که بین مقادیر پیوستگی وجود دارد و از لحاظ تئوریک میتواند با درجه ای کوچک تر و کوچک تر اندازه گیری شود. به عبارت دیگر، معیاری است که می تواند بین دو مقدار خود تمامی اعداد حقیقی ممکن را اختیار کند. وزن و درآمد نمونه متغیر پیوسته هستند.

## چهارچوب های اندازه گیری

دو چهارچوب پایه برای اندازه گیری و ارزیابی وجود دارد:

۱- مبتنی بر مقادیر نرمال (Norm referenced)

۲- مبتنی بر ضابطه (Criterion referenced)

معیارهای مبتنی بر مقادیر نرمال آنهایی هستند که قضاوت در باره عملکرد فرد را نسبت به مقادیر نرمال گروه انجام می دهند. مفاهیم آماری میانگین و انحراف معیار، مرتبط با این معیارها هستند. در اندازه گیری های مبتنی بر مقادیر نرمال، داده های جمع آوری شده با جدول نمرات خام نرمال مقایسه می شوند. معمولاً نمرات خام به طریقی استاندارد میشوند که آنها را نسبت به میانگین و انحراف معیار نمونه ای که در آن مقادیر نرمال بدست آمده اند مقایسه می کنند.

بسیاری از اندازه گیری های کلینیکی مبتنی بر مقادیر نرمال هستند. فشار خون و نبض نسبت به دامنه ای از مقادیر نرمال ارزیابی می شوند، عملکرد عضله می تواند با میانگین عملکرد برای گروه های همسان شده از لحاظ سن و جنس مقایسه شود، و وضعیت رشد ممکن است نسبت به نرم های تثبیت شده برای کودکانی که رشد طبیعی دارند مقایسه شود.

معیار مبتنی بر ضابطه (Criterion-referenced) معیاری است که در آن قضاوت در باره عملکرد فرد نسبت به ضابطه تعیین شده انجام می شود. وقتی که معلم ۷۵٪ را به عنوان حداقل نمره برای گذراندن یک درس تعیین می کند، این یک اندازه گیری مبتنی بر ضابطه است. اگر همه دانشجویان بیش از ضابطه ۷۵٪ نمره کسب کنند، همه آنها در آن درس قبول می شوند. اگر نمره فقط ۲۵٪ دانشجویان بیشتر از ضابطه تعیین شده باشد، بنابراین فقط ۲۵٪ آنها در آن درس قبول می شوند. در کلینیک، بیماران باید ضابطه تعیین شده را بدست بیاورند تا از بیمارستان مرخص شوند، یا از وسایل کمکی در منزل استفاده کنند، یا به ورزشکار مصدوم، اجازه شرکت در مسابقات داده شود.

### پایایی اندازه گیری

کلینیسین ها و پژوهشگران، اندازه گیری های مختلف با پایایی متفاوت را در کلینیک و آزمایشگاه انجام می دهند. صرف نظر از تخصص و حیطه فعالیت، اندازه گیری ها به دلایل مختلفی انجام می شود (برای مثال، به عنوان بخشی از ارزیابی بیمار، کمک به تصمیم گیری درباره طرح درمان، برای فیدبک دادن به بیماران و دیگران، و کمک به پیش بینی و قضاوت در باره پیامد درمان).

پایایی اندازه گیری ها و استفاده از آنها به طور پایا توسط کلینیسین ها و پژوهشگران، موضوعی مهم است که کمک می کند تا ارزش این اندازه گیری ها تعیین شود. آگاهی

از پایایی و چگونگی اندازه‌گیری آن کمک می‌کند که یافته‌های کلینیکی و نتایج تحقیقات منتشر شده قابل درک باشند. گرچه پایایی امری است که عموماً مورد توجه می‌باشد، اما سطح پایایی مورد نیاز برای این که از لحاظ کلینیکی قابل قبول باشد تعریف نشده است. مانند تست فرضیه، سطوح پایایی که از لحاظ آماری معنی دار است به این معنی نیست که از لحاظ کلینیکی نیز قابل قبول باشد. بنابراین سطوح پایایی باید با احتیاط تفسیر گردند. پایایی یک تست، عموماً اختصاصی جمعیت است و سطح آن بسته به جمعیت ممکن است متفاوت باشد. توافق رایج این است که هیچ تخمین منفردی از پایایی برای نشان دادن تصویری کامل از آن کافی نیست، و انواع مختلف پایایی باید با هم استفاده شوند.

"پایایی" یعنی نمرات یک تست، تا چه حد عاری از خطاهای اندازه‌گیری هستند. به عبارت دیگر، پایایی پیوستگی یا تکرارپذیری اندازه‌گیری‌ها را نشان می‌دهد. اصطلاحات معادل پایایی عبارتند از دقت (Accuracy)، ثبات (Stability)، و پیوستگی (Consistency). استفاده از رویکردهای مختلف برای ارزیابی پایایی (Reliability) و اعتبار (Validity) پژوهش، یک راه تولید یافته‌های مفید و معتبر است. در تعیین پایایی و اعتبار پژوهش، کاهش خطا هدف اصلی است. در ادامه‌ی این قسمت، تئوری‌های پایایی، اجزا، و معیارها معرفی می‌شوند.

## تئوری‌های پایایی

دو تئوری پایه در اندازه‌گیری عبارتند از تئوری کلاسیک اندازه‌گیری (Classical measurement theory) و تئوری تعمیم‌پذیری (Generalizability) که دیدگاه‌های متفاوتی برای پایایی فراهم می‌کنند. تئوری کلاسیک اندازه‌گیری بر این فرض استوار است که هر اندازه‌گیری یا نمره بدست آمده، شامل یک جزء واقعی (True) و یک جزء

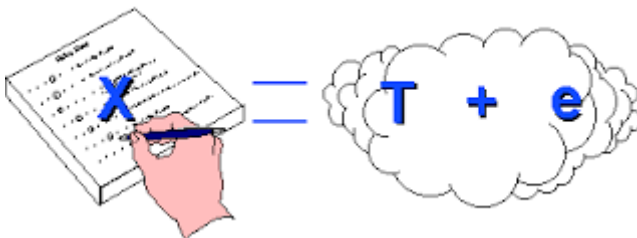
خطا (Error) است (شکل ۴-۳). به علاوه، هر فرد یک نمره واقعی منفرد در اندازه گیری مورد نظر دارد. نسبت نمره واقعی (T) به مجموع نمره واقعی و خطا (T + E)، ضریب پایایی (Reliability coefficient) نام دارد  $R = T/T + E$ .

وقتی که نمره مشاهده شده به نمره واقعی نزدیک می شود، پایایی افزایش می یابد، به طوری که با خطای صفر، پایایی کامل وجود دارد و ضریب مساوی ۱ است، زیرا نمره مشاهده شده با نمره واقعی یکی است.

بالعکس، با افزایش خطای اندازه گیری پایایی کاهش می یابد، به طوری که با حداکثر خطا، پایایی وجود ندارد و ضریب به صفر نزدیک می شود. البته چیزی به عنوان حداقل سطح قابل قبول وجود ندارد که بتوان آن را برای همه معیارها اعمال کرد، چون بسته به کاربرد تست، پایایی فرق خواهد کرد.

از آنجایی که ما هرگز نمی دانیم نمره واقعی برای هر معیار چیست، بنابراین از رابطه ی بین اندازه گیری های مکرر برای تخمین خطاهای اندازه گیری استفاده می کنیم. اگر جزء خطا کوچک باشد، اندازه گیری را پایا می دانیم و بنابراین اجازه تخمین پیوسته مقدار مورد نظر را می دهد. در تئوری اندازه گیری کلاسیک، همه ی تغییرپذیری ها در نمره فرد، به عنوان خطای اندازه گیری تفسیر می شود.

|                  |   |            |   |          |
|------------------|---|------------|---|----------|
| نمره اندازه گیری | + | نمره واقعی | = | نمره خطا |
|------------------|---|------------|---|----------|



شکل ۴-۳: هر اندازه گیری شامل یک جزء واقعی و یک جزء خطا است.

در تئوری تعمیم پذیری منابع مختلفی از تغییرپذیری برای هر معیار وجود دارد. اندازه گیری ها به طریقی بررسی می شوند که به پژوهشگر اجازه می دهد خطاهای اندازه گیری را به منابع تغییرپذیری یا فاست ها تقسیم کند.

برای درک تفاوت بین این دو رویکرد، اندازه گیری وضعیت جلو آمده ی سر با وسیله ای که اجازه اندازه گیری به سانتی متر را می دهد در نظر بگیرید. بر اساس تئوری اندازه گیری کلاسیک، هر فرد یک نمره واقعی برای وضعیت به جلو آمده سر دارد و واریاسیون در نمرات فرد، خطاهای اندازه گیری هستند. بر عکس، بر اساس تئوری تعمیم پذیری، تفاوت نمرات ممکن است به هر تعداد از عوامل مختلف مربوط باشد. برای این نمونه، ممکن است عوامل مورد نظر پژوهشگر شامل سطح ریلکسیشن فرد، سطح راحتی فرد با معاینه کننده ای که اندازه گیری ها را انجام می دهد، مهارت معاینه کننده، و دقت اندازه گیری باشند (شکل ۴-۴). به نظر می رسد که رویکرد تعمیم پذیری برای بررسی اندازه گیری ها مناسب باشد زیرا راهی برای سنجش منابع مختلف تغییرپذیری فراهم می کند.



شکل ۴-۴: واریاسیون در اندازه گیری وضعیت به جلو آمده سر می تواند ناشی از خطای

اندازه گیری (تئوری کلاسیک) یا عوامل مختلف (تئوری تعمیم پذیری) باشد.

## اجزای پایایی

تست پایایی معمولاً به یکی از دلایل زیر انجام می شود: "پایایی دستگاه" یعنی پایایی دستگاه اندازه گیری، "پایایی ریتَر" یعنی پایایی پژوهشگر/مشاهده گر/کلینیسین، "پایایی پاسخ" یعنی پایایی/ثبات متغیری که اندازه گیری می شود. بنابراین، چندین جزء پایایی که اغلب مورد بررسی قرار می گیرند عبارتند از:

۱- پایایی دستگاه (Instrument reliability)

۲- پایایی اینتر ریتَر (Intrarater reliability)

۳- پایایی اینتر ریتَر (Interrater reliability)

گرچه اغلب جدا کردن کامل این اجزا مشکل است، اما لازم است مفهوم این اجزا را بدانیم تا بتوانیم تعیین کنیم که کدام جزء یا ترکیب اجزا بررسی می شود.

