

کتاب جامع

بهداشت عمومی

فصل ۸ / گفتار ۱۱ / دکتر محسن جانقربانی

طراحی یک تجربه: کار آزمایشی‌های بالینی ۲

فهرست مطالب

۱۲۶۶	طراحی یک تجربه: کار آزمایشی بالینی ۲
۱۲۶۷	اهداف درس
۱۲۶۷	مقدمه
۱۲۶۷	پیگیری و رعایت منشور
۱۲۷۱	اندازه‌گیری پیامد
۱۲۷۱	پیامدهای بالینی در مقابل پیامدهای جایگزین
۱۲۷۲	ویژگی‌های آماری
۱۲۷۲	تعداد متغیرهای وابسته
۱۲۷۳	قضاوت درباره پیامدها
۱۲۷۳	اثرات مغایر
۱۲۷۴	تحلیل یافته‌ها
۱۲۷۴	تحلیل با قصد درمان
۱۲۷۶	پایش کارآزمایی‌های بالینی
۱۲۷۹	طرح‌های دیگر کارآزمایی کور تصادفی شده
۱۲۷۹	سایر طرح‌های تقسیم تصادفی شده
۱۲۸۰	طرح‌های بین گروهی تقسیم تصادفی نشده
۱۲۸۲	طرح‌های درون گروهی
۱۲۸۳	قصد اجرای یک کارآزمایی
۱۲۸۴	خلاصه
۱۲۸۴	منابع

طراحی یک تجربه: کار آزمایشی‌های بالینی ۲

Designing an Experiment: Clinical Trials

دکتر محسن جانقربانی
دانشگاه علوم پزشکی اصفهان

اهداف درس

پس از یادگیری این مطالب، فراگیرنده قادر خواهد بود:

- نحوه پیگیری و رعایت منشور یک کار آزمایشی بالینی را طراحی و تشریح نماید
- پیامد یافته‌های یک کار آزمایشی بالینی را اندازه‌گیری و تحلیل نماید
- نحوه پایش یافته‌های کار آزمایشی را طراحی و تشریح کند
- انواع روش‌های کار آزمایشی بالینی تصادفی کلاسیک را توضیح دهد.

مقدمه

در بخش گذشته، کارآزمایی تصادفی شده و کورسازی شده بحث شد و در خصوص چگونگی انتخاب شرکت کنندگان، اندازه‌گیری متغیرهای پایه، تقسیم تصادفی و مداخله، توضیحات لازم داده شد. در این بخش، نحوه به حداکثر رساندن پیگیری و رعایت منشور، اندازه‌گیری پیامد و تحلیل یافته‌ها بحث می‌شود. در کارآزمایی‌های بالینی، کاری روی شرکت کنندگان انجام می‌شود و این بخش نیاز به پایش یافته‌ها را طی کارآزمایی نشان می‌دهد. این بخش با مروری بر برخی از روش‌های دیگر کارآزمایی تصادفی شده کلاسیک پایان می‌یابد.

پیگیری و رعایت منشور

اگر تعداد قابل توجهی از شرکت کنندگان در مطالعه، مداخله تحت بررسی را دریافت نکنند، منشور را رعایت نکنند، یا در طی پیگیری گم شوند، قدرت یافته‌های کارآزمایی احتمالاً کم یا سوگرا می‌شود. راهکارهای به

حداکثر رساندن پیگیری و تمکین در جدول ۱ خلاصه شده‌اند.

تاثیر مداخله (و قدرت کارآزمایی)، به اندازه شرکت کنندگانی که آن را دریافت نکرده‌اند کاهش می‌یابد. پژوهشگر باید تلاش کند دارو یا مداخله رفتاری را انتخاب کند که به سادگی اعمال یا مصرف شود و به طور کامل قابل تحمل باشد. اگر یک مداخله رفتاری توسط شرکت کنندگان، نیاز به ساعت‌ها تمرین داشته باشد، احتمالاً رعایت آن ضعیف خواهد بود. داروهایی را که می‌توان روزانه یک دوز واحد آن را مصرف کرد، آسان‌تر به خاطر سپرده می‌شوند و در نتیجه ترجیح داده می‌شوند. منشور باید شامل شروطی باشد که رعایت دستورها را افزایش دهد، نظیر آموختن به شرکت کنندگان که قرص‌ها را به طور معمول صبح‌ها در وقت معینی مصرف کنند و به آن‌ها جعبه قرصی داده شود که روزهای هفته روی آن مشخص شده‌اند.

همچنین، لازم است بهترین روش سنجش تمکین از مداخله، نظیر استفاده از روش‌های گزارش شخصی، شمارش قرص‌ها و سطوح متابولیت‌های ادرار و سرم در نظر گرفته شود. این اطلاعات می‌تواند شرکت کنندگانی را که تمکین نمی‌کنند بشناسد، در نتیجه اگر در پایان مطالعه تفاوتی بین گروه‌ها نبود در توجیه این یافته به پژوهشگر کمک نماید.

پژوهشگر می‌تواند قبل از گرفتن رضایت، آنچه در مطالعه انجام می‌شود را شرح دهد و با برنامه ریزی مراجعه‌ها در زمانی راحت و با تعداد کافی کارمند برای جلوگیری از معطلی، با تلفن زدن به شرکت کننده یک روز قبل از هر مراجعه و با جبران هزینه‌های رفت و آمد و سایر هزینه‌های اضافی، رعایت مراجعه‌ها و سنجش‌های مطالعه را افزایش دهد.

عدم پیگیری شرکت کنندگان در کارآزمایی و اندازه‌گیری پیامد مورد نظر می‌تواند باعث یافته‌های سوگرا شود، روایی (اعتبار) یافته‌ها را کم کند و قدرت آماری را کاهش دهد. برای مثال، یک کارآزمایی افشانه بینی کلسی تونین برای کاهش خطر شکستگی‌های استئوپروزی گزارش نمود که درمان، خطر شکستگی را ۳۶٪ کاهش می‌دهد (۱). ولی حدود ۶۰٪ آن‌هایی که تقسیم تصادفی شدند در حین پیگیری گم شدند، و معلوم نشد آیا در این شرکت کنندگان شکستگی روی داده است یا خیر. از آنجا که تعداد کل شکستگی‌ها کم بود، حتی چند شکستگی در شرکت کنندگانی که طی پیگیری گم شدند می‌توانست یافته‌های کارآزمایی را تغییر دهد. این ابهام، اعتبار یافته‌های مطالعه را کم کرد (۲).

حتی اگر شرکت کنندگان منشور را نقض کنند یا مداخله کارآزمایی را قطع نمایند، باید پیگیری شوند، به طوری که بتوان از پیامدهای آن‌ها در تحلیل‌های با قصد درمان استفاده نمود. در بسیاری از کارآزمایی‌ها، شرکت کنندگانی که با شرکت در کارآزمایی دیگر، منشور را نقض می‌کنند، مداخله را قطع می‌کنند، یا اثرات جانبی را گزارش می‌کنند که باعث قطع پیگیری می‌شود، می‌تواند باعث یافته‌های سوگرا یا غیرقابل تفسیر شود. برای مثال، دارویی را در نظر بگیرد که باعث اثر جانبی علامتداری می‌شود و اغلب باعث قطع داروی مطالعه می‌گردد. اگر شرکت کنندگانی که داروی مطالعه را قطع کرده‌اند برای پیامد، پیگیری نشوند میزان رویدادها در گروه درمان فعال به سمت پایین سوگرا خواهد شد. اگر اثر جانبی با پیامد اصلی مرتبط باشد، این سوگرایی می‌تواند روی یافته‌های اصلی اثر جدی بگذارد.

جدول ۱ - راهکارهای به حداکثر رساندن پیگیری و رعایت منشور

مثال	اصل
لازم است قبل از تقسیم تصادفی دو یا چند بار ملاقات انجام داد آن‌هایی را که قبل از تقسیم تصادفی در دوره درگیرسازی تمکین نمی‌کنند حذف نمایید آن‌هایی را که احتمال دارد نقل مکان کنند یا تمکین نمی‌کنند حذف کنید	افرادی را انتخاب کنید که احتمال دارد مداخله و منشور را رعایت کنند
بجای دو قرص از یک قرص استفاده کنید ملاقات‌ها به اندازه کافی باشد تا تماس نزدیک برقرار شود، ولی آنقدر تکرار نشود که خسته کننده شود. برنامه‌های ملاقاتها را در شب یا آخر هفته تنظیم کنید یا اطلاعات را با پست الکترونیکی گردآوری کنید کارمند کافی داشته باشید تا از معطلی جلوگیری شود هزینه‌های رفت و آمد را تامین کنید با افراد رابطه شخصی برقرار کنید	مداخله را آسان سازید تماس‌های مطالعه را راحت و لذت بخش کنید
آزمون‌های غیرتهاجمی و آگاهی بخش را انتخاب کنید نتایج جالب آزمون را به شرکت کننده بدهید و مشاوره مناسب انجام دهید	سنجش‌های مطالعه را بی‌درد و جالب کنید
هرگز پیگیری افراد را بخاطر عدم رعایت منشور، رویدادهای مغایر یا اثرات جانبی قطع نکنید برای شرکت کنندگان کارت‌های تبریک تولد و تعطیلی بفرستید پیام‌های پست الکترونیکی یا خبرنامه بفرستید روی اهمیت علمی تمکین و پیگیری تاکید کنید	افراد را به ادامه شرکت در کارآزمایی تشویق کنید
افراد در تماس با فرد را ترغیب کنید و از خدمات ردیابی استفاده کنید	افرادی که حین پیگیری گم شده‌اند را بیابید

برخی راهکارهای پیگیری کامل مشابه آن‌هایی است که برای مطالعه‌های همگروهی گفته شد. در آغاز مطالعه، باید شرکت کنندگان را از اهمیت پیگیری آگاه نمود و پژوهشگران باید نام، نشانی و شماره تلفن یک یا دو نفر از آشنایان نزدیک را که همیشه می‌دانند شرکت کننده کجاست ثبت کنند. این کار علاوه بر اینکه توانایی پژوهشگر را برای ارزیابی وضع حیاتی زیاد می‌کند، ممکن است با تلفن کردن به کسانی که به طور مطلق شرکت در مطالعه را رد کرده‌اند به مقیاس‌های پیامد جانشین دسترسی پیدا کنند. در کارآزمایی مطالعه قلب و جایگزینی استروژن، ۸۹٪ زنان برای آخرین معاینه مراجعه نمودند، با ۸٪ دیگر آخرین تماس تلفنی برای تعیین پیامد برقرار شد و با تماس تلفنی، نامه‌های سفارشی، تماس با بستگان نزدیک و خدمات ردیابی اطلاعاتی در خصوص وضع حیاتی هر یک از شرکت کنندگان بدست آمد (۳).

طرح کارآزمایی باید طوری باشد که رعایت مداخله و تکمیل تمام مراجعه‌های پیگیر و سنجش‌ها را تا حد ممکن برای شرکت کنندگان آسان سازد. مراجعه‌های طولانی و پراضطراب، برخی از شرکت کنندگان را از مراجعه باز می‌دارد. بیشتر احتمال دارد شرکت کنندگان به مراجعه‌هایی پاسخ گویند که مستلزم آزمون‌های غیرتهاجمی است، نظیر توموگرافی رایانه‌ای پرتو الکترونی قلب، بجای آزمون‌های مهاجمی نظیر آنژیوگرافی عروق تاجی. گردآوری اطلاعات پیگیری بوسیله تلفن یا وسایل الکترونیکی ممکن است تمکین شرکت کنندگانی را که مراجعه‌ها را مشکل یافته‌اند بهبود بخشد. از طرف دیگر، اگر برخی پاداش‌های اجتماعی یا بین فردی برای شرکت کننده وجود نداشته باشد، شرکت کنندگان ممکن است علاقه خود را به کارآزمایی از دست بدهند. شرکت کنندگان ممکن است از مراجعه‌های ماهانه مطالعه خسته شوند و اگر تنها سالی یک بار مراجعه کنند ممکن است علاقه خود را از دست بدهند. با مثبت و لذت بخش کردن تجربه کارآزمایی برای شرکت کنندگان، پیگیری نیز اصلاح می‌شود: سنجش‌ها و روش‌های کارآزمایی را بدون درد و جالب طراحی کنید، آزمون‌هایی را انجام دهید که در غیر این صورت در اختیار نمی‌باشند، یافته‌های آزمون‌ها را به شرکت کنندگان بدهید (اگر یافته روی پیامد تاثیر نداشته باشد)، خبرنامه‌ها را بفرستید، یا پیام‌های قدرانی بفرستید، کارت‌های تبریک تولد و تعطیلی و هدیه‌های ارزان بفرستید و با کارمند مطالعه روابط شخصی قوی برقرار کنید.

دو جنبه طراحی که مخصوص کارآزمایی‌ها است و ممکن است تمکین و پیگیری را بهبود بخشند عبارتند از: مراجعه‌های غربالگری قبل از تقسیم تصادفی و دوره درگیرسازی. از شرکت کنندگان بخواهید قبل از تقسیم تصادفی در یک یا دو مراجعه غربالگری شرکت کنند تا شرکت کنندگانی که در می‌یابند نمی‌توانند چنین مراجعه‌هایی را کامل کنند حذف شوند. رمز کار در اینجا این است که سدهای به اندازه کافی بلندی برای ورود به مطالعه برقرار کنید تا کسانی که بعداً تمکین نخواهند کرد را حذف کند، ولی آنقدر بلند نباشد که شرکت کنندگانی را که تمکین رضایت بخشی خواهند داشت نیز حذف کند.

دوره درگیرسازی ممکن است برای افزایش نسبت شرکت کنندگان در مطالعه که مداخله و روش‌های پیگیری را رعایت می‌کنند طرح مفیدی باشد. در آغاز مطالعه، تمام شرکت کنندگان روی دارونما گذاشته می‌شوند. بعداً در زمانی خاص (معمولاً چند هفته)، آن‌هایی که مداخله را رعایت کرده‌اند به طور کور تقسیم تصادفی می‌شوند که دریافت دارونما را ادامه دهند یا شروع به دریافت داروی فعال کنند. در این شیوه، قبل از تقسیم تصادفی با حذف شرکت کنندگانی که تمکین نمی‌کنند می‌توان قدرت مطالعه را افزایش داد و برآورد بهتر اثرات کلی مداخله را ممکن ساخت. ولی معلوم نیست آیا قبل از تقسیم تصادفی درگیرسازی دارونما موثرتر است یا لزوم تکمیل یک یا چند مراجعه غربالگری.

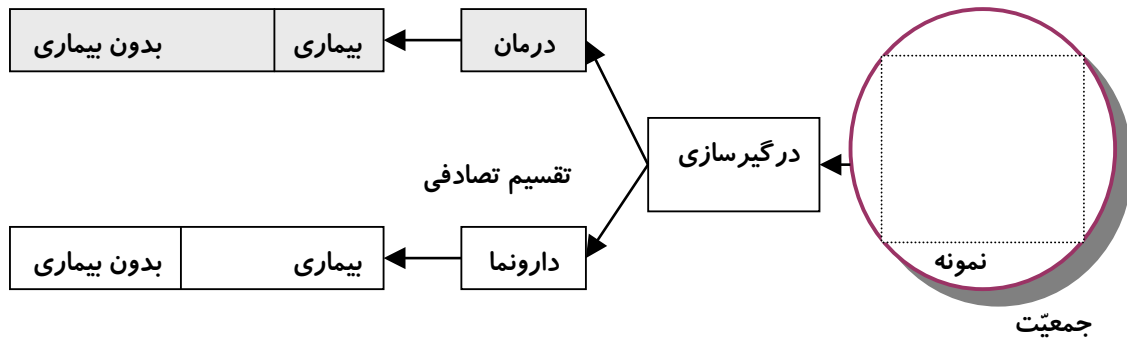
نوعی طرح درگیرسازی در شکل ۱ نشان داده شده است که در دوره درگیرسازی بجای دارونما از داروی فعال استفاده می‌شود. در طرح درگیرسازی داروی فعال علاوه بر افزایش تمکین در بین شرکت کنندگان، افرادی انتخاب می‌شوند که مداخله را تحمل می‌کنند و به آن پاسخ می‌دهند. از پاسخ به یک متغیر سرراهی (یعنی متغیری که بین مداخله و پیامد قرار دارد)، به عنوان معیار تقسیم تصادفی استفاده می‌شود. برای مثال، در یک کارآزمایی اثر داروی ضد آریتمی بر میرایی، پژوهشگر ممکن است تنها شرکت کنندگانی را تقسیم تصادفی کند که آریتمی آنها بدون اثرات جانبی به طور رضایت بخش فروکش کرده است (۴). این طرح با افزایش نسبت افراد گروه مداخله که

به مداخله پاسخ می‌دهند، قدرت را به حداکثر می‌رساند. همچنین با تقلید از تمایل پزشک به ادامه مصرف یک دارو، تنها وقتی که وی شواهدی دال بر اینکه کار می‌کند را می‌بیند، قابلیت تعمیم اصلاح می‌شود. وقتی افرادی که مداخله را تحمل نمی‌کنند یا پاسخ نمی‌دهند از یک کارآزمایی حذف شوند، نتایج به کسانی که حذف شده‌اند قابل تعمیم نمی‌باشد.

همچنین، ممکن است میزان اثرات جانبی در بین آنهایی که عضوگیری شده‌اند کمتر از آن میزان در بین تمام آنهایی باشد که روی مداخله گذاشته شده‌اند. در کارآزمایی اثر کارودیلول (Carvedilol) بر میرایی در مبتلایان به نارسایی احتقانی قلب از یک دوره درگیر سازی فعال دو هفته‌ای استفاده شد. در طی درگیرسازی، نارسایی احتقانی قلب ۱۷ نفر بدتر شد و ۷ نفر جان خود را از دست دادند (۵). این افراد در کارآزمایی تقسیم تصادفی نشدند، و این اثرات معکوس درمان دارویی به عنوان پیامد در نظر گرفته نشد.

اندازه‌گیری پیامد

اغلب در انتخاب مقیاس پیامد، پژوهشگر مجبور است بین مناسبت بالینی با قابلیت اجرا و هزینه تعادل برقرار کند.