## پایایی دستگاه

پایایی خود دستگاه ممکن است مورد ارزیابی قرار گیرد. سه طبقه ی مورد استفاده برای اندازه گیری عبارتند از:

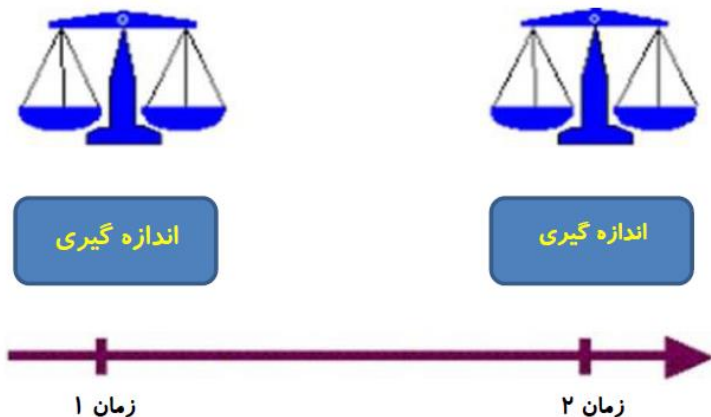
- بیوفیزیولوژیکی
- خود اظهاری (Self-report)
- مشاهده ای

ابزارهای مختلف برای انواع اندازه گیری ها مورد استفاده قرار می گیرند، و رویکرد مناسب برای تعیین پایایی دستگاه، به نوع دستگاه بستگی دارد.

## اندازه گیری بیوفیزیولوژیکی

اندازه گیری های بیوفیزیولوژیکی با استفاده از ابزارهای مکانیکی یا الکتریکی نظیر دینامومتر، گونیامتر، اسپیرومتر، و الکترومیوگراف به دست می آیند. دستگاه ها قبل از به کارگیری برای اندازه گیری، باید کالیبره شوند. در این موارد، دستگاه با یک "استاندارد طلایی" مقایسه می شود. بدنبال کالیبراسیون، پایایی دستگاه ها معمولاً به وسیله انجام اندازه گیری های مکرر در دامنه ای از مقادیر که انتظار می رود در کاربرد واقعی دستگاه یافت شوند، ارزیابی می شود. ارزیابی همبستگی نمرات در دو یا چند بار استفاده از تست، پایایی آزمون-بازآزمون نام دارد (Test-retest reliability) (شکل ۴-۵).

برای مثال در پژوهشی، پایایی آزمون-بازآزمون دینامومتر دستی (نیروسنج) با استفاده مکرر از بارهای معلوم ۱۰ تا ۶۰ کیلوگرمی تعیین شد (شکل ۴-۶). اگر فرمت خروجی دستگاه آنالوگ باشد (یعنی تست کننده مقدار را با بررسی مقیاس بر روی دستگاه تعیین می کند)، در این صورت جدا کردن پایایی دستگاه از توانایی معاینه کننده در خواندن صحیح مقیاس غیرممکن خواهد بود. اگر خروجی دستگاه با فرمت دیجیتال باشد، جدا کردن پایایی دستگاه از پایایی معاینه کننده آسان تر است چون خروجی دیجیتال جایی برای تفسیر توسط معاینه کننده باقی نمی گذارد. البته در این حالت، فرض بر این است که دستگاه به طور صحیح کالیبره شده است.



شکل ۴-۵: پایایی آزمون-بازآزمون با دو اندازه گیری در دو زمان مختلف.

## اندازه گیری های خود-اظهاری

اندازه گیری های خوداظهاری با استفاده از ابزارهایی انجام می شوند که افراد نمره خودشان از پدیده مورد بررسی را تعیین می کنند. پژوهش های نوشتاری، تست های استاندارد، مقیاس های درد، پرسشنامه های اختصاصی بیماری هایی مثل کمردرد، و مصاحبه ها، ممکن است شامل موارد خود اظهاری یا کاملاً خوداظهاری باشند (پرسشنامه های اختصاصی بیماری ها برای مثال کمردرد، گردن درد، زانو درد). شکل ۴-۷ نمونه ای از پرسشنامه خوداظهاری شاخص درجه بندی عملکردی (Index Functional Rating) برای کمردرد و گردن درد را نشان می دهد.



شکل ۴-۶: دینامومتر و وضعیت تست.

اشکال پایایی برای معیارهای خوداظهار و همچنین تست های استاندارد شده عبارتند از:

- پایایی آزمون-بازآزمون، که در آن افراد حداقل در دو زمان یا بیشتر تست میشوند.
- پایایی فرم های موازی (Parallel-form reliability) که گاهی اوقات "پایایی جایگزینی" Alternate reliability نامیده می شود. در این حالت، نسخه های مختلف یک تست، به گروهی از افراد ارائه می شود. همبستگی نمرات به دست آمده از دو نسخه، برای ارزیابی پیوستگی نتایج نسخه های جایگزین محاسبه میشود. اگر شما بخواهید پایایی تستی برای ارزیابی تفکر نقادانه را بررسی کنید، ابتدا باید مجموعه ای از آیتم های مربوط به تفکر نقادانه را ایجاد کنید. سپس سوالات را به

طور تصادفی به دو مجموعه سوال تقسیم کنید که نمایانگر اشکال موازی تست است.

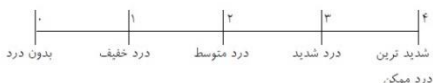
- پایایی به روش دو نیمه کردن (Split-half reliability) که در آن بخش هایی از یک تست با یکدیگر مقایسه می شوند.
- پایایی ثبات درونی (Internal consistency reliability) که در آن پیوستگی پاسخ ها به آیتم های یک تست ارزیابی می شود.

### شاخص درجه بندی عملکردی

فقط برای استفاده در مشکلات گردن و / یا کمر

جهت ارزیابی صحیح مشکل شما، ما باید بدانیم که مشکلات گردن و / یا کمر شما تا چه حد بر توانایی انجام فعالیت های روزانه شما تاثیر داشته است. برای هر یک از موارد زیر، لطفاً دور عددی را که به بهترین وجه وضعیت کنونی شما را توصیف می کند، دایره بکشید.

۱. شدت درد



۲. خوابیدن



۳. مراقبت شخصی (شست و شو، لباس پوشیدن و غیره)

شکل ۴-۷: نسخه فارسی پرسشنامه شاخص درجه بندی عملکردی

## اندازه گیری های مشاهده ای

اندازه گیری های مشاهده ای نیازمند یک ابزار انسانی، با دانش سیستماتیک از آنچه که مشاهده می شود، می باشد. معاینه کننده یا یک مشاهده گر ساده است یا نقش فعال تری داشته و از بیمار می خواهد که مجموعه ای از اعمال را انجام دهد. مشاهده الگوهای حرکتی یا سایر استراتژی های حرکتی توسط معاینه کننده یا استفاده از چک لیست ها

برای سازماندهی مشاهدات نمونه، اندازه گیری های مشاهده ای هستند. صرفنظر از فرمت مورد استفاده در اندازه گیری های مشاهده ای، برای افزایش پایایی، مشاهدات باید از لحاظ رفتاری تعریف و عملیاتی شوند.

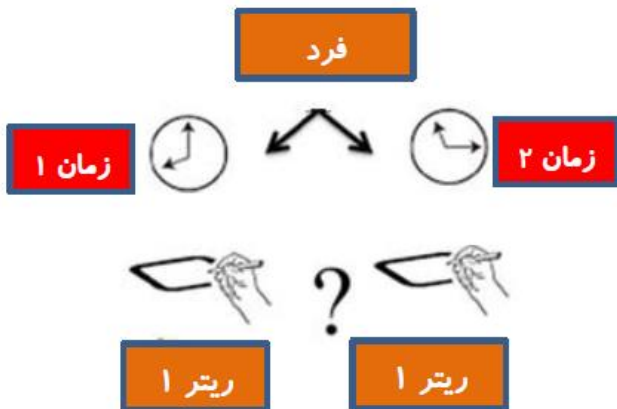
انجام تست دستی قدرت عضلات، طبقه بندی راه رفتن یا جابجایی مستقل بیماران، تعیین وضعیت رشد کودکان، و مستند کردن شرایط محیطی خانه یا کار سالمندان یا بیماران، نمونه های دیگری از معیارهایی هستند که فقط نیازمند یک ابزار انسانی با آگاهی از آنچه که مشاهده می شود می باشند. پایایی ثبات درونی برای مقیاس های مشاهده ای با آیتیم های متعدد (مانند تست های نوشتاری) می تواند بررسی گردد. چون تست کننده ابزار یک فرد است، تعیین پایایی معیارهای مشاهده ای با پایایی اینتراریتر و اینتراریتر ارتباط دارد.

### پایایی اینتراریتر (Intrarater Reliability)

پایایی اینتراریتر طبق تعریف عبارت است از پیوستگی نمرات اختصاص داده شده توسط یک ریتور در تست فرد در دو زمان مختلف (شکل ۴-۸). اگر پژوهشگری از ویدئو کلیپ برای آنالیز عملکردی مثل الگوهای گفتاری در دو زمان مختلف استفاده کند، یعنی پژوهشگر آن را در دو زمان مختلف تماشا کرده و نمره دهد، از آنجایی که رفتاری که در دو زمان ارزیابی می شود یکسان است، هر تغییری در نمرات به خطاهای اندازه گیری مربوط است.

برای اکثر اندازه گیری های انجام شده در کلینیک ها، ما توانایی نداریم که حرکت مورد نظر را عیناً مانند ویدئو بازتولید کنیم. بنابراین، در یک پژوهش، برای بررسی پایایی اینتراریتر در اندازه گیری قدرت عضلات بازکننده زانوی بیماران با استفاده از دینامومتر دستی، بیمار باید حرکت را دو یا سه مرتبه انجام دهد. در چنین شرایطی، هرگونه تغییر

پذیری در اندازه گیری به خطای ریترا یا عملکرد متفاوت فرد نسبت داده می شود. البته جدا کردن این دو اغلب مشکل است.