شکل ۱ - در یک کارآزمایی تصادفی شده با یک دوره درگیرسازی برای آزمودن تمکین، پژوهشگر (الف) نمونه‌ای از جمعیت انتخاب می‌کند، (ب) متغیرهای پایه را اندازه می‌گیرد، (ج) شرکت کنندگان را تقسیم تصادفی می‌کند، (د) مداخله‌ها را اعمال می‌نماید، (ه) همگروه‌ها را پیگیری می‌کند، (و) متغیرهای وابسته را اندازه می‌گیرد

پیامدهای بالینی در مقابل پیامدهای جایگزین

مقیاس‌های از نظر بالینی مرتبط، نظیر مرگ، سکت قلبی، بستری شدن در بیمارستان و کیفیت زندگی، پرمعنی‌ترین پیامدهای کارآزمایی‌ها هستند. از نشانگرهای جایگزین برای خطر پیامد، نظیر کلسترول برای خطر بیماری عروق تاجی قلب، وقتی استفاده می‌شود که آزمودن یک داروی جدید در مراحل به نسبت اولیه است و

منابع برای یک مطالعه بزرگ با پیامدهای بالینی، خیلی محدود است. نشانگرهای جایگزین حداقل باید مقبولیت بیولوژیکی داشته باشند و با پیامد مورد نظر رابطه داشته باشند، برای مثال، معمولاً از چگالی استخوان به عنوان جایگزینی برای خطر شکستگی استفاده می‌شود، زیرا نشان داده شده که چگالی کم استخوان در استئوپروز با افزایش خطر شکستگی همراه است. ولی ثابت نشده، در آن‌هایی که درمان باعث تغییرات مطلوب در نشانگر جایگزین می‌شود، پیامدهای بالینی بهتری ایجاد خواهند کرد. متاسفانه، نمونه‌های بسیاری از کارآزمایی‌ها وجود دارند که از نشانگرهای جایگزین برای پیامدهای بالینی استفاده شده و یافته‌های گمراه کننده ایجاد نموده‌اند. برای مثال، چندین مطالعه نشان داده‌اند که آریتمی بطنی در مبتلایان به سکته قلبی، خطر مرگ را افزایش می‌دهد. کارآزمایی‌های بعدی نیز نشان داده‌اند که داروهای خاصی می‌توانند آریتمی بطنی را سرکوب کنند (پیامد جایگزین) متاسفانه، کارآزمایی سرکوب آریتمی قلبی (CAST [Cardiac Arrhythmia Suppression Trial]) نشان داد که هرچند این داروها دفعات آریتمی شدید را کم می‌کنند، میزان میرایی در بیماران درمان شده بالاتر بود (۴).

ویژگی‌های آماری

مقیاس پیامد باید طوری باشد که بتوان به طور صحیح و دقیق آن را ارزیابی کرد. وزن نوزاد در هنگام تولد مثالی از یک پیامد است که این معیارها را دارد، مثال دیگری که این معیارها را ندارد، ناتوانی مادرزادی در یادگیری است، یک متغیر رفتاری که انتهای نامشخص یک طیف را بیان می‌کند. متغیرهای وابسته پیوسته این مزیت را بر متغیرهای دو حالتی دارند که قدرت مطالعه را زیاد می‌کنند، در نتیجه به اندازه نمونه کوچک تری نیاز دارند.

اگر یک پیامد دو حالتی غیرقابل اجتناب باشد، قدرت مطالعه بیشتر به تعداد رویدادها بستگی دارد تا به تعداد کل شرکت کنندگان (۶). برای مثال، در کارآزمایی مطالعه قلب و جایگزینی استروژن، قدرت مطالعه با ۲۷۶۳ زن شرکت کننده در کارآزمایی تعیین نشد، بلکه با ۳۴۸ نفری که پیامد اصلی (سکته قلبی غیرکشنده یا مرگ از بیماری عروق تاجی قلب) را تجربه کردند، تعیین شد (۲). یک پیامد دو حالتی شایع تر، نظیر سندرم بیماری عروق تاجی حاد (سکته قلبی غیرکشنده، مرگ از بیماری عروق تاجی، و بستری برای آنژین ناپایدار) که در ۵۶۸ زن روی داد، را می‌توان با قدرت به مراتب بیشتری آزمود.

تعداد متغیرهای وابسته

اغلب داشتن چندین متغیر وابسته که جنبه‌های مختلف پدیده مورد نظر را اندازه می‌گیرند مطلوب است. در کارآزمایی مطالعه قلب و استروژن جایگزین، رویدادهای بیماری عروق تاجی قلب به عنوان نقطه فرجام اصلی انتخاب شد. تمام سکته‌های قلبی غیرکشنده، مرگ از بیماری عروق تاجی قلب، رگ دار شدن مجدد، بستری بخاطر آنژین ناپایدار و نارسایی احتقانی قلب، سکته مغزی و حمله گذرای کم خونی، رویدادهای ترومبوآمبولی وریدی و تمام علل مرگ بررسی شدند و قضاوت شد این متغیرها اثرات قلبی عروقی هورمون درمانی را بهتر توصیف می‌کنند (۳). ولی برای برنامه ریزی اندازه نمونه و دوره مطالعه و برای اجتناب از مسائل تفسیر آماری فرضیه‌های چندگانه یک نقطه فرجام اصلی واحد طراحی شد.

قضاوت درباره پیامدها

بیشتر پیامدهایی که افراد گزارش می‌کنند، نظیر سابقه سکته مغزی یا ترک سیگار صددرصد درست نیست. گزارش فردی پیامدها که برای کارآزمایی مهم است باید در صورت امکان تایید شود. معمولاً، معیارهای قضاوت در باره رویداد بیماری نظیر سکته مغزی، عبارتند از (الف) ایجاد معیارهای شفاف برای پیامد (نقص جدید نرولوژیک با ضایعه مشابه توموگرافی رایانه‌ای یا اسکن تصویربرداری مغناطیسی)، (ب) گردآوری مدارک بالینی لازم برای ارزیابی (خلاصه ترخیص و گزارش‌های پرتونگاری)، و (ج) هر مورد بالقوه توسط خبرگان بررسی و قضاوت شود آیا معیارهای تشخیصی در نظر گرفته شده‌اند یا خیر. آن‌هایی که اطلاعات را گردآوری می‌کنند و در باره موارد داوری می‌نمایند باید از تخصیص درمان بی اطلاع باشند.

اثرات مغایر

پژوهشگر باید مقیاس‌هایی را برای پیامد در نظر بگیرد که رویداد اثرات مضر را که ناشی از مداخله می‌باشند، تشخیص دهد. یکی از هدف‌های اصلی اغلب کارآزمایی‌های بالینی، حتی آن‌هایی که درمان‌های به ظاهر بی ضرر، نظیر یک برنامه آموزش بهداشت را می‌آزمایند این است که آیا اثرات مفید یک مداخله بر اثرات مضر آن می‌چربد یا خیر.

طیف اثرات مضر می‌تواند از علائم به نسبت جزئی نظیر بثورات پوستی یا رویدادهای شبه آنفلوآنزا، تا عوارض شدید و مرگبار باشد. پژوهشگر باید این مسئله را در نظر بگیرد که هم ماهیت فرجام و هم اندازه نمونه لازم برای تشخیص اثرات مضر ممکن است متفاوت از آن‌هایی باشد که برای تشخیص منافع است. متأسفانه، معمولاً تشخیص اثرات جانبی کمیاب، بدون توجه به اینکه کارآزمایی چقدر بزرگ است، غیرممکن است و تنها پس از آنکه آن مداخله به طور وسیع استفاده شد، می‌توان به آن پی برد.

اگر اثرات بالقوه مضر یک درمان جدید معلوم نباشد، پژوهشگر باید در مراحل اولیه آزمودن درمان جدید پرسش‌هایی وسیع و باز راجع به تمام انواع اثرات بالقوه مضر بنماید. در کارآزمایی‌های بزرگ، ارزیابی و کد گذاری تمام رویدادهای بالقوه مضر می‌تواند خیلی گران و وقت گیر، با نتایج مهم کم باشد. پژوهشگران باید ضمن ارزیابی کافی زیان‌های بالقوه مداخله راهکارهایی برای به حداقل رساندن این بار اضافی در نظر گیرند. برای مثال، رویدادهای شایع، نظیر عفونت دستگاه تنفسی، و اختلال‌های گوارشی را می‌توان در زیر مجموعه‌ای از شرکت کنندگان یا در یک زمان محدود بررسی کرد. رویدادها یا اثرات بالقوه مضر مهم که بخاطر پژوهش‌های قبلی مورد انتظارتند را می‌توان با پرسش‌های خاص به طور صحیح تر و موثرتر بررسی کرد. برای مثال، چونکه رابدومیولیزیس (Rhabdomyolysis) یک اثر جانبی گزارش شده درمان با استاتین‌ها است، باید در هر کارآزمایی یک استاتین جدید علائم و نشانه‌های میوزیت پرسیده شود. هرگاه از داده‌های یک کارآزمایی برای تصویب یک داروی جدید استفاده می‌شود، طرح کارآزمایی باید انتظارات قانونی برای گزارش رویدادهای مضر را برآورده سازد.

تحلیل یافته‌ها

معمولا تحلیل آماری فرضیه اصلی یک کارآزمایی بالینی، آسان است. اگر پیامد دو حالتی است، ساده ترین روش، مقایسه نسبت‌های گروه‌های مطالعه با استفاده از آزمون مجذور کای است. وقتی پیامد پیوسته است از آزمون t ، یا اگر توزیع پیامد طبیعی نباشد، از روش غیرپارامتری می‌توان استفاده کرد. در بیشتر کارآزمایی‌های بالینی دوره پیگیری هر شرکت کننده متفاوت است، که استفاده از روش‌های بقا را ضروری می‌کند. مدل‌های آماری پیچیده تر، نظیر تحلیل مخاطرات نسبی کاکس، می‌توانند آن را تکمیل نمایند و به طور همزمان توزیع‌های نامناسب شانس متغیرهای مخدوش کننده در آغاز مطالعه را تطبیق نمایند. جزئیات فنی این روش‌ها و نحوه و زمان استفاده از آن‌ها در کتاب‌های دیگر شرح داده شده است (۷).

دو موضوع مهم که در تحلیل یافته‌های کارآزمایی بالینی باید در نظر گرفته شوند عبارتند از برتری روش تحلیل با قصد درمان و نقش کمکی تحلیل‌های زیر - گروهی.

تحلیل با قصد درمان

برای تحلیل، پژوهشگر باید تصمیم بگیرد با "تقاطع‌ها" چه کند. یعنی شرکت کنندگانی که به گروه درمان فعال تخصیص یافته‌اند و درمان را مصرف نمی‌کنند یا آن را ادامه نمی‌دهند و آن‌هایی که به گروه شاهد تخصیص یافته‌اند و سرانجام درمان فعالی می‌گیرند. تحلیل با قصد درمان پیامدها را براساس تخصیص تصادفی هر مورد مقایسه می‌کند، بدون توجه به اینکه مداخله تخصیص یافته را گرفته است یا خیر. تحلیل‌های با قصد درمان ممکن است اثر کامل درمان را کمتر از حد برآورد نمایند، ولی در کارآزمایی‌های بالینی در مقابل علل مهم تر نتایج سوگرا در کارآزمایی‌های بالینی محافظت ایجاد می‌کنند.

راه دیگری برای تحلیل‌های با قصد درمان این است که تنها کسانی در تجزیه و تحلیل وارد شوند که مداخله را رعایت کرده‌اند. برای مثال برای انجام تحلیل‌های تمام مدت منشور تنها شرکت کنندگانی در هر دو گروه در نظر گرفته می‌شوند که بیش از ۸۰٪ داروی تحت مطالعه تخصیص یافته خود را دریافت کرده‌اند یا تنها آن‌هایی که قابل ارزیابی هستند (یعنی، نسبت معینی از داروی مطالعه را گرفته‌اند، نسبت معینی از مراجعه‌ها را انجام داده‌اند و تجاوزه‌های دیگری از منشور نداشته‌اند). این کار به نظر معقول می‌رسد، زیرا افراد تنها هنگامی که در واقع تحت مداخله قرار گیرند می‌توانند تحت تاثیر واقع شوند، ولی مسئله این است که تمکین کنندگان ممکن است با کسانی که تمکین نمی‌کنند به جهاتی که می‌تواند بر پیامد موثر باشد، متفاوت باشند. در کارآزمایی مداخله‌های استروژن - پروژستین پس از یائسگی، ۸۷۵ زن یائسه به طور تصادفی به چهار رژیم استروژن یا استروژن به اضافه پروژستین مختلف و دارونما تخصیص یافتند (۸). پس از ۳ سال ۳۰٪ زنانی که بدون مخالفت به گروه استروژن تخصیص یافته بودند بخاطر هیپرپلازی آندومتر، که منادی سرطان آندومتر است، درمان را قطع کردند. اگر این زنان از تحلیل تمام مدت منشور حذف شوند، ممکن است رابطه درمان استروژن و سرطان آندومتر دیده نشود.

عیب عمده روش تحلیل با قصد درمان این است که شرکت کنندگانی که انتخاب شده‌اند تا مداخله تخصیص یافته را نگیرند، کم و بیش در برآورد اثرات آن مداخله منظور می‌شوند. در نتیجه قطع یا تقاطع قابل

توجه بین درمان‌ها باعث می‌شود که تحلیل‌های با قصد درمان اندازه‌تاثیر درمان را کمتر از حد برآورد نمایند. به این دلیل، اغلب، یافته‌های کارآزمایی‌ها را با هر دو روش تحلیل با قصد درمان و تحلیل تمام مدت منشور ارزیابی می‌کنند. اگر نتایج هر دو تحلیل متفاوت باشد، معمولاً یافته‌های تحلیل‌های با قصد درمان برتر هستند، زیرا ارزش تقسیم تصادفی را حفظ می‌کنند، و بر خلاف تحلیل‌های تمام مدت منشور، تنها می‌توانند برآورد اثر را در جهت محافظه کارانه (به نفع فرضیه صفر) سوگرا نمایند. تنها وقتی که سنجش‌های پیگیر کامل باشند بدون توجه به اینکه شرکت کنندگان درمان را رعایت کرده‌اند یا خیر می‌توان یافته‌ها را به هر دو طریق تحلیل نمود.

تحلیل‌های زیر - گروهی به صورت مقایسه بین گروه‌های تقسیم تصادفی شده در زیر - مجموعه‌ای از همگروه کارآزمایی تعریف می‌شود. این تحلیل‌ها از شهرت دوگانه‌ای (هم خوب و هم بد) برخوردارند، زیرا به آسانی از آن‌ها سوء استفاده می‌شود و می‌توانند به نتیجه‌گیری‌های اشتباه منجر شوند. با این حال، در صورت دقت، می‌توانند اطلاعات کمی مفیدی فراهم نموده و استنتاج‌های ممکن از یک کارآزمایی بالینی را بسط دهند. برای حفظ ارزش تقسیم تصادفی، زیر - گروه‌ها باید بر اساس سنجش‌هایی که قبل از شروع درمان طبقه بندی شده‌اند تعریف شوند. برای مثال، یک کارآزمایی آلدورنات (Alendronate) برای پیشگیری از شکستگی‌های استئوپروزی به این نتیجه رسید که دارو خطر شکستگی را در بین زنان با چگالی استخوان کم ۱۴٪ کاهش می‌دهد. تحلیل‌های از قبل برنامه ریزی شده بر حسب زیر - گروه‌های چگالی استخوان که در آغاز مطالعه اندازه‌گیری شده بود نشان داد در زنانی که چگالی استخوان آن‌ها بیش از ۲/۵ انحراف معیار کمتر از طبیعی بود، درمان موثر بوده است (۳۶٪ کاهش خطر شکستگی) ($P < 0/01$ در مقابل، در زنان با چگالی استخوان بیشتر در آغاز مطالعه، درمان موثر نبود (۹). ذکر این نکته مهم است که ارزش تقسیم تصادفی در هر یک از زیر - گروه‌ها حفظ شده است. میزان شکستگی در زنان تقسیم تصادفی شده به آلدورنات با میزان در زنانی که به دارونما تقسیم تصادفی شده بودند در هر زیر - گروه مقایسه شد، آن‌هایی که چگالی استخوان کم داشتند (با سنجش‌هایی که قبل از تقسیم تصادفی انجام شده بود تعریف شد) و آن‌هایی که چگالی استخوان زیادتر داشتند با هم مقایسه شدند.

ولی تحلیل‌های زیر - گروهی به چند دلیل باعث ایجاد یافته‌های گمراه کننده می‌شوند. طبق تعریف، زیر - گروه‌ها از کل جمعیت کارآزمایی کوچکترند و ممکن است برای یافتن تفاوت‌های مهم قدرت کافی نداشته باشند. وقتی قدرت کافی برای یافتن یک تاثیر وجود ندارد، پژوهشگران باید از این ادعا که دارو در زیر - گروه "موثر نبوده" خودداری کنند. پژوهشگران اغلب یافته‌ها را در تعداد زیادی از زیر - گروه‌ها بررسی می‌کنند و احتمال یافتن شانس تاثیر متفاوتی از مداخله در یک زیر - گروه افزایش می‌یابد. به طور مطلوب، تحلیل‌های زیر - گروه‌های برنامه ریزی شده باید قبل از آغاز کارآزمایی تعریف شوند و تعداد زیر - گروه‌های تجزیه و تحلیل شده باید همراه با یافته‌های مطالعه گزارش شوند. در یک روش محافظه کارانه لازم است ادعاهای راجع به پاسخ‌های مختلف در زیر - گروه‌ها با شواهد آماری مبنی بر این که بین درمان و ویژگی زیر - گروه تعامل وجود دارد، حمایت شود. برای مثال، در مطالعه آلدورنات تعامل معنی داری بین چگالی استخوان در آغاز مطالعه و تاثیر درمان بر خطر شکستگی یافت شد ($P < 0/01$) و از این نتیجه‌گیری حمایت کرد که آلدورنات در زنان مبتلا به استئوپروز موثر

است، ولی در زنان با چگالی استخوان بیشتر موثر نیست. تحلیل‌های زیر - گروهی بر اساس عوامل پس از تقسیم تصادفی از ارزش تقسیم تصادفی حفاظت نمی‌کنند و اغلب یافته‌های گمراه‌کننده ایجاد می‌کنند. تحلیل تمام مدت منشور که به افرادی محدود می‌شود که درمان تقسیم تصادفی شده را رعایت می‌کنند مثالی از این نوع تحلیل زیر - گروهی است.

پایش کارآزمایی‌های بالینی

چرا یک کارآزمایی بالینی را پایش می‌کنیم؟ یک تفاوت مهم بین کارآزمایی‌های بالینی و مطالعه‌های مشاهده‌ای این است که در کارآزمایی بالینی عملی روی شرکت‌کنندگان انجام می‌شود. به دلایل اخلاقی، پژوهشگر باید مطمئن شود که شرکت‌کنندگان در معرض یک مداخله زیانبار قرار نمی‌گیرند، از یک مداخله مفید مضایقه نمی‌شود، یا اگر پرسش پژوهشی را نمی‌توان پاسخ داد کارآزمایی ادامه نمی‌یابد.

مبرم‌ترین دلیل پایش کارآزمایی‌های بالینی این است که مطمئن شویم مداخله به طور غیرمنتظره زیان‌آور نباشد. اگر زیان‌ها بر منافع بچربند، کارآزمایی باید متوقف شود. دوم، اگر مداخله موثرتر از آنی باشد که هنگام طراحی کارآزمایی برآورد شده بود، آنگاه منفعت را می‌توان زود در کارآزمایی مشاهده کرد. هرگاه منفعت معلومی ثابت شده باشد، ادامه کارآزمایی و تاخیر در ارائه مداخله به شرکت‌کنندگانی که روی دارونما هستند و افراد دیگری که می‌توانند از آن سود ببرند، ممکن است اخلاقی نباشد. سوم، اگر امکان پاسخ دادن به پرسش پژوهش وجود ندارد، ادامه شرکت افراد در یک کارآزمایی که به زمان و تلاش نیاز دارد و ممکن است باعث ناراحتی یا خطر شود، غیراخلاقی است. اگر یک کارآزمایی برای ۵ سال برنامه ریزی شده ولی پس از ۴ سال تفاوت اندکی در میزان رویدادهای پیامد در گروه‌های درمان شده و درمان نشده وجود داشته باشد، آنگاه "قدرت شرطی" (احتمال پاسخ به پرسش پژوهشی بانتایج حاصل) خیلی کوچک می‌شود و برای توقف کارآزمایی باید اقدام کرد. به علاوه، چون که کارآزمایی‌های بالینی پرهزینه هستند، توقف آن‌ها به محض یافتن پاسخ پرسش، هزینه‌ها را کم می‌کند.