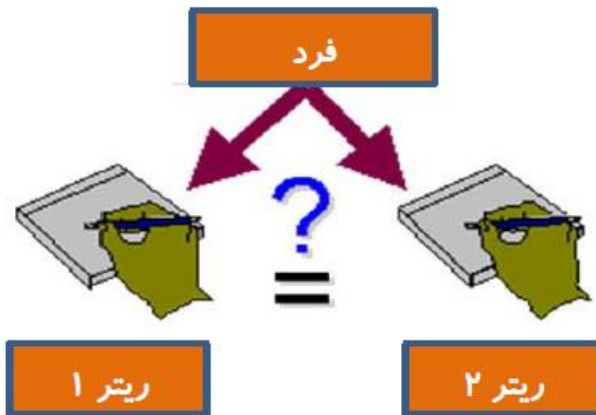
شکل ۴-۸: پایایی اینتراریتر، همبستگی بین نمرات یک ریترا که در دو زمان مختلف اندازه گیری شده.

## پایایی اینتراریتر (Interrater Reliability)

پایایی اینتراریتر عبارت است از پیوستگی در عملکرد بین ریتراهای مختلف (حداقل دو نفر) در ارزیابی و اختصاص نمرات به یک گروه از افراد در یک نقطه از زمان (شکل ۴-۹). بنابراین، در پایایی اینتراریتر، دو یا چند ریترا در باره عملکرد یک گروه از افراد در یک زمان قضاوت می کنند. مشاهده همزمان واکنش های تعادلی بیمار به عنوان بخشی از ارزیابی، و نمره دادن توسط دو یا چند کلینیسین و مقایسه بین نمرات آنها معیاری از پایایی اینتراریتر است.

اگر دو کلینیسین همزمان حرکات خودبخودی نوزاد را به عنوان بخشی از ارزیابی، مشاهده کرده و نمره بدهند، مقایسه نمرات آنها معیاری از پایایی اینتراریتر است چون آنها دقیقاً حرکات یکسانی را مشاهده و نمره داده اند. اگر آنها کودک را در دو زمان مختلف

مشاهده می کردند، جدا کردن تغییرپذیری ناشی از تفاوت معاینه کننده ها، از تغییرپذیری ناشی از تفاوت های واقعی در رفتار کودک از زمانی به زمان دیگر، غیرممکن بود. توجه داشته باشید که در بررسی پایایی اینترریتر، اگر دو یا چند ریتر بتوانند دقیقاً همان پاسخ ها را مشاهده کنند مثل وقتی که ریترها از مشاهده عملکرد افراد در ویدئوتیپ قضاوت کرده و نمره می دهند، در اینصورت همزمانی اندازه گیری ها اهمیتی ندارد. انجام پژوهش های پایایی اینترریتر و اینتراریتر برای تست ها رایج است. نمونه پژوهش های پایایی در اندازه گیری شدت اسپاستیسیته با مقیاس های کلینیکی در بیماران مبتلا به سکته مغزی قابل مشاهده است.



شکل ۴-۹: پایایی اینترریتر، همبستگی بین نمرات دو ریتر اندازه گیری شده در یک زمان.

تکنیک مثلث سازی (Triangulation) شکلی از پایایی اینترریتر است که برای مستند کردن پیوستگی نتایج پژوهش های کیفی استفاده می شود. مثلث سازی شامل مقایسه پاسخ ها از چندین منبع مختلف می باشد که در عمل به عنوان ریترهای مختلف پدیده مورد ارزیابی هستند. به عبارت دیگر، در مثلث سازی از داده های به دست آمده از منابع متعدد،

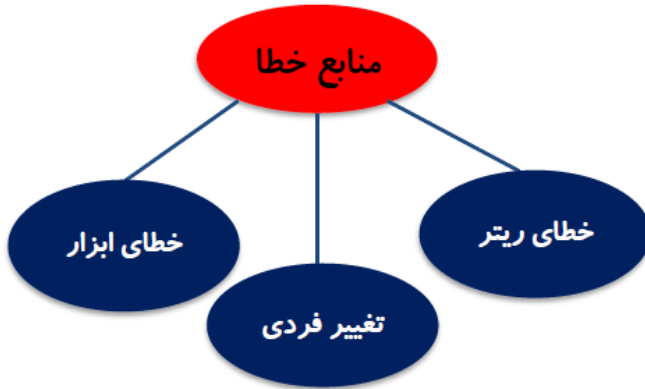
مانند مصاحبه با پزشکان، پرستاران و بیماران، در باره یک موضوع یکسان استفاده میشود. هدف مثلث سازی کنترل سوگیری (بایاز) است.

### پایایی درون فردی (Intrasubject Reliability)

"تغییرپذیری قانون زندگی است، همانطور که دو صورت یکسان نیستند، بنابراین دو بدن یکسان نیستند، و هیچ دو فردی عکس العمل یکسان ندارند. فقط تحت شرایط غیرطبیعی که ما به عنوان بیماری می شناسیم رفتار یکسانی دارند."

ویلیام اوسلر

در پایایی درون فردی، ارتباط پایایی با تغییرات واقعی در عملکرد فرد از زمانی به زمان دیگر بررسی می شود. بعضی اندازه گیری ها ممکن است فقط به این دلیل پایا به نظر نرسد که پدیده اندازه گیری شده ذاتا متغیر است. برای مثال، اندازه گیری های منفرد اسپاستیسیته در بیماران دچار ضایعات نورولوژیک ممکن است تکرارپذیر نباشد چون اسپاستیسیته یک پدیده در حال تغییر است. اندازه گیری خالص از تغییرپذیری یک فرد غیرممکن است، مگر این که یک ابزار کاملاً پایا و یک معاینه کننده کاملاً پایا داشته باشیم. بنابراین، اکثر محاسبات پایایی آزمون-بازآزمون منعکس کننده ترکیبی از خطای ابزار، خطای تست کننده، و تغییرپذیری واقعی فرد است (شکل ۴-۱۰).



شکل ۴-۱۰: منابع خطا در پایایی آزمون-بازآزمون.

## سنجش پایایی

پایایی به دو روش نسبی (Relative) و مطلق (Absolute) اندازه گیری می شود. در پایایی نسبی، رابطه بین دو یا چند مجموعه از اندازه گیری های مکرر بررسی می شود. در پایایی مطلق، تغییرپذیری نمرات در اندازه گیری های مکرر بررسی می شود.

## پایایی نسبی

پایایی نسبی بر اساس این ایده است که در اندازه گیری های پایا، موقعیت فرد در اندازه گیری های مکرر حفظ خواهد شد. برای مثال، انتظار می رود افرادی که در اولین اندازه گیری، نمره بالا دارند، حتی با تغییر نمره واقعی در طی زمان، نمراتشان نزدیک به نمرات بالای گروه باقی بماند. درجه پایایی نسبی، با محاسبه ی ضرایب همبستگی، و تعیین میزان ارتباط بین اندازه گیری های مکرر یک متغیر محاسبه می شود. ضرایب همبستگی، ثبات موقعیت فرد در گروه را ارزیابی میکنند. برای انواع مختلف داده ها، از ضرایب همبستگی مختلفی استفاده می شود (جدول ۴-۲).

جدول ۴-۲: ضرایب همبستگی

| نام ضریب همبستگی         | نوع داده | تعداد اندازه گیری ها                 |
|--------------------------|----------|--------------------------------------|
| پیرسون                   | پیوسته   | دو                                   |
| همبستگی درون گروهی (ICC) | پیوسته   | دو یا بیشتر                          |
| اسپیرمن                  | رتبه ای  | دو                                   |
| کندال تائو               | رتبه ای  | دو                                   |
| کاپا کوهن                | اسمی     | معمولا دو، می تواند بیشتر از دو باشد |

ضریب همبستگی بیانگر ارتباط کامل بین اندازه گیری های مکرر است. برای تفسیر مقادیر ضرایب همبستگی و درجه پایایی، ضوابط مختلفی وجود دارد. در بعضی متون، مقادیر همبستگی ۰/۸-۱ را بسیار پایا، و ۰/۶۹ و کمتر را پایایی ضعیف توصیف می کنند. برای پایایی حداقل استاندارد ۰/۷، و برای تفسیر ضرایب همبستگی و ICC (Intraclass Correlation Coefficient) ضوابط زیر پیشنهاد شده است:

- پایایی عالی بیشتر از ۰/۷۵ (Excellent reliability)
- پایایی خوب ۰/۶۰ تا ۰/۷۵ (Good reliability)
- پایایی متوسط ۰/۴۰ تا ۰/۵۹ (Fair reliability)

## مشکلات ضرایب همبستگی

- فرمول های زیادی برای ضرایب همبستگی وجود دارد که موجب می شود ضرایب مختلف برای داده های مشابه وجود داشته باشد.
- توافق عمومی در باره مناسب بودن فرمول های مختلف وجود ندارد.

- سطوح قابل قبول پایایی برای قضاوت در باره تغییر فردی یا گروهی فرق میکند.
  - مقدار ضریب همبستگی تحت تاثیر دامنه نمرات برای محاسبه ضریب است. ضرایب همبستگی، پیوستگی موقعیت فرد در گروه را ارزیابی می کند. اگر تغییرپذیری معیار مورد نظر در یک گروه، کم باشد، پایه ریاضی کمی برای تعیین موقعیت نسبی وجود دارد و همبستگی بین اندازه گیری های مکرر پایین خواهد بود. بنابراین، در شرایط برابر، ضریب همبستگی پایایی اینترریتر محاسبه شده برای گروهی از بیماران با مقادیر دامنه حرکتی خم کردن زانو بین  $70^{\circ}$  و  $90^{\circ}$  کمتر از گروهی با مقادیر دامنه بین  $30^{\circ}$  و  $90^{\circ}$  خواهد بود.
  - اکثر ضرایب همبستگی، برای آشکار کردن خطاهای سیستماتیک مناسب نیستند. خطای سیستماتیک خطایی است که قابل پیش بینی است. برای مثال، فرض کنید در اولین اندازه گیری محیط اندام، پژوهشگر از یک نوع متر نواری و در اندازه گیری دوم از متر نواری دیگری استفاده کند که فاقد سانتی متر اول است. به این ترتیب یک خطای اندازه گیری سیستماتیک ۱ سانتی متری در اندازه گیری دوم وجود خواهد داشت. گرچه، اگر موقعیت هر فرد در گروه حفظ شده باشد، ضرایب همبستگی علی رغم تفاوت مطلق بالا، تغییر نخواهد کرد.
- به دلیل پنج مشکل فوق، استفاده از فقط یک ضابطه برای تعیین پایایی قابل قبول، مورد سوال واقع شده است. به علاوه، جزء پایایی مورد بررسی، بر تفسیر ضریب همبستگی اثر گذار است. برای مثال، معمولاً انتظار می رود که نمرات ثبت شده در یک روز، در مقایسه با نمرات ثبت شده در طی یک دوره طولانی، تغییرپذیری کمتری داشته باشند. به همین ترتیب، ضریب پایایی اینترریتر عموماً بالاتر از ضریب پایایی اینترریتر است. در نهایت اگر پژوهشگری می خواهد تصمیم بگیرد که از دو ابزار اندازه گیری، کدام یک را استفاده کند و دریافته است که یکی از ابزارها پایایی اینترریتر ۰/۹۹ و پایایی اینترریتر دیگری ۰/۸۸ است، پس پایایی ۰/۸۸ غیرقابل قبول به نظر می رسد. یا اگر پژوهشگری بخواهد بین دو ابزار با ضرایب پایایی اینترریتر ۰/۴۵ و ۰/۶۰ یکی را انتخاب کند، ضریب ۰/۶۰ قابل

قبول است. بنابراین، به علت محدودیت های تعیین پایایی نسبی با ضرایب همبستگی، پژوهشگران باید اطلاعات نسبی را با اطلاعات مطلق کامل کنند.