قبل از خاتمه یک کارآزمایی خاص، ممکن است کارآزمایی‌های دیگر به پرسش پژوهشی پاسخ دهند. مطلوب آن است، که بیش از یک کارآزمایی شواهدی راجع به پرسش پژوهشی خاص فراهم کنند، ولی اگر چنین کارآزمایی شواهد قطعی در اختیار قرار گیرد، پژوهشگر باید توقف کارآزمایی را در نظر بگیرد. در برنامه ریزی هر کارآزمایی باید نحوه انجام پایش ضمنی منظور شود. قبل از شروع مطالعه باید رهنمودها و روش‌های پایش به طور مکتوب شرح داده شوند. نکاتی که در این رهنمودها در نظر گرفته می‌شوند در جدول ۲ ارائه شده‌اند.

همیشه توقف کارآزمایی باید با تصمیم به دقت سنجیده شده‌ای انجام شود، به طوری که مسئولیت‌های اخلاقی نسبت به شرکت‌کنندگان و پیشرفت علمی دانش را متعادل کند. هر گاه یک کارآزمایی زود خاتمه یابد، شانس بدست آوردن یافته‌های قطعی تر از بین می‌رود. اغلب تصمیم‌گیری پیچیده است و باید خطرهای بالقوه برای شرکت‌کنندگان در مقابل منافع احتمالی سنجیده شود. بنابراین، لازم است کمیته‌ای که کارآزمایی را پایش می‌کند شامل پزشکان، طرفداران شرکت‌کنندگان، آمارشناسان و خبرگان در انجام کارآزمایی‌ها باشد. به طور طبیعی این خبرگان از خارج از کارآزمایی هستند و در کارآزمایی، درگیر نمی‌باشند، در نتیجه منافع شخصی یا مالی در ادامه آن ندارد.

جدول ۲ - پایش یک کارآزمایی بالینی

عناصر پایش
عضوگیری
تمکین
تقسیم تصادفی
کورسازی
پیگیری
متغیرهای مهم
پیامدها
اثرات زیان آور
مخدوش کنندگان بالقوه
چه کسی پایش خواهد کرد
اگر کارآزمایی کوچک با خطر جزئی، پژوهشگران کارآزمایی در غیر این صورت هیئت مستقل پایش داده‌ها
تغییر منشور در اثر پایش
ختم کارآزمایی
اصلاح کارآزمایی
توقف شاخه‌ای از کارآزمایی
افزودن سنجش‌های جدید لازم برای پایش ایمنی
ادامه ندادن شرکت کنندگان پر خطر
بسط به موقع کارآزمایی
افزایش نمونه کارآزمایی
هرچند وقت یک بار پایش انجام می‌شود
به اندازه کافی تا به هدف‌های پایش برسیم
تنها تا وقتی که داده‌های جدید قابل توجهی وجود دارد
روش‌های آماری برای پایش

در توقف یک کارآزمایی، آزمون‌های معنی داری آماری اطلاعات مهم، ولی غیر قاطع فراهم می‌کنند. ثبات روندهای زمانی باید ارزیابی شوند، ثبات تاثیرات بر پیامدهای مربوطه باید ارزیابی شوند، و تاثیر توقف زود مطالعه بر اعتبار یافته‌ها باید به دقت در نظر گرفته شود (مثال ۱).

برای پایش یافته‌های ضمنی یک کارآزمایی روش‌های آماری زیادی وجود دارد. تحلیل مکرر یافته‌های یک کارآزمایی نوعی آزمودن چندگانه است و در نتیجه احتمال ارتکاب خطای نوع اول را افزایش می‌دهد. برای

مثال ۱ - کارآزمایی‌هایی که زود خاتمه یافته‌اند

مطالعه کانادایی ضد انعقاد فیبرینولیتیک دهلیزی (۱۰): فیبرینولیتیک دهلیزی یک عامل خطر سکتته مغزی و رویدادهای آمبولی است. مطالعه کانادایی ضد انعقاد فیبرینولیتیک دهلیزی یک کارآزمایی دو سوکور، تقسیم تصادفی شده، با شاهد دار و نما، برای ارزیابی تاثیر وارفارین (Warfarin) در کاهش میزان سکتته مغزی، آمبولیسم سیستمیک، یا خونریزی داخل مغزی یا کشنده در مبتلایان به فیبرینولیتیک غیر روماتیسمی بود. مطالعه طراحی شده بود تا ۶۶۰ بیمار را در نظر بگیرد و آن‌ها را برای ۳/۵ سال تحت درمان پیگیری کند. در طی کارآزمایی (پس از اینکه ۳۸۳ بیمار تقسیم تصادفی شدند و برای میانگین ۲ سال پیگیری شدند)، یافته‌های دوکارآزمایی تقسیم تصادفی شده دیگر گزارش گردید و کاهش معنی داری در رویدادهای آمبولی و میزان کمتری از رویدادهای خون ریزی دهنده عمده در مبتلایان به فیبرینولیتیک دهلیزی که با وارفارین درمان شده بودند را نشان داد. کمیته رهبری مطالعه کانادایی ضد انعقاد فیبرینولیتیک تصمیم گرفت که شواهد دال بر سود وارفارین به اندازه کافی می‌باشد تا بدون بررسی مقدماتی داده‌ها کارآزمایی متوقف شود.

کارآزمایی سرکوب آریتمی قلبی (۴): رویداد دیپولاریزاسیون نارس بطنی در زنده ماندگان از سکتته قلبی یک عامل خطر مرگ ناگهانی است. کارآزمایی سرکوب آریتمی قلبی تاثیر درمان ضد آریتمی (اینکائید [Encainide])، فلکائینید [Flecainide]، یا موری سیزین [Moricizine]) را در مبتلایان به آریتمی بطنی بدون علامت یا با علائم خفیف پس از سکتته قلبی بر خطر مرگ ناگهانی ارزیابی کرد. پس از متوسط ۱۰ ماه پیگیری، افرادی که با داروی فعال درمان شده بودند در مقایسه با آن‌هایی که به دارونما تخصیص یافته بودند میرایی کلی بیشتری (۷/۵٪ در مقابل ۳٪) و میزان‌های مرگ از آریتمی بیشتری (۴/۵٪ در مقابل ۱/۵٪) داشتند. برنامه ریزی شده بود که کارآزمایی تا ۵ سال ادامه یابد، ولی پس از ۱۸ ماه متوقف شد.

پروژه داروی کرورن (۱۱، ۱۲): پروژه داروی کرورن یک کارآزمایی تقسیم تصادفی، و کورسازی شده برای تعیین این بود که آیا ۵ مداخله مختلف پایین آورنده کلسترول (استروژن کونژوگه، ۵ میلی گرم در روز، استروژن ۲/۵ میلی گرم در روز، کلوفیبرات ۱/۸ گرم در روز، دکستروتیروکسین ۶ میلی گرم در روز، نیاسین ۳ گرم در روز) میزان میرایی ۵ ساله را کاهش می‌دهد یا خیر. پروژه داروی کرورن ۸۳۴۱ مرد مبتلا به سکتته قلبی را که حداقل برای ۵ سال پیگیری می‌شدند در نظر گرفت. پس از متوسط ۱۸ ماه پیگیری، بخاطر سکتته قلبی غیرکشنده اضافی (۶/۲٪ درمقایسه با ۳/۲٪) و رویدادهای ترومبوآمبولی وریدی (۳/۵٪ درمقایسه با ۱/۵٪) شاخه دوز بالای استروژن متوقف شد. این تصمیم گیری با این حقیقت تقویت شد که استروژن بالا با آتروفی بیضه‌ای، ژینکوماستی، حساسیت پستان‌ها، و کاهش میل جنسی نیز همراه است. در همان زمان، دکستروتیروکسین در زیر - گروه مردانی که در نوار قلبی پایه ضربان بطنی نارس مکرر نشان داده بودند متوقف شد، زیرا میزان مرگ در این زیر - گروه ۳۸/۵٪ در مقایسه با ۱۱/۵٪ در زیر - گروه مشابهی که دارونما دریافت می‌کردند بود. به زودی پس از آن درمان دکستروتیروکسین بخاطر میزان میرایی اضافی در گروه تحت درمان در تمام افراد متوقف شد. دو سال قبل از پایان برنامه ریزی شده مطالعه، شاخه ۲/۵ میلیگرم استروژن در روز نیز بخاطر عدم وجود شواهدی دال بر تاثیر مفید و افزایش خطر رویداد ترومبوآمبولی وریدی در مردان تحت درمان متوقف شد.

مطالعه سلامت پزشکان (۱۳): مطالعه سلامت پزشکان یک کارآزمایی تصادفی شده تاثیرآسپیرین (۳۲۵ میلی گرم یک روز در میان) بر میرایی از بیماری قلبی - عروقی بود. کارآزمایی پس از ۴/۸ سال از ۸ سال پیگیری برنامه ریزی شده متوقف گردید. در گروه تحت درمان کاهش معنی داری از سکتته قلبی وجود داشت (خطر نسبی برای سکتته قلبی غیرکشنده = ۰/۵۶). ولی تعداد مرگ‌های ناشی از بیماری قلبی عروقی در هر دو گروه مساوی بود. میزان مرگ از بیماری قلبی عروقی که در مطالعه مشاهده شد خیلی کمتر از حد انتظار بود (۸۸ پس از ۴/۵ سال پیگیری در مقابل ۷۳۳ مورد انتظار) و کارآزمایی بخاطر قدرت شرطی برای کشف اثر مطلوب درمان آسپیرین بر مرگ از بیماری قلبی عروقی که به سطح بسیار پایینی افت کرد، متوقف شد.