در پایایی نسبی، درجه ی ثبات موقعیت افراد در اندازه گیری های مکرر آنها در یک نمونه، بررسی می شود. جدول ۴-۳ و ۴-۴ حداکثر فشار دمی را دو اندازه گیری با فاصله ۴۸ ساعت نشان می دهد. در جدول ۴-۳ گرچه تفاوت های بین دو اندازه گیری از ۱۶- تا ۲۲+ سانتیمتر آب متغیر است، اما رتبه ها بدون تغییر مانده اند. یعنی در روزهای اول و دوم، فرد شماره ۴ بالاترین، فرد شماره ۱ دومین، فرد شماره ۵ سومین، حداکثر فشار دمی را دارند. این شکل از پایایی، اغلب بوسیله ضریب همبستگی نوع، یعنی ضریب همبستگی پیرسون ارزیابی می شود. برای داده های جدول ۴-۳، مقدار ضریب همبستگی پیرسون که با  $r$  نشان داده می شود ۰/۹۴ است که بیانگر همبستگی بالا و بنابراین پایایی خوب است.

جدول ۴-۳: داده های اندازه گیری مکرر حداکثر فشار دمی با پایایی نسبی خوب.

| رتبه |       |       | حداکثر فشار دمی |       |       |
|------|-------|-------|-----------------|-------|-------|
| فرد  | روز ۱ | روز ۲ | تفاوت           | روز ۱ | روز ۲ |
| ۱    | ۱۱۰   | ۱۲۰   | + ۱۰            | ۲     | ۲     |
| ۲    | ۹۴    | ۱۰۵   | + ۱۱            | ۴     | ۴     |
| ۳    | ۸۶    | ۷۰    | - ۱۶            | ۵     | ۵     |
| ۴    | ۱۲۰   | ۱۴۲   | + ۲۲            | ۱     | ۱     |
| ۵    | ۱۰۷   | ۱۰۷   | ۰               | ۳     | ۳     |

در جدول ۴-۴، گرچه تفاوت های بین دو اندازه گیری، مشابه با مقادیر جدول ۴-۳ به نظر می رسد (یعنی ۱۵- تا ۲۲+)، اما رتبه تغییر کرده است. فرد ۴ بالاترین حداکثر فشار دمی

را در روز ۱، اما دومین رتبه را در روز ۲ دارد، فرد ۱ دومین رتبه را در روز ۱، اما پایین ترین رتبه را در روز ۲ دارد، و غیره. برای داده های جدول ۴-۴،  $I = 0.51$  است که همبستگی پایین و در نتیجه پایایی ضعیف را نشان می دهد. بنابراین، ضرایب همبستگی، اطلاعاتی در باره رابطه بین دو متغیر می دهد، و نه لزوماً در باره نزدیکی آنها.

جدول ۴-۴: داده های اندازه گیری مکرر حداکثر فشار دمی با پایایی نسبی ضعیف.

| رتبه  |       | حداکثر فشار دمی |       |       |     |
|-------|-------|-----------------|-------|-------|-----|
| روز ۲ | روز ۱ | تفاوت           | روز ۲ | روز ۱ | فرد |
| ۵     | ۲     | - ۱۵            | ۹۵    | ۱۱۰   | ۱   |
| ۳     | ۴     | + ۱۳            | ۱۰۷   | ۹۴    | ۲   |
| ۴     | ۵     | + ۱۱            | ۹۷    | ۸۶    | ۳   |
| ۲     | ۱     | ۰               | ۱۲۰   | ۱۲۰   | ۴   |
| ۱     | ۳     | + ۲۲            | ۱۲۹   | ۱۰۷   | ۵   |

همانطور که پیش تر ذکر شد، ضرایب همبستگی اطلاعاتی در باره درجه ارتباط بین دو مجموعه داده، یا پیوستگی موقعیت در دو توزیع می دهد. به شرطی که موقعیت های نسبی هر فرد، از تستی به تست دیگر یکسان باقی بماند، همبستگی بالایی به دست خواهد آمد. البته، یک ضریب همبستگی، هرگونه خطاهای سیستماتیک را آشکار نخواهد کرد. بنابراین ممکن است دو مجموعه نمره داشته باشیم که کاملاً همبستگی دارند اما تکرار پذیر نباشند (جدول ۴-۵). با داده های فرضی ارائه شده در جدول ۴-۵، علی رغم تفاوت سیستماتیک ۴۰ سانتیمتر آب برای هر فرد، ضریب همبستگی پیرسون کامل است ( $I=1$ ).

بنابراین، همبستگی فقط می گوید چگونه دو مجموعه از نمرات با یکدیگر تغییر می کنند، و وسعت توافق بین آنها را روشن نمی کند. اغلب پژوهشگران لازم است بدانند که مقادیر واقعی به دست آمده در دو اندازه گیری یکسان هستند، و فقط متناسب بودن آنها با یکدیگر کافی نیست. گرچه تحقیقات منتشر شده ای وجود دارند که تنها از همبستگی به عنوان شاخص پایایی استفاده کرده اند، ولی نتایج آنها می تواند گمراه کننده باشد. امروزه پیشنهاد شده است که از آنها دیگر نباید به طور جداگانه استفاده کرد.

### ضریب ICC (Intra-class Correlation Coefficient)

برای غلبه بر محدودیت های ضرایب همبستگی کلاسیک، ICC پیشنهاد شده است. ضریب ICC یک شاخص منفرد است که با استفاده از مقادیر واریانس به دست آمده از طریق تقسیم واریانس کل به واریانس بین فردی (Between subject) و درون فردی (Within subject) به دست آمده است. بنابراین، درجه پیوستگی (Consistency) و توافق (Agreement) بین اندازه گیری ها را نشان می دهد.

نسخه های زیادی از ICC وجود دارد که هر یک از آنها برای موقعیت های خاصی مناسب هستند. استفاده از ICC بیانگر این است که هر جزء واریانس، به طور مناسبی از داده های کافی (حداقل ۲۵ درجه آزادی)، و از نمونه ای که نماینده جمعیت مورد نظر است (و نتایج برای آنها مورد استفاده قرار خواهد گرفت) محاسبه شده است. در این نمونه، درجات آزادی عبارت است از تعداد افراد ضربدر تعداد اندازه گیری ها.

گرچه مانند سایر ضرایب پایایی، سطح قابل قبول استاندارد پایایی با استفاده از ICC وجود ندارد. اما مقدار ۰/۷۰ به عنوان حداقل پایایی استاندارد برای حداقل ۵۰ بیمار پیشنهاد شده است.

جدول ۴-۵: داده های اندازه گیری مکرر حداکثر فشار دمی با ضریب همبستگی پیرسون بالا، اما پایایی مطلق ضعیف.

| رتبه |       | حداکثر فشار دمی |       |       |       |
|------|-------|-----------------|-------|-------|-------|
| فرد  | روز ۱ | روز ۲           | تفاوت | روز ۱ | روز ۲ |
| ۱    | ۱۱۰   | ۱۵۰             | + ۴۰  | ۲     | ۲     |
| ۲    | ۹۴    | ۱۳۴             | + ۴۰  | ۴     | ۴     |
| ۳    | ۸۶    | ۱۲۶             | + ۴۰  | ۵     | ۵     |
| ۴    | ۱۲۰   | ۱۶۰             | + ۴۰  | ۱     | ۱     |
| ۵    | ۱۰۷   | ۱۴۷             | + ۴۰  | ۳     | ۳     |

دامنه مقادیر ICC بین صفر تا ۱ بوده ، و مقادیر نزدیک به یک، نشان دهنده پایایی بالاتر می باشند. استفاده از ICC وقتی مفیدتر خواهد بود که سه مجموعه مشاهدات یا بیشتر از یک نمونه منفرد یا از نمونه های مستقل به دست آمده باشد. ضریب ICC عدم مزایایی نیز دارد که باید به آن توجه داشت و بنابراین استفاده جداگانه از آن مناسب نیست. همانطور که پیش تر توضیح داده شد، هر ضریب پایایی، نسبت واریانس بین افراد، به مجموع واریانس خطا و واریانس فردی است. اگر واریانس بین افراد بالا باشد (یعنی داده ها از یک جمعیت ناهمگون باشد) پایایی به طور غیرقابل اجتنابی بالا به نظر خواهد رسید. بنابراین اگر ICC برای داده های یک گروه از افراد با دامنه وسیعی از ویژگی اندازه گیری شده استفاده شود، پایایی بالاتر از وقتی خواهد بود که به داده های گروهی با دامنه باریکی از همان ویژگی اعمال شده باشد.

## پایایی مطلق

اگر هر تست، برای اندازه گیری یک نمونه منفرد به تعداد نامحدودی استفاده شود، انتظار می رود پاسخ هایی تولید کند که از کارآزمایی به کارآزمایی دیگر کمی تغییر می کنند. از لحاظ تئوریک، این پاسخ ها می توانند رسم شوند و توزیع آنها از یک منحنی نرمال پیروی خواهد کرد (با میانگین معادل با نمره واقعی، و خطاها در بالا و پایین میانگین). هر قدر پاسخ اندازه گیری پایا تر باشد، تغییرپذیری خطا اطراف میانگین کمتر خواهد بود. بنابراین، انحراف معیار خطاهای اندازه گیری انعکاسی از پایایی پاسخ تست است که خطای معیار اندازه گیری (Standard Error of Measurement, SEM) نام دارد. مقدار SEM از فردی به فرد دیگر فرق می کند، اما فرمول آن برای محاسبه در سطح گروه عبارت است از:  $SEM = \sqrt{1 - ICC} \times \sigma$  (که  $\sigma$  انحراف معیار مشاهده شده در مجموعه نمرات تست و ICC ضریب پایایی برای آن داده است). معیار SEM شاخص پایایی مطلق است و هر قدر کوچک تر باشد پایایی بیشتر است. البته، SEM برای داده های فاصله ای مناسب است، چون با داده های نسبتی، ممکن است با افزایش مقادیر اندازه گیری شده، مقدار خطای تصادفی افزایش یابد.

پایایی مطلق، بیانگر میزان تغییرنمره در اندازه گیری های مکرر است، یعنی هر قدر کمتر تغییر کنند، پایایی بالاتر خواهد بود. معیار مورد استفاده برای اندازه گیری پایایی مطلق، علاوه بر خطای معیار اندازه گیری (SEM)، حدود توافق ۹۵٪ بلند و آلتمن می باشد (Bland and Altman's 95% limits of agreement). تفاوت های سیستماتیک باید به عنوان بخشی از خطای اندازه گیری مورد توجه قرار گیرند زیرا ما می خواهیم آنها را از تغییرات "واقعی" ناشی از درمان متمایز کنیم. بنابراین SEMagreement ارجح است. معیار خطای اندازه گیری (SEM) می تواند به کوچکترین تغییر آشکار شدنی (smallest detectable change, SDC;  $1.96 \times SEM \times \sqrt{2}$ ) تبدیل شود که نشان دهنده

کوچکترین تغییر درون فردی در نمره که می تواند به عنوان تغییر "واقعی"، بالای خطای اندازه گیری، در یک فرد تفسیر شود (SDCind). اندازه گیری SDC در گروهی از افراد (SDCGroup) را می توان از تقسیم SDCind بر  $\sqrt{n}$  محاسبه کرد. تست تی زوجی (Paired t-test) و تکنیک های آنالیز واریانس، روش های آماری برای آشکار کردن بایاز سیستماتیک بین گروه های داده ها است. البته این تستها فقط اطلاعاتی در باره تفاوت های سیستماتیک بین میانگین های دو مجموعه از داده ها می دهند، و نه در باره تفاوت های فردی. بنابراین برای بررسی پایایی، چنین تست هایی نباید بطور جداگانه استفاده شوند، بلکه باید با سایر روش ها مثل تست های توافق بلند و آلتمن بکار روند.

یک کلینیسین یا پژوهشگر برای اظهار نظر معنی دار در باره تغییر وضعیت بیمار، باید بداند چه مقدار تغییرپذیری در نمرات را می توان انتظار داشت که فقط به دلیل خطاهای اندازه گیری باشد. در یک پژوهش، برای بررسی پایایی روش بانداز به شکل ۸ انگلیسی برای اندازه گیری ادم مچ پا، ضریب پایایی اینترایتر (ضریب ICC) برای اندازه گیری محیط مچ پا ۰/۹۹ و خطای معیار اندازه گیری (SEM) محاسبه شده ۰/۴۵ سانتیمتر بود. این یافته نشان می دهد که در حدود ۹۶٪ موارد، می توان انتظار داشت که نمره واقعی محیط مچ پا در دامنه  $\pm 0.90$  cm اندازه گیری مشاهده شده، بیفتد ( $\pm 2$  SEM نمره مشاهده شده)، و انتظار داشت که اندازه گیری مکرر در محدوده  $\pm 1.27$  cm اندازه گیری اول بیفتد ( $\pm 2[\sqrt{2}]$  SEM) اندازه گیری اول).

بنابراین، ضریب همبستگی و معیارهای پایایی مطلق، انواع مختلفی از اطلاعات در باره یک معیار فراهم می کنند. ضریب همبستگی بسیار بالا بیانگر این است که پایایی نسبی معیار بالا است، و موقعیت نسبی افراد گروه در اندازه گیری های مکرر باید تقریباً بطور کامل حفظ شده باشد. اما SEM نشان می دهد که ما با این اندازه گیری ممکن است چقدر خطا بر حسب واحدهای اندازه گیری انتظار داشته باشیم. آگاهی از SEM یک

معیار، کلینسین ها را قادر می سازد که تغییرات کلینیکی را بر اساس خطای اندازه گیری قابل انتظار از معیار مورد نظر خود تفسیر کنند.

## اعتبار اندازه گیری

اعتبار اندازه گیری (Measurement Validity) عبارت است از مناسب بودن، معنی دار بودن، و مفید بودن استنباطات اختصاصی از نمرات تست. پایایی شرط لازم برای اعتبار است، اما کافی نیست. یک معیار ناپایا همچنین یک معیار نامعتبر است، چون اندازه گیری با مقدار زیادی خطا، معنی یا کاربرد کمی دارد. یک معیار پایا فقط وقتی اعتبار دارد که علاوه بر تکرارپذیر بودن، اطلاعات معنی داری فراهم کند.

اعتبار پژوهش، میزان باورپذیر بودن و مفیدبودن نتایج پژوهش را نشان می دهد. توجه داشته باشید که گرچه دو نوع اعتبار به ترتیب به اندازه گیری و طرح پژوهش (Research design) مربوط هستند، اما ارتباط آنها با کاربرد یافته ها (و نه خود یافته ها) مشابه هستند. بنابراین، اعتبار اندازه گیری، یک کیفیت مرتبط با یک ابزار یا تست مشخص نیست، بلکه یک کیفیت مرتبط با روش استخراج نتایج تست است.

برای مثال، تعداد لغات مورد استفاده یک کودک، نشانه ای کلی از اعتبار عملکرد شناختی در کودکان بدون اختلالات ارتباطی است. ولی همین معیار نمی تواند شاخص معتبری از عملکرد شناختی برای کودکانی باشد که در تولید صداهای معین مشکل دارند و لغاتی که برای صحبت کردن انتخاب و به کار می برند را محدود می کنند. اعتبار اندازه گیری اغلب به چندین طبقه تقسیم می شود: اعتبار سازه (Construct)، اعتبار محتوی (Content)، و اعتبار ملاکی (Criterion).

## اعتبار سازه

اعتبار یا روایی سازه (Construct validity) اعتبار سازه های انتزاعی یا مفاهیم پایه ی معیارهاست. به عبارت دیگر، اعتبار سازه میزان اندازه گیری مفاهیم و فرضیات نظری پایه ای است که یک تست بر آنها استوار است. برای مثال، قدرت (Strength) سازه ای است که آن را می توان به عنوان مفهومی برای توانایی حرکت یک قسمت از بدن بر خلاف جاذبه، توانایی تولید گشتاور، توانایی بلند کردن یک وزنه معین به تعداد دفعات معین در یک دوره زمانی معین، یا توانایی انجام بعضی کارهای عملکردی تعریف کرد. تست های دستی قدرت عضلات، تست های ایزو کینتیک، تست های انجام کار، و تست های عملکردی (Functional)، هر یک معیاری معتبر برای یکی از مفاهیم و تعاریف قدرت یا عمل عضله هستند.

برای به حداکثر رساندن اعتبار سازه، پژوهشگران ابتدا باید در باره سازه هایی که میخواهند اندازه گیری کنند بسیار روشن باشند. اگر "قدرت" یک سازه مهم در پژوهش باشد، مفهوم آن به عنوان قدرت عملکردی، قدرت استاتیک، قدرت اکستریک، یا بعضی جنبه های دیگر این سازه، ممکن است بسیار گسترده تعریف شده باشد. وقتی که سازه پایه مورد نظر مشخص گردید و تعریف شد، باید از لحاظ عملی تعریف شود تا قابل اندازه گیری باشد. تعریف عملی یک سازه، توصیفی اختصاصی روشی است که سازه مورد نظر در پژوهش، ارائه شده یا اندازه گیری می شود.

در پژوهشی، برای بررسی خستگی افراد مبتلا به ام اس پژوهشگران ابزار اندازه گیری را بر پایه تعریف عملی خستگی به عنوان "فقدان انرژی فیزیکی و یا ذهنی که خود فرد یا مراقب او آن را درک می کند و مانع فعالیت های معمول و دلخواه بیمار می شود" انتخاب کردند. مقیاس استفاده شده در پژوهش، Fatigue Impact Scale بود که با سازه مورد نظر انطباق داشت و شامل زیرمقیاس های مربوط به اثر فیزیکی، شناختی، و اجتماعی خستگی بود. گرچه ارائه تعاریف عملی برای اعتبار سازه ضروری است، اما اعتبار سازه را

تضمین نمی‌کند. برای مثال، ممکن است بحث شود که در پژوهش ذکر شده، پژوهشگران ادراکات افراد مبتلا به اسکروز متعدد و مراقبین آنها (برای کسانی که در همه فعالیت‌ها مستقل نبودند) را تعیین نکردند و به این ترتیب در گنجاندن همه عناصر تعریف عملی در طراحی اندازه‌گیری‌های پژوهش موفق نبودند. ارائه تعاریف عملی مورد استفاده در یک پژوهش، به پژوهشگران اجازه می‌دهد که عقیده خودشان از اعتبار اندازه‌گیری را شکل دهند.

علاوه بر وظیفه هوشمندانه تعیین انطباق بین سازه‌ها و اندازه‌گیری آنها، راه دیگر تعیین اعتبار سازه استفاده از مقایسات "گروه-شناخته شده" (known-group) است که نوعی از اعتبار تمایزی می‌باشد. در این رویکرد، گروه‌هایی از افراد که انتظار می‌رود در یک اندازه‌گیری، عملکرد متفاوتی داشته باشند، مقایسه می‌شوند تا تعیین شود که آیا تفاوت‌های فرض شده تحقق می‌یابند. در یک پژوهش، برای بررسی اعتبار مقیاس اصلاح شده اصلاح شده اشورث (Modified Modified Ashworth Scale) در ارزیابی شدت اسپاستیسیتة عضلات خم‌کننده مچ دست پس از سکته مغزی، پژوهشگران نشان دادند که میانگین معیارهای نوروفیزیولوژیکی  $Hslp/Mslp$  و  $Hmax/Mmax$  در بیماران با درجات شدیدتر اسپاستیسیتة، بالاتر بود. به این ترتیب اعتبار سازه مقیاس MMAS بطور موفقیت‌آمیزی در تمایز بین بیماران با درجات مختلف اسپاستیسیتة نشان داده شد. مثالی دیگر اعتبار سازه، مقیاسی است که قادر باشد افراد سالمند ساکن در محیط‌های مختلف (مثل خانه سالمندان، مهمانسرا، و مرکز مراقبت از سالمندان) را متمایز کند.