مثال، اگر برای هر آزمون از $\alpha = 0/05$ استفاده شود و در طی کارآزمایی چهار مرتبه و در پایان یک مرتبه یافته‌ها تحلیل شوند، احتمال ارتکاب خطای نوع اول به حد ۱۴٪ افزایش می‌یابد (۱۴). برای حل این مسئله، معمولاً روش‌های آماری برای پایش ضمنی، α را برای هر آزمون (α_i) کاهش می‌دهد، به طوری که کل آلفا مساوی $0/05$ شود. برای تصمیم‌گیری درباره "نحوه مصرف α " روش‌های متعددی وجود دارد.

طرح‌های دیگر کارآزمایی کور تصادفی شده

سایر طرح‌های تقسیم تصادفی شده

انواعی از کارآزمایی‌های تقسیم تصادفی شده کلاسیک وجود دارند که در صورت مساعد بودن شرایط مفیدند.

هدف‌های طرح فاکتوریل پاسخ دادن به دو موضوع پژوهشی مجزا با یک همگروه واحد از شرکت کنندگان است (شکل ۲). یک مثال خوب مطالعه سلامت پزشکان است که برای آزمودن اثر آسپیرین بر سکت قلبی و اثر بتاکاروتن بر سرطان، طراحی شده بود (۱۵). شرکت کنندگان به طور تصادفی به چهار گروه تقسیم شدند، ولی هر یک از دو فرضیه را می‌توان با مقایسه دو نیمه از همگروه تحت مطالعه آزمود. نخست، تمام کسانی که روی آسپیرین بودند با تمام کسانی که روی دارونمای آسپیرین بودند مقایسه شدند (بدون توجه به این حقیقت که نصف هر یک از این گروه‌ها بتاکاروتن دریافت می‌کردند)، سپس تمام کسانی که روی بتاکاروتن بودند با تمام کسانی که روی دارونمای بتاکاروتن بودند مقایسه شدند (اکنون بدون توجه به این حقیقت که نصف از هر یک از این گروه‌ها آسپیرین می‌گرفتند). پژوهشگر دو کارآزمایی کامل را به قیمت یک کارآزمایی تمام می‌کند.

طرح فاکتوریل فوق العاده موثر است. محدودیت عمده آن احتمال تعامل بین درمان‌ها و پیامدها است. در مثال فوق، هر تاثیر بتاکاروتن بر سکت قلبی پیامد نصف شرکت کنندگانی را که آسپیرین دریافت می‌کنند تغییر خواهد داد، قدرت را کم و تفسیر را مشکل خواهد نمود.

طرح‌های فاکتوریل در واقع می‌توانند برای مطالعه چنین تعامل‌هایی بکارروند، ولی این کارآزمایی‌ها با نمونه‌های با حجم زیاد نیاز دارند، پیچیده تر هستند و اجرای آن‌ها مشکل تر است و تفسیر نتایج مشکل می‌باشد. بهترین نقش طرح‌های فاکتوریل در پژوهش بالینی مطالعه دو موضوع پژوهشی به نسبت مستقل است.

تقسیم تصادفی جفت‌های جور شده یک راهکار برای متعادل نمودن متغیرهای مخدوش کننده اولیه است که مستلزم انتخاب جفت‌های افرادی است که برای عوامل مهمی نظیر سن و جنس جور شده‌اند، سپس به طور تصادفی مشخص می‌کنیم کدام عضو هر جفت می‌تواند به کدام گروه مطالعه تخصیص یابد. یک جنبه بویژه جالب این طرح این است که برای نشان دادن تفاوت اثرات درمان و شاهد در دو قسمت از یک شخصی در یک زمان استفاده شود. برای مثال، در مطالعه رتینوپاتی دیابتی یک چشم هر شرکت کننده به طور تصادفی به درمان با فتوکواگولاسیون اختصاص یافت در حالی که چشم دیگر وی به عنوان شاهد عمل می‌کرد (۱۶).

تقسیم تصادفی گروهی یا خوشه‌ای مستلزم این است که پژوهشگر بجای تقسیم تصادفی افراد، دسته‌ها

یا گروه‌های افراد را که به طور طبیعی تشکیل می‌شوند تقسیم تصادفی کند. یک مثال خوب کارآزمایی است که بازیکنان ۱۲۰ تیم بیسبال دانشکده را انتخاب کرد، و نصف تیم را به طور تصادفی به یک مداخله که ترک جویدن توتون را تشویق می‌کرد تخصیص داد و میزان به طور معنی دار کمتری از جویدن توتون را در بازیکنان تیمی که مداخله را دریافت کرده بود مشاهده کرد (۱۷). اعمال مداخله در گروه‌های افراد ممکن است از مبادرت به درمان در یک فرد قابل اجراتر و هزینه - تاثیر بیشتری داشته باشد، و ممکن است موضوع‌های پژوهشی در خصوص اثرات برنامه‌های بهداشتی را در جمعیت بهتر مشخص نماید. اعمال برخی مداخله‌ها، نظیر رژیم غذایی کم چربی، در یک عضو خانواده بدون اعمال آن در سایر افراد خانواده مشکل است. همین طور، شرکت کنندگانی که یک مداخله قابل انتقال را دریافت می‌کنند ممکن است این توصیه را با آشنایانی که به گروه شاهد تخصیص یافته‌اند در میان بگذارند. برای مثال، پزشکی که در یک مجتمع پزشکی کار می‌کند و به طور تصادفی به یک مداخله آموزشی تخصیص یافته است احتمال زیادی دارد این مداخله را با همکارانش در میان بگذارد. یک عیب تقسیم تصادفی گروهی این واقعیت است که برآورد اندازه نمونه و تجزیه و تحلیل‌ها پیچیده تر است (۱۸).

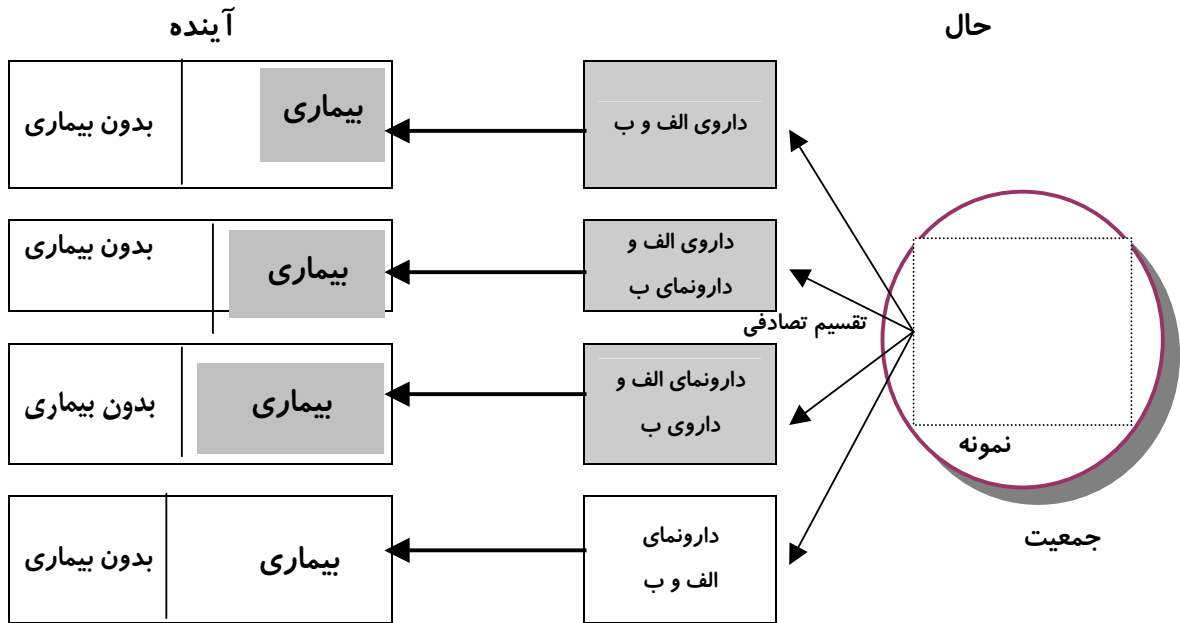
طرح‌های بین گروهی تقسیم تصادفی نشده

کارآزمایی‌هایی که گروه‌های تقسیم تصادفی نشده را باهم مقایسه می‌کنند در کنترل تاثیر متغیرهای مخدوش کننده خیلی کمتر از کارآزمایی‌های کور تصادفی شده رضایت بخش هستند. متغیرهایی را که در آغاز مطالعه اندازه‌گیری شده و به طور نامساوی بین دو گروه مطالعه توزیع گردیده‌اند می‌توان با روش‌های تحلیلی تطبیق کرد، ولی این راهکار مسئله متغیرهای مخدوش کننده اندازه‌گیری نشده را در نظر نمی‌گیرد. چالمرز یافته‌های مطالعه‌های تقسیم تصادفی شده و نشده موضوع پژوهشی مشابهی را مرور کرد (۱۹)، حتی پس از تطبیق آماری برای اختلاف‌های متغیرهای آغاز مطالعه، مزایای ظاهری مداخله در مطالعه‌های تقسیم تصادفی نشده خیلی زیاده‌تر بود. این مطالعه و تحلیل‌های دیگر (۲۰) نشان می‌دهند که در مطالعه‌های بالینی تقسیم تصادفی نشده مسئله مخدوش شدن می‌تواند جدی باشد و نمی‌توان با تطبیق آماری اثر آن‌ها را کاملاً از بین برد.