## اعتبار محتوی

اعتبار محتوا (Content validity) نشان می‌دهد که یک معیار تا چه میزان نمایانگر کاملی از محتوی مورد نظر است. اعتبار محتوا بیشتر در ابزارهای خوداظهاری یا مشاهده‌ای مطرح است تا ابزارهای بیوفیزیولوژیکی. وقتی که دانشجویان بعد از امتحان بیان می‌کنند

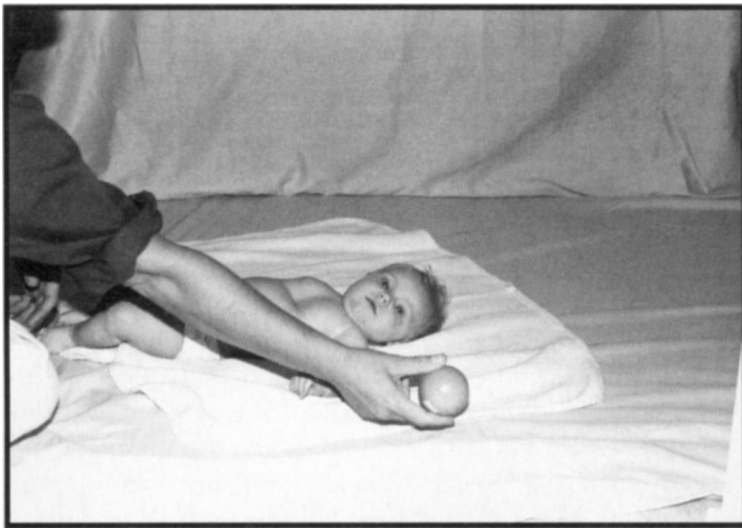
که تعداد زیادی از سوالات، از یک مبحث خاص بوده، در حقیقت در باره اعتبار محتوایی امتحان صحبت می کنند، یعنی آنها سوال می کنند که آیا تاکید امتحان بر روی موضوعی خاص، نمایانگر صحیحی از محتوای درس است؟

در یک پژوهش، شکلی از اعتبار محتوای تست TIMP (Test of Infant Motor Performance) به نام "ارتباط اکولوژیکی" آیتم های تست مورد بررسی قرار گرفت. در ارتباط اکولوژیکی، ارتباط بین نیازمندی های محیطی نوزادان در زندگی روزانه و نیازمندی های ناشی از انجام آیتم های تست TIMP بررسی گردید. آنها دریافتند که ۹۸٪ آیتم های تست با نیازمندی های محیطی در فعالیت های روزانه طبیعی نوزاد منطبق است (شکل 4-11). برای بررسی اعتبار محتوی، پژوهشگران ارائه دهنده تست همچنین از کمیته صاحب نظران (Expert panel) خواستند که آیتم های تست را بررسی کنند تا مطمئن گردند که نیازمندی های محیطی ممکن در تست گنجانده شده اند. وقتی که پژوهشگری یک پرسشنامه یا یک مقیاس عملکردی را طراحی می کند، باید ارزیابی اعتبار محتوا توسط همتایان صاحب نظر در آن حیطه را داشته باشد یا آن را در محیطی طبیعی به عنوان بخشی از تست پایلوت (Pilot test) ارزیابی کند. این روش های ارزیابی ممکن است به اضافه کردن آیتم ها، حذف آیتم های نامربوط یا مازاد، یا باز ارزیابی موضوعات مشخصی که بر آنها تاکید شده منجر شود.

## اعتبار ملاکی

اعتبار ملاکی (Criterion validity)، میزان همبستگی سیستماتیک یک معیار با معیارها یا پیامدهای دیگر می باشد. همانطور که در پایایی نسبی، نمرات حاصل از انجام مکرر یک تست و اندازه گیری مقایسه می شوند، در اعتبار ملاکی نیز، نمرات حاصل از به کارگیری معیارهای مختلف مقایسه می شوند. پایه ریاضی برای تعیین درجه ارتباط دو معیار مختلف، با پایه ریاضی برای تعیین ارتباط بین اندازه گیری مکرر یک تست، مشابه

است. بنابراین، ضرایب همبستگی مورد استفاده برای تعیین پایایی نسبی، اغلب برای سنجش اعتبار ملاکی نیز استفاده می شوند. مفاهیم اپیدمیولوژیکی ویژگی (Specificity) و حساسیت (Sensitivity) نیز برای تعیین اعتبار ملاکی مفید هستند. اعتبار ملاکی بر پایه زمانبندی معیارهای مختلف، به اعتبار همزمان (Cocurrent validity) و اعتبار پیش بینی کننده (Predictive validity) تقسیم می شود.



شکل ۴-۱۱: نمونه ای از تست *Infant Motor Performance* در ارزیابی اعتبار محتوی اکولوژیکی.

نوزاد در وضعیت طاق باز خوابیده با سر بطور کامل چرخیده به راست یا چپ. با استفاده از یک توپ قرمز در حال حرکت به عنوان یک محرک بینایی، نوزاد با چشم توپ را دنبال می کند و چرخش سر به سمت مقابل برانگیخته می شود.

## اعتبار ملاکی همزمان (Cocurrent validity)

اعتبار ملاکی همزمان، نوعی اعتبار یا روایی وابسته به ملاک است که در آن یک ابزار یا رویکرد جدید با یک استاندارد اندازه گیری، در شرایطی که هر دو ابزار در یک زمان برای اندازه گیری استفاده شده اند، مقایسه می گردند. مقایسه روشی جدید برای آشکار کردن شکستگی میچ پا با وجود ۱۵ mm افزایش میچ پا در عکس رادیوگرافی ساده در مقایسه با "استاندارد طلایی" سی تی اسکن، نمونه ای از بررسی اعتبار ملاکی همزمان است. در اعتبار سنجی مقیاس Back Pain Functional Scale (BPFS) به زبان فارسی، پژوهشگران برای بررسی اعتبار ملاکی همزمان، از همبستگی نمرات نسخه فارسی BPFS با معیار درد Numerical Rating Scale در مبتلایان به کمردرد مزمن استفاده کردند. اعتبار ملاکی همزمان نسخه فارسی BPFS، با همبستگی معنی دار بین دو مقیاس نشان داده شد. گرچه باید توجه داشت که اغلب استاندارد طلایی واقعی که معیار جدید نسبت به آن بتواند مقایسه شود وجود ندارد.

## اعتبار پیشینی کننده (Predictive validity)

اعتبار ملاکی پیشینی کننده به این نکته اشاره دارد که آیا تست انجام شده در یک نقطه از زمان، پیش بینی کننده وضعیت در آینده است. در پژوهش مربوط به اعتبار پیش بینی کننده تست TIMP، نمرات به دست آمده از این تست تا ۴ ماهگی نوزادان، با نمرات تست BOTMP (Bruininks-Oseretsky Test of Motor Proficiency) همان کودکان در ۴ تا ۷ سالگی مقایسه شد. نتیجه نشان داد که بیشتر از ۸۹٪ کودکان بوسیله TIMP به طور صحیح طبقه بندی شدند. یعنی ۸۹٪ کودکان "مثبت واقعی" (طبقه بندی شده در گروه در معرض خطر تاخیر رشدی بوسیله TIMP و بعداً در معرض خطر تاخیر رشدی بوسیله BOTMP طبقه بندی شدند) یا "منفی واقعی" بودند (طبقه بندی شده در گروه در معرض خطر تاخیر رشدی نبودن بوسیله TIMP و بعداً طبقه بندی شده به عنوان

تاخیر رشدی نداشتن بوسیله BOTMP). فقط ۱۱٪ کودکان "مثبت کاذب" (طبقه بندی شده در معرض خطر بوسیله TIMP و تاخیر نداشتن بوسیله BOTMP) یا "منفی کاذب" (طبقه بندی شده در معرض خطر نبودن بوسیله TIMP اما تاخیر نداشتن بوسیله BOTMP) بودند. تعیین اعتبار پیش بینی کننده یک معیار غربالگری (Screening measure) مثل TIMP یک فرایند اساسی اما زمانبر است. فرضیه بسیاری از تست های غربالگری این است که آنها اجازه می دهند بعضی پدیده ها که معمولاً تا مراحل بعدی بیماری آشکار نمیشوند، شناسایی گردند. مفید بودن چنین معیارهایی را نمی توان تعیین کرد، مگر این که اعتبار پیش بینی کننده آنها معلوم شود. در کاربرد کلینیکی، دانشمندان و کلینیسین ها علاقه مند به اعتبار پیشبینی کننده تست های مختلف برای پیش آگهی منطقی وضعیت بیماران هستند.

## پاسخ دهی به تغییرات

مسئله مهم سوم، پاسخ دهی به تغییر (Responsive to change) یک اندازه گیری است. کلینیسین ها عموماً امیدوار هستند که در بیماران تغییری ایجاد کنند، بنابراین به معیارهایی که تغییرات را در افراد نشان می دهند علاقه مند هستند. دو نفر را در نظر بگیرید، هر دو آنها اهدافی مربوط به تغییر وزن دارند. یکی از آنها مردی است که ۱۰۰ کیلو گرم وزن دارد و می خواهد وزنش را تا ۸۵ کیلوگرم کاهش دهد. دیگری یک نوزاد نارس است که ۲۵۰۰ گرم وزن دارد و می جنگد تا وزنش به ۴۰۰۰ گرم برسد. برای مردی که رژیم دارد ترازویی با فواصل ۱ کیلوگرم پاسخ دهی کافی برای مستند کردن کاهش وزنش دارد، ولی برای نوزاد نارس، یک ترازوی معمولی پاسخ دهی کافی برای اندازه گیری تغییرات وزن اندکی که انتظار می رود رخ بدهد را نخواهد داشت، بنابراین یک ترازوی دقیق تر که بتواند تغییرات در حد گرم را اندازه گیری کند مورد نیاز است. پس با توجه به ویژگی های افراد مورد بررسی، باید روشهایی برای اندازه گیری انتخاب شوند که پاسخ

دهی به تغییرات داشته باشند. پاسخ دهی به تغییرات، همچنین "حساسیت به تغییر" (Sensitivity to change) نامیده می شود.