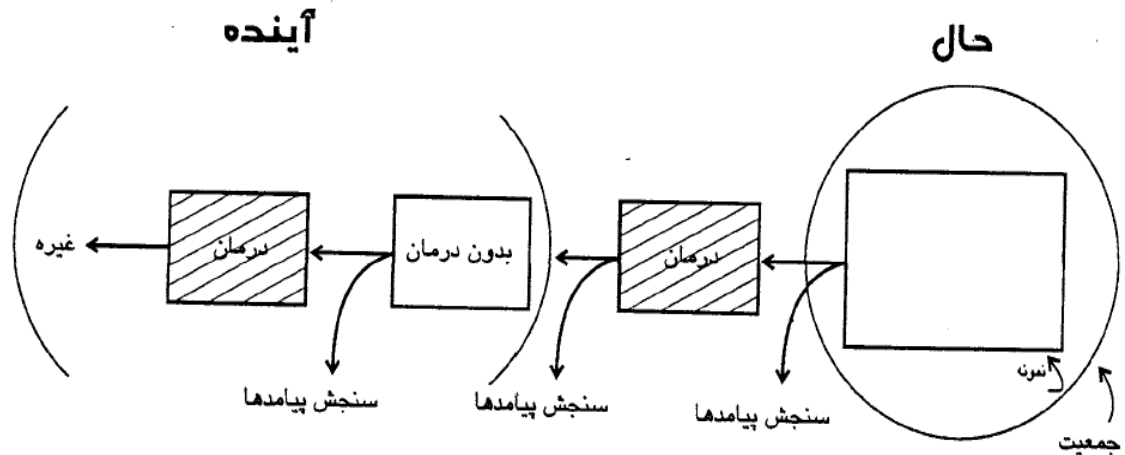
گاهی افراد با یک مکانیسم شبه تصادفی به گروه‌های مطالعه تخصیص می‌یابند. برای مثال، افراد یک در میان (یا با شماره پرونده زوج) به گروه درمان اختصاص می‌یابند. گاهی چنین طرح‌هایی مزایای تدارکاتی دارند، ولی قابلیت پیشگویی تخصیص گروه مطالعه به پژوهشگر این امکان را می‌دهد تا با دستکاری ترتیب تقدم و تاخر یا شایستگی انتخاب افراد جدید در مداخله اعمال نفوذ کند.

گاهی، پژوهشگر بر طبق ضوابط بالینی معینی افراد را به گروه‌های مطالعه اختصاص می‌دهد. برای مثال، بیماران دیابتی برحسب اینکه مایل باشند چهار تزریق در روز را قبول کنند ممکن است برای دریافت چهار مرتبه انسولین در روز یا انسولین با اثر طولانی یک مرتبه در روز تخصیص یابند.

مسئله این است کسانی که مایل هستند چهار تزریق انجام دهند ممکن است توصیه‌های بهداشتی دیگر را نیز بیشتر رعایت کنند و این می‌تواند علت هر اختلافی باشد که در پیامدهای دو برنامه درمانی مشاهده می‌شود. گاهی طرح‌های تقسیم تصادفی نشده بر اساس این اعتقاد غلط که اخلاقی تر هستند انتخاب می‌شوند. در واقع، تنها مطالعه‌هایی اخلاقی هستند که آنقدر خوب طراحی شده باشند تا احتمال معقول پاسخ صحیح دادن



شکل ۲ - در یک کارآزمایی فاکتوریل تقسیم تصادفی شده پژوهشگر (الف) نمونه‌ای از جمعیت انتخاب می‌کند، (ب) متغیرهای پایه را اندازه می‌گیرد، (ج) به طور تصادفی دو داروی فعال و شاهد‌های آن‌ها را به چهار گروه تخصیص می‌دهد، (د) مداخله را اعمال می‌کند، (ه) همگروه‌ها را پیگیری می‌کند، (و) متغیرهای وابسته را اندازه می‌گیرد.



شکل ۳ - در یک مطالعه مجموعه زمانی، پژوهشگر (الف) نمونه‌ای از جمعیت انتخاب می‌کند، (ب) متغیرهای پایه و پیامد را اندازه می‌گیرد، (ج) مداخله را در تمام همگروه اعمال می‌کند، (د) همگروه را پیگیری می‌کند، (ه) دوباره متغیرهای وابسته را اندازه می‌گیرد، (و) مداخله را حذف می‌کند و دوباره متغیرهای وابسته را اندازه می‌گیرد (اختیاری)

به موضوع پژوهشی را داشته باشند و طرح‌های تقسیم تصادفی شده بیش از طرح‌های تقسیم تصادفی نشده احتمال دارد به نتیجه قطعی منجر شوند. به علاوه مبنای اخلاقی هر کارآزمایی این ابهام است که آیا مداخله مفید یا مضر است، ابهامی به نام تعادل (Equipoise) که اگر قرار است اصلاً کارآزمایی انجام شود باید وجود داشته باشد.

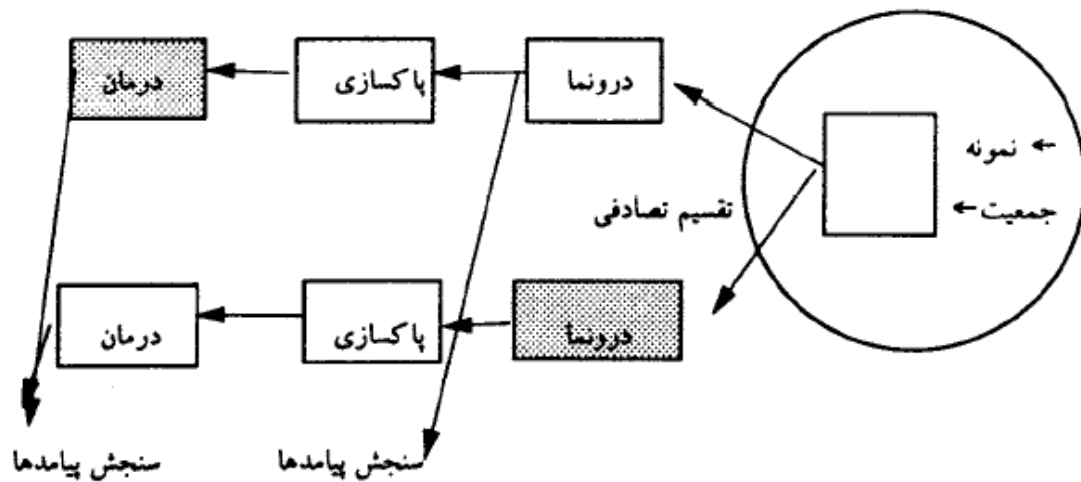
طرح‌های درون گروهی

طرح‌هایی که شامل تقسیم تصادفی نباشند، می‌توانند برای برخی از انواع پرسش‌ها انتخاب مفیدی باشند (شکل ۳). در یک طرح مجموعه زمانی، برای ارزیابی اثر یک درمان هر شرکت کننده به عنوان شاهد خودش عمل می‌کند. این بدین معنی است که نه تنها ویژگی‌های ذاتی نظیر سن، جنس و عوامل ارثی متعادل شده‌اند (نظیر مطالعه‌های بین گروهی) بلکه به عنوان متغیرهای مخدوش کننده در واقع حذف شده‌اند.

عیب عمده طرح‌های درون گروهی نبودن یک گروه شاهد همزمان است. تاثیر ظاهری مداخله ممکن است ناشی از اثرات آموزشی (شرکت کنندگان آزمون‌های عملکرد ادراکی را در پیگیری بهتر انجام می‌دهند، زیرا از آزمون آغاز مطالعه آموخته‌اند)، برگشت به میانگین (افرادى که به علت فشارخون بالا در آغاز مطالعه انتخاب شده بودند به سادگی بخاطر بی ثباتی تصادفی در فشار خون در پیگیری معلوم شد فشار خون پایین تری دارند)، یا روندهای زمانی (فراوانی عفونت‌های تنفسی در پیگیری کمتر بود، چون که کارآزمایی در فصل آنفلوآنزا آغاز شده بود) باشد. گاهی طرح‌های درون گروهی از یک راهکار شروع و توقف مکرر درمان استفاده می‌کنند. اگر شروع و توقف مکرر روش‌های مداخله الگوی مشابهی در پیامد ایجاد کند، از این موضوع که این تغییرات ناشی از درمان است به شدت حمایت می‌شود. این روش تنها وقتی مفید است که متغیر وابسته نسبت به مداخله به سرعت و قابل برگشت پاسخ دهد (برای مثال تاثیر مصرف الکل بر سطح لیپوپروتئین کلسترول پرچگالی).

طرح ضربدری هم ویژگی‌های طرح‌های بین گروهی و هم طرح‌های درون گروهی را دارد (شکل ۴). نصف شرکت کنندگان به طور تصادفی دوره شاهد را شروع می‌کنند و سپس به داروی فعال بر می‌گردند، در حالی که نصف دیگر به عکس عمل می‌کنند. این روش (یا مربع لاتین برای بیش از دو گروه درمانی) تحلیل‌های بین گروهی و نیز درون گروهی را مقدور می‌سازد. مزایای این طرح قابل توجه است: احتمال مخدوش شدن را به حداقل می‌رساند، زیرا هم شرکت کننده به عنوان شاهد خودش عمل می‌کند و به طور قابل توجهی قدرت کارآزمایی را زیاد می‌کند به طوری که شرکت کنندگان کمتری نیاز است. با این حال، معایب آن نیز قابل توجه است: دوره مطالعه را دو برابر می‌کند و پیچیدگی تجزیه و تحلیل و تفسیر ناشی از مسئله اثرات انتقالی به آن افزوده می‌شود. اثر انتقالی تاثیر باقیمانده مداخله بر پیامد در طی دوره‌ای است که مداخله قطع شده است.