پاسخ دهی به تغییر، طبق تعریف عبارت است از میزانی که از لحاظ عملی یا نظری، تغییرات معنی دار در "وضعیت" فرد یعنی تغییرات واقعی در مقادیر مشاهده شده انعکاس می یابد. پاسخ دهی یک معیار، به پایایی و اعتبار آن معیار مربوط است. اگر معیاری خیلی پایا نبوده و SEM بزرگ داشته باشد، پس تغییرات آن باید بزرگتر و بیشتر از خطای اندازه گیری باشند تا بیانگر تغییرات واقعی در وضعیت کلینیکی بیماران یا مشارکت کنندگان در پژوهش باشد. بالعکس، برای معیاری خیلی پایا با SEM کوچک، حتی تغییرات کوچک در نمرات، می توانند نمایانگر تغییرات واقعی در وضعیت فرد باشند، و در نتیجه "پاسخ دهی" به تغییرات دارد.

معیاری که انطباق مفهومی خوبی با سازه ای که اندازه گیری می شود نداشته باشد (یعنی اعتبار سازه ندارد)، بعید است که پاسخ دهی به تغییرات داشته باشد. برای مثال، شاخص بارتل برای فعالیت های زندگی روزانه (Barthel Index) به عنوان ابزار ژنریک ارزیابی پیامدها برای افراد دچار ضایعه نخاعی استفاده شده است. ولی اختصاصی بودن آیتم های آن برای مستند کردن تغییرات کوچک در عملکرد بعضی افراد کوادری پلژی (فلج چهار اندام) مورد سوال قرار گرفته است. پژوهشگران در پاسخ به این نگرانی ها، برای بیماران کوادری پلژی شاخصی اختصاصی تولید کردند که Quadriplegia Index of Function (QIF) نام دارد و پاسخ دهی آن به تغییرات را در مقایسه با شاخص بارتل بررسی کردند. آنها دریافتند که میانگین درصد بهبودی، از پذیرش تا ترخیص در یک بیمارستان توانبخشی، برای افراد کوادری پلژی و بر اساس QIF ۴۶٪، و برای شاخص بارتل فقط ۲۰٪ بود، که پاسخ دهی به تغییر بهتر برای QIF را نشان می داد.

پاسخ دهی به تغییر، همچنین تابع تعداد مقادیری است که یک مقیاس می تواند بگیرد. اگر یک مقیاس بتواند مقادیر کمی بگیرد (مثلا مقیاس درجه بندی A, B, C, D, E، باید تغییرات نسبتا بزرگ در پرفورمانس ایجاد شود تا تغییری در درجه آن ثبت گردد.

بالعکس، اگر یک مقیاس درجه بندی شامل درجات + و - باشد، پس تغییرات نسبتاً کوچک در پرفورمانس ممکن است از درجه  $B-$  به  $B$  و  $B+$  تغییر پیدا کند.

اثرات سقف و کف (Ceiling and floor effects) همچنین بر پاسخ دهی یک معیار موثر است. اثر کف (Floor effect) وقتی رخ می دهد که نمرات فرد، نمرات کف یک مقیاس (کمترین مقادیر) است و هیچگونه کاهش بیشتر در مقدار مورد اندازه گیری نمی تواند ثبت شود. اثرات سقف (Ceiling effect) در نمرات بالای یک مقیاس رخ میدهد، به طوری که هیچگونه بهبودی بیشتری نمی تواند ثبت شود و علیرغم این که نمره نهایی فرد، "زیر" نمره کف یا "بالای" نمره سقف است، ممکن است تغییر واقعی با مقیاس مذکور به عنوان بدون تغییر ثبت شود. در بررسی رابطه بین معیارهای ژنریک کیفیت زندگی مربوط به سلامت و عملکرد راه رفتن، پژوهشگران اثرات سقف و کف بزرگی در بعضی زیرمقیاس های SF-36 در جمعیت سربازان دچار نوروپاتی محیطی مشاهده کردند. برای مثال، آنها دریافتند که ۴۶٪ نمونه های آنها نمره کف زیرمقیاس نقش محدودیت فیزیکی و ۳۰٪ آنها نمره سقف زیرمقیاس عملکرد اجتماعی را گرفتند. از آنجایی که این طرح، پژوهشی طولی برای بررسی اثر فیزیکی و اجتماعی نوروپاتی محیطی بوده است، آنها قادر به آشکار کردن کاهش در محدودیت های نقش فیزیکی، یا بهبودی عملکرد اجتماعی برای تعداد زیادی از افراد شرکت کننده در پژوهش نبودند. وقتی پژوهشگران یا کلینیسین ها می خواهند از معیارهایی استفاده کنند که پاسخ دهی به تغییر داشته باشند، باید معیارهایی را انتخاب کنند که پایایی و اعتبار خوب در جمعیت مورد بررسی یا تحت درمان داشته باشند. چنین معیارهایی بعید است که اثرات سقف و کف بزرگی برای جمعیت مورد نظر داشته باشند. مانند پایایی و اعتبار، پاسخ دهی به تغییر یک ویژگی ثابت برای معیار نیست، بلکه بر اساس شرایط و جمعیتی که در آن استفاده می شود، تغییر می کند.

دو اصطلاح وجود دارد که برای نشان دادن کاربرد کلینیکی تغییرات استفاده می شوند. این اصطلاحات عبارتند از " حداقل تفاوت آشکار کردنی" ( Minimal Detectable Difference, MDD) و "حداقل تفاوت مهم از لحاظ کلینیکی" ( Minimal Clinically Importance Difference, MCID).

حداقل تفاوت آشکار کردنی یا حداقل تغییر آشکار کردنی ( Smallest Detectable Change) عبارت است از کوچکترین تغییری که از خطای اندازه گیری بیشتر است. حداقل تفاوت آشکار کردنی بطور ریاضی محاسبه می شود. حداقل تفاوت مهم از لحاظ کلینیکی نیز معیاری از کوچکترین تغییر معنی دار برای بیمار است. البته، این معنی دار بودن فردی است و به تعداد دیدگاه های مختلف بستگی دارد. آیا بیمار تغییر را معنی دار می داند؟ یا مراقب بیمار (عضو خانواده یا فیزیوتراپیست) تغییر را معنی دار می داند؟

## خلاصه

اندازه گیری یک فرایند سیستماتیک برای تمایز چیزها از هم است. اندازه گیری ها میتوانند به وسیله مقیاس های اسمی، رتبه ای، فاصله ای، و نسبی، از نوع گسسته یا پیوسته، و با چهارچوب مبتنی بر نرم یا مبتنی بر ضابطه شناسایی شوند. اجزای پایایی عبارتند از دستگاه، تست کننده، و تغییرپذیری فردی. معیارهای نسبی پایایی ضرایب همبستگی هستند، و معیار مطلق پایایی، خطای معیار اندازه گیری (SEM) است. اعتبار یا روایی، معنی دار بودن و سودمندی کاربرد یک اندازه گیری است. اجزای اعتبار عبارتند از اعتبار سازه، اعتبار محتوا، و اعتبار ملاکی. پاسخ دهی به تغییر، میزان انعکاس تغییر وضعیت فرد، در تغییر مقدار اندازه گیری شده می باشد. پاسخ دهی به تغییر، به پایایی و اعتبار یک معیار بستگی دارد و اختصاصی جمعیتی است که در آنها اندازه گیری انجام می شود.

Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. *Phys Ther* 76:248-59,1996.

Ansari NN, Tarameshlu M, Ghelichi L. Dysphagia In Multiple Sclerosis Patients: Diagnostic And Evaluation Strategies. *Degener Neurol Neuromuscul Dis* 10:15-28,2020.

Ansari NN, Feise RJ, Naghdi S, et al. The functional rating index: reliability and validity of the Persian language version in patients with neck pain. *Spine* 37:E844-8,2012.

Ansari NN, Feise RJ, Naghdi S, et al. The functional rating index: reliability and validity of the Persian language version in patients with low back pain. *Spine* 36:E1573-7,2011.

Ansari NN, Fathali M, Naghdi S, et al. Effect of pulsed ultrasound on chronic rhinosinusitis: a case report. *Physiother Theory Pract* 26:558-63,2010.

Ansari NN, Naghdi S, Hasson S, et al. Inter-rater reliability of the Modified Modified Ashworth Scale as a clinical tool in measurements of post-stroke elbow flexor spasticity. *NeuroRehabilitation* 24:225-9,2009.

Ansari NN, Naghdi S, Younesian P, et al. Inter- and intrarater reliability of the Modified Modified Ashworth Scale in patients with knee extensor poststroke spasticity. *Physiother Theory Pract* 24:205-13,2008.

Baty J, Hwang H, Ding Z, et al. The effect of a carbohydrate and protein supplement on resistance exercise performance, hormonal response, and muscle damage. *Strength Cond Res* 21:321-9,2007.

Blake ML. Inferencing processes after right hemisphere brain damage: effects of contextual bias. *J Speech Lang Hear Res* 52:373-84,2009.

Bohannon RW, Cooper J. Total knee arthroplasty: evaluation of an acute care rehabilitation program. *Arch Phys Med Rehabil* 74:1091-4,1993.

Bruton A, Conway JH, Holgate ST. Reliability: What is it and how is it measured?' *Physiotherapy* 86:94-9,2000.

Brady MC, Godwin J, Enderby P, et al. Speech and Language Therapy for Aphasia After Stroke: An Updated Systematic Review and Meta-Analyses. *Stroke* 47:e236-e237,2016.

Cai Z, Yang Y, Zhang J. Effects of physical activity on the progression of diabetic nephropathy: a meta-analysis. *Biosci Rep* 41: BSR20203624,2021.

Carter RE, Lubinsky J, Domholdt E. *Rehabilitation Research: Principles and Applications*, 4th ed, Missouri, 2011, Elsevier.

Case-Smith J. Effectiveness of school-based occupational therapy intervention on handwriting, *Am J Occup Ther* 56:17-25,2002.

Chin A, Paw MJM, de Jong N, et al. Physical exercise and/or enriched foods for functional improvement in frail, independently living elderly: a randomized controlled trial. *Arch Phys Med Rehabil* 82:811-7,2001.

Clark TWI, Janzen DL, Logan PM, et al. Improving the detection of radiographically occult ankle fractures: positive predictive value of an ankle joint effusion. *Clin Radiol* 51:632-6,1996.

Clemson L, Cumming RG, Roland MA. Case-control study of hazards in the home and risk of falls and hip fractures. *Age Aging* 25:97-101,1996.

Collier SE, Thomas JJ. Range of motion at the wrist: a comparison study of four wrist extension orthoses and the free hand. *Am J Occup Ther* 56:180-4,2002.