برای مثال، پس از یک دوره درمان با داروی مدر (دیورتیک) فشارخون ممکن است ماه‌ها به سطح اولیه باز نگردد. برای کاهش اثر انتقالی، پژوهشگر می‌تواند یک دوره "پاکسازی" بدون درمان را به این امید وارد کند که متغیر وابسته قبل از شروع مداخله بعدی به حالت طبیعی برگردد، ولی دانستن اینکه آیا تمام اثرات انتقالی از بین رفته‌اند مشکل است. به طور کلی، مطالعه‌های ضربدری تنها وقتی انتخاب مناسبی هستند که تعداد افراد تحت مطالعه محدود باشد و باور کنیم اثرات انتقالی مسئله‌ای نیست.



شکل ۴ - طرح مجموعه زمانی

قصد اجرای یک کارآزمایی

به طور کلی در صورت امکان به پرسش‌های پژوهشی باید با کارآزمایی‌های تقسیم تصادفی شده پاسخ داده شود. مزیت عمده یک کارآزمایی تقسیم تصادفی شده استعداد آن برای کنترل تاثیر متغیرهای مخدوش کننده است و در نتیجه پاسخ‌های قطعی تر فراهم می‌کند. برای برخی پرسش‌های پژوهشی، یک کارآزمایی ممکن است سریع تر و کم هزینه تر از مطالعه‌های مشاهده‌ای باشد، بویژه اگر متغیر وابسته پیوسته باشد و به سرعت به مداخله پاسخ دهد. برای مثال، نشان دادن رابطه بین چربی رژیم غذایی و کلسترول سرم در یک مطالعه مشاهده‌ای مشکل است (بخاطر خطاهای سنجش متغیر رژیم غذایی)، ولی انجام آن در یک کارآزمایی به نسبت آسان است. برای برخی پرسش‌های پژوهشی جهت کنترل مخدوش کنندگان و اطمینان از اینکه منفعت آن بر خطرش می‌چربد بوضوح یک کارآزمایی لازم است. برای مثال، مطالعه‌های مشاهده‌ای مدام نشان داده‌اند افرادی که بتا - کاروتن مصرف می‌کنند خطر سرطان کمتری دارند، ولی چهار کارآزمایی بزرگ این منفعت را نیافتند (۲۱)، یافته‌های مطالعه‌های مشاهده‌ای ممکن است ناشی از مخدوش شدن باشد، زیرا افرادی که ویتامین‌ها را مصرف می‌کنند ممکن است از آن‌هایی که مصرف نمی‌کنند بیشتر نگران سلامتی خود باشند.

به هر حال، کارآزمایی‌ها معمولاً وقت گیر و گران هستند و اغلب شرکت کنندگان را در معرض خطر یا ناراحتی قرار می‌دهند. بنابراین، تا وقتی که به اندازه کافی راجع به مداخله بدانیم که پیشنهاد کند یک کارآزمایی تعیین کننده مقدور است، کارآزمایی‌ها نباید انجام شوند. چنین اطلاعاتی شامل تعریف دقیق مداخله (درمان، مشاوره، روش جراحی یا مقدار دارو، دوره و طریق مصرف)، سود احتمالی مداخله (برای برآورد اندازه نمونه و دوره کارآزمایی) و اثرات زیان آور احتمالی مداخله (برای حفظ ایمنی کافی شرکت کنندگان) می‌باشند. هر گاه بخاطر فقدان تقسیم تصادفی، کورسازی، یا تعداد کافی شرکت کنندگان پاسخ قطعی بعید باشد نباید یک کارآزمایی بالینی انجام داد.

خلاصه

۱ - اگر تعداد قابل توجهی از شرکت کنندگان در مطالعه مداخله تحت بررسی را دریافت نکنند منشور مطالعه را رعایت نکنند، یا در حین پیگیری گم شوند، یافته‌های کارآزمایی احتمالاً کمتر از حد قدرت دارند، سوگرا هستند، یا غیرقابل تفسیرند.

۲ - سنجش‌های از نظر بالینی مرتبط، نظیر مرگ، سکنه قلبی، بستری در بیمارستان و کیفیت زندگی، پرمعنی‌ترین پیامدهای کارآزمایی‌ها هستند. پژوهشگر باید تا حد ممکن، سنجش‌های پیامدی را در نظر بگیرد که رویداد اثرات زیان آور که ممکن است ناشی از مداخله باشد، را پیدا کند.

۳ - تحلیل‌های با قصد درمان روش اصلی است که مزیت کنترل مخدوش کنندگان را به وسیله تقسیم تصادفی دارد. تحلیل‌های تمام منشور، روش ثانویه‌ای است که برآوردی از اندازه تاثیر را در افراد رعایت کننده به دست می‌دهد، باید با احتیاط تفسیر شوند.

۴ - با دقت کافی، تحلیل‌های زیر - گروهی می‌توانند اطلاعات کمکی مفیدی فراهم نمایند و استنتاج حاصل از یک کارآزمایی را بسط دهند. برای حفظ ارزش تقسیم تصادفی، زیر - گروه‌ها باید بر اساس سنجش‌هایی تعریف شوند که قبل از شروع درمان انجام شده‌اند، و تحلیل‌ها باید پیامدهای بین زیر - مجموعه‌های تخصیص تصادفی شده گروه‌های مطالعه را مقایسه کنند.

۵ - تفاوت مهم بین کارآزمایی‌های بالینی و مطالعه‌های مشاهده‌ای این است که در یک کارآزمایی بالینی، عملی روی شرکت کنندگان انجام می‌شود. پایش ضمنی یک کارآزمایی باید اطمینان دهد که شرکت کنندگان با مداخله زیان آور مواجه نشده‌اند، مداخله مفید دریغ نشده، یا در صورتی که نتوان پرسش پژوهشی را پاسخ داد کارآزمایی ادامه نیابد.

۶ - چندین نوع طرح کارآزمایی تقسیم تصادفی شده وجود دارند که می‌توانند تحت شرایط مساعد کارآیی را به طور قابل توجهی افزایش دهند.

الف) طرح فاکتوریل امکان اجرای دو کارآزمایی مستقل را به قیمت یک کارآزمایی فراهم می‌کند.

ب) تقسیم تصادفی جفت‌های جور شده متغیرهای مخدوش کننده آغاز مطالعه را متعادل می‌کند.

ج) تقسیم تصادفی گروهی مطالعه موثر دسته‌هایی را که به طور طبیعی وجود دارند میسر می‌سازند.

د) طرح‌های مجموعه زمانی یک گروه (تقسیم تصادفی نشده) دارد که در طی دوره مداخله‌های مختلف پیامدهای هر فرد را با خودش مقایسه می‌کند.

ه) طرح ضربدری اگر اثرات انتقالی مسئله‌ای نباشند می‌توانند مخدوش کننده را کنترل کنند و اندازه نمونه را به حداقل برسانند.

منابع

1. Chesnut CH, Silverman S, Andriano K, et al. A randomized trial of nasal spray salmon calcitonin in postmenopausal women with established osteoporosis: the PROOF study. *Am J Med*, 2000; 109:330-331.
2. Cummings S., Chapurlat R. What PROOF proves about calcitonin and clinical trials. *Am J*

Med, 2000; 109:267-276.

3. Hulley S, Grady D ,Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. JAMA 1998; 280:605-13.

4. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. N Engl J Med 1989; 321:406-12.

5. Pfeffer M, Stevenson L. Beta - adrenergic blockers and survival in heart failure. N Engl J Med 1996; 334:1396-7.

6. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984; 3:404-20

7. Friedman LM, DeMets DL, Furberg C. Fundamentals of clinical trials, 3rd ed. St. Louis: Mosby-Year Book, 1996.

8. Writing Group for the PEPI Trial. Effect of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. JAMA 1995; 273:199-208.

9. Cummings S, Black D , Thompson D, et al. Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. JAMA 1998; 280:2077-82.

10. Laupacis A, Connolly SJ , Gent M, et al. How should results from completed studies influence ongoing clinical trials? The CAFA Study experience. Ann Intern Med 1991; 115:818-22.

11. The Coronary Drug Project. Initial findings leading to modifications of its research protocol. JAMA 1970; 214:1303-13.

12. The Coronary Drug Project. Findings leading to discontinuation of the 2.5mg day estrogen group. The coronary Drug Project Research Group. JAMA 1973; 226:652-7.

13. Findings from the aspirin component of the ongoing Physicians Health Study. N Engl J Med 1988; 318:262-4.

14. Armitage P., McPherson C., Rowe B., Repeated significance test on accumulation data. J R Stat Soc 1969; 132A:235-44.

15. Hennekens C, Eberlein K. A randomized trial of aspirin and beta-carotene among U.S. physicians. Prev Med 1985; 14:165-8.

16. Diabetic Retinopathy Study Research Group. Preliminary report on: effects of photocoagulation therapy. Am J Ophthalmol 1976; 81:383- 96.

17. Walsh M, Hilton J., Masouedis C., et al. Smokeless tobacco cessation intervention for college athletes: results after 1 year. Am J Pub Health 1999;89: 228-34.

18. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. Am J Epidemiol 1981; 114: 906-14.

19. Chalmers T, Celano P, Sacks H, et al. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309: 1358-61.

20. Pocock S. Current issues in the design and interpretation of clinical trials. Br Med J 1985; 296: 39-42.

21. Marshall J. Beta-carotene: a miss for epidemiology. J Natl Cancer Inst 1999, 91: 2068-9.