Cushman DG, Dumas HM, Haley SM, et al: Re-admissions to inpatient paediatric pulmonary rehabilitation, *Pediatr Rehabil* 5:133-9,2002.

de Goede CT, Keus SHJ, Kwakkel G, et al. The effects of physical therapy in Parkinson's disease: a research synthesis. *Arch Phys Med Rehabil* 82:509-15,2001.

Domholdt E. *Physical Therapy Research: Principles and Applications*, 2nd ed, Philadelphia, W.B. Saunders Co, 2000.

Dunn KS. The effect of physical restraints on fall rates in older adults who are institutionalized.

*J Gerontol Nurs* 27:40-8,2001.

Fakhari Z, Ansari NN, Naghdi S, et al. A single group, pretest-posttest clinical trial for the effects of dry needling on wrist flexors spasticity after stroke. *NeuroRehabilitation* 40:325-36,2017.

Fallat L, Grimm DJ, Saracco JA. Sprained ankle syndrome: prevalence and analysis of 639 acute injuries, *J Foot Ankle Surg* 37:280-5,1998.

Findley TW, Daum MC. Research in physical medicine and rehabilitation. III. The chart review, or how to use clinical data for exploratory retrospective studies, *Am J Phys Med Rehabil* 68:150-157,1989.

Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2003.

Guccione AA, Felson DT, Anderson J, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. *Am J Public Health* 84:351-358,1994.

Harding JE, Miles FK, Becroft DM, et al. Chest physiotherapy may be associated with brain damage in extremely premature infants. *J Pediatr* 132:440-4,1998.

Hyde ML. Reasonable psychometric standards for selfreport outcomes measures in audiological rehabilitation. *Ear Hear* 21:24S-36S,2000.

Janson C, Chinn S, Jarvis D, et al. Effect of passive smoking on respiratory symptoms, bronchial responsiveness, lung function, and total serum IgE in the European Community Respiratory Health Survey: a cross-sectional study. *Lancet* 358:2103-9,2001.

GeerdinkJJ, Hopkins B, Beek WJ, et al. The organization of leg movements in preterm and full-term infants after term age. *Dev Psychobiol* 29:335-51,1996.

Jahnsen R, Villien L, StanghelleJK, et al. Fatigue in adults with cerebral palsy in Norway compared with the general population. *Dev Med Child Neurol* 45:296-303,2003.

Jansen CWS, Minerbo G. A comparison between early dynamically controlled mobilization and immobilization after flexor tendon repair in zone 2 of the hand: preliminary results. *J Hand Ther* 3:20-25,1990.

Lake DA. Student performance and perceptions of a lecture- based course compared with the same course utilizing group discussion. *Phys Ther* 81:896-902,2001.

Lascaratos JG, Trompoukis C, Segas JV, et al. Professor Nicolas Taptas (1871-1955): a pioneer of post-laryngectomy voice rehabilitation. *Laryngoscope* 113:702-705,2003.

Law M, Stewart D, Pollack N, et al. Guidelines for critical review form- quantitative studies. Accessed August 19,2021: Availble at <https://canchild.ca/system/tenon/assets/attachments/000/000/366/original/quantguide.pdf>.

Leucht S, Kissling W, Davis JM. How to read and understand and use systematic reviews and meta-analyses. *Acta Psychiatr Scand* 2009 119:443-50, 2009.

Lu-Yao GL, Baron JA, Barrett JA, et al. Treatment and survival among elderly Americans with hip fractures: a population-based study. *Am J Public Health* 84:1287-1291,1994.

Mack R, Salmoni A, Viverais-Dressler ZG, et al. Perceived risks to independent living: the views of older, community- dwelling adults. *Gerontologist* 37:729-736,1997.

Mahmood SS, Levy D, Vasan RS, et al. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet* 383:999-1008, 2014.

Manchikanti L, Benyamin RM, Helm S, et al. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: part 3: systematic reviews and meta-analyses of randomized trials. *Pain Physician* 12:35-72,2009.

McCrum-Gardner E. Sample size and power calculations made simple. *IJTR* 17:10-14, 2010.

Palisano RJ. The case report, case study, and single subject design. *Phys Occup Ther Pediatr* 27:1-5,2007.

Mathiowetz V. Test-retest reliability and convergent validity of the Fatigue Impact Scale for persons with multiple sclerosis. *Am J Occup Ther* 57:389-395, 2003.

Moore JH, Gieck JH, Saliba EN, et al: The biophysical effects of ultrasound on median nerve distal latencies *Electromyogr Clin Neurophysiol* 40:169-180,2000.

Mawdsley RH, Hoy DK, Erwin PM. Criterion-related validity of the figure-of-eight method of measuring ankle edema. J Orthop Sports Phys Ther 30:149-153,2000.

Morris JA, Mackenzie EJ, Edelstein S1. The effect of preexisting conditions on mortality in trauma patients, JAMA 263:1942-1946,1990.

Murney ME, Campbell SK. The ecological relevance of the Test of Infant Motor Performance elicited scale items. Phys Ther 78:479-489,1998.

Murphy S, Tickle-Degnen L. The effectiveness of occupational therapy-related treatments for persons with Parkinson's disease: a meta-analytic review. Am J Occup Ther 55:385-392,2000.

Naghdi S, Ansari NN, Mansouri K, et al. Neurophysiological examination of the Modified Modified Ashworth Scale (MMAS) in patients with wrist flexor spasticity after stroke. Electromyogr Clin Neurophysiol 48:35-41,2008.

Naghdi S, Ansari NN, Farhadi M. J Phys Ther Sci 20:233-238,2008.

Nakhostin Ansari N, Naghdi S, Forogh B, et al. Development of the Persian version of the Modified Modified Ashworth Scale: translation, adaptation, and examination of interrater and intrarater reliability in patients with poststroke elbow flexor spasticity. Disabil Rehabil 34:1843-7,2012.

Nakhostin Ansari N, Naghdi S, Habibzadeh F, et al. Persian translation and validation of the Back Pain Functional Scale. Physiother Theory Pract 34:3,223-30,2018.

Nilsson L, CarlssonJ, Danielsson A, et al. Walking training of patients with hemiparesis at an early stage after stroke: a comparison of walking training on a treadmill with body weight support and walking training on the ground. Clin Rehabil 15:515-27,2001.

Novak I, Cusick A, Lowe K. A pilot study on the impact of occupational therapy home programming for young children with cerebral palsy. *Am J Occup Ther* 61:463-8,2007.

Page S: Exploring competitive orientation in a group of athletes participating in the 1996 paralympic trials, *Percept Mot Skills* 91:491-502,2000.

Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet* 351:47-52,1998.

Powers CM, Boyd LA, Fontaine CA, et al. The influence of lower-extremity muscle force on gait characteristics in individuals with below-knee amputations secondary to vascular disease. *Phys Ther* 76:369-77,1996.

Prencipe M, Ferretti C, Casini AR, et al. Stroke, disability, and dementia: results of a population survey. *Stroke* 28:531-6,1997.

Pulvermuller F, Neininger B, Elbert T, et al. Constraint induced therapy of chronic aphasia after stroke. *Stroke* 32:1621-6,2001.

Robey RR. A meta-analysis of clinical outcomes in the treatment of aphasia. *J Speech Lang Hear Res* 41:172-87,1998.

Richardson LA, Buckenmeyer PJ, Bauman BD, et al. Contemporary cardiac rehabilitation: patient characteristics and temporal trends over the past decade. *J Cardiopulm Rehabil* 20:57-64, 2000.

Roberts P, Priest H. Reliability and validity in research. *Nurs Stand* 20:41-5,2006.

Rosa L, Rosa E, Samer L, et al. A close look at therapeutic touch. *JAMA* 279:1005-10,1998.

Runeson L, Haker E. Iontophoresis with cortisone in the treatment of lateral epicondylalgia (tennis elbow)-a doubleblind study. Scand J Med Sci Sports 12:136-42,2002.

Schatz P, Hillary F, Moelter S, et al. Retrospective assessment of rehabilitation outcome after traumatic brain injury: development and utility of the functional independence level. J Head Trauma Rehabil 17:510-25,2002.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psycho I Bull 86:420-428, 1979.

Smith DG, Domholdt E, Coleman KL, et al. Ambulatory activity in men with diabetes: relationship between self-reported and real-world performance-based measures. J Rehabil Res Dev 41:571-580,2004.

Son KJ, Son HR, Park B, et al. A Community-Based Intervention for Improving Medication Adherence for Elderly Patients with Hypertension in Korea. Int J Environ Res Public Health 16:721,2019.

Soriano JB, Ancochea J, Miravittles M, et al. Recent trends in COPD prevalence in Spain: a repeated cross-sectional survey 1997-2007. Eur Respir J 36:758-65,2010.

Stratford PW, Norman GR, McIntosh JM. Generalizability of grip strength measurements in patients with tennis elbow, Phys Ther 69:276-81,1989.

Taylor VM, Deyo RA, Ciol M, et al. Patient-oriented outcomes from low back surgery: a community-based study, Spine 25:2445- 2452, 2000.

Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60(1):34-42, 2007.

Thomas M, McKinley RK, Freeman E, et al. Prevalence of dysfunctional breathing in patients treated for asthma in primary care: cross sectional survey. *BMJ* 2001 322:1098-100,2001.

Timm KE. Postsurgical knee rehabilitation: a five year study of four methods and 5381 patients. *Am J Sports Med* 16:463-8,1988.

Tong AYC, Man DWK. The validation of the Hong Kong Chinese Version of the Lawton Instrumental Activities of Daily Living scale for institutionalized elderly persons. *OTJR Occup Particip Health* 22:132-42,2002.

Ulione MS. Health promotion and injury prevention in a child development center. *Pediatr Nurs* 12:148-54,1997.

van Dijk CN, Lim LSL, Marti RK, et al. Physical examination is sufficient for the diagnosis of sprained ankles. *J Bone Joint Surg Br* 78:958-62,1996.

Visintin M, Barbeau H, Korner-Bitensky N, et al. A new approach to retrain gait in stroke patients through body weight support and treadmill stimulation. *Stroke* 29: 1122-8,1998.

Waagfjord J, Levangie P, Certo CM. Effects of treadmill training on gait in a hemiparetic patient. *Phys Ther* 70:549-58,1990.

Wiley ME, Damiano DL. Lower-extremity strength profiles in spastic cerebral palsy. *Dev Med Child Neurol* 40:100-7,1998.

Zubrick SR, Taylor CL, Rice ML, et al. Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *J Speech Lang Hear Res* 50:1562-92,2007.