# فصل ۷
# توجه به بازی شانس

## بازی شانس و قانون اعداد بزرگ

شواهد مطمئنی دربارهٔ تأثیرات درمان‌ها بر پیشگیری از وقوع سوگیری متکی است (و مدیریت سوگیری‌هایی که از آن‌ها جلوگیری نشده است). تا زمانی که این ویژگی‌های آزمون‌ها بی‌طرفانه به دست نیاید، هیچ‌گونه دست‌کاری در نتایج پژوهش‌ها نمی‌تواند مشکلات باقی‌مانده و عواقب خطرناک و گاهی اوقات کشندهٔ آن‌ها را حل کند (ن. ک: فصل‌های ۱ و ۲). حتی زمانی که اقدامات برای کاهش سوگیری موفقیت‌آمیز باشد، باز هم فرد می‌تواند به‌واسطهٔ نقش شانس گمراه شود.

اگر سکه‌ای را بارها و بارها پرتاب کنند، بعید نیست به‌دفعات روی شیر یا خط سکه مشاهده شود. اگر سکه را باز هم بیشتر پرتاب کنند، در پایان تعداد تقریباً برابری شیر یا خط خواهند داشت. به همین ترتیب، هنگام مقایسهٔ دو درمان، هر تفاوتی در نتایج ممکن است به‌سادگی منعکس‌کنندهٔ نقش شانس باشد. فرض کنیم ۴۰ درصد بیماران پس از دریافت درمان A و ۶۰ درصد از بیماران مشابه پس از دریافت درمان B فوت کنند. براساس جدول ۱، اگر ۱۰ بیمار هر دو درمان را دریافت کنند، چه نتیجه‌ای حاصل می‌شود. تفاوت در تعداد مرگ‌ومیر دو نوع درمان با عنوان «نسبت خطر» (risk ratio) بیان‌شده است. نسبت خطر در این مثال ۰٫۶۷ است.

بر پایهٔ این ارقام کوچک، آیا منطقی است که درمان A را بهتر از درمان B بدانیم؟

احتمالاً خیر. شاید برخی افراد به‌دلیل اثر شانس در یک گروه بهتر شده‌اند. اگر مقایسه در دیگر گروه‌های کوچکی از بیماران تکرار شوند، تعداد مرگ‌ومیر در هر گروه ممکن است معکوس شود (۶ در برابر ۴) یا برابر شود (۵ در برابر ۵) یا نسبت‌های دیگری که آن هم تنها به‌دلیل شانس به دست می‌آید.

**جدول ۱. بررسی برآورد مطمئن از تفاوت درمان‌های A و B**

| | درمان A | درمان B | نسبت خطر (A:B =) |
|---|---|---|---|
| **تعداد فوت‌شدگان** | ۴ | ۶ | ۰،۶۷ (= ۴: ۶) |
| **مجموع** | ۱۰ | ۱۰ | |

اگر پس از اینکه صد بیمار هریک از درمان‌های مذکور را دریافت کردند، همان نسبت از بیماران در هر گروه درمانی (۴۰ درصد و ۶۰ درصد) فوت کنند، چه انتظاری خواهید داشت (جدول ۲)؟ تفاوت (نسبت خطر) همان است که در جدول ۱ آمده است (۰/۶۷)، با این حال ۴۰ مورد مرگ در مقایسه با ۶۰ مورد مرگ نسبت به ۴ مورد مرگ در مقایسه با ۶ مورد مرگ، تفاوت چشمگیرتری دارد و نقش شانس در این حالت کمرنگ‌تر می‌شود؛ بنابراین برای اینکه مقایسه‌های درمانی شانس ما را گمراه نکند، باید مبتنی بر بررسی و ارزیابی تعداد زیادی از بیماران فوت‌شده، بدترشده یا بهبودیافته یا بدون تغییر نتیجه‌گیری کرد. این امر گاهی اوقات به «قانون اعداد بزرگ» شناخته می‌شود.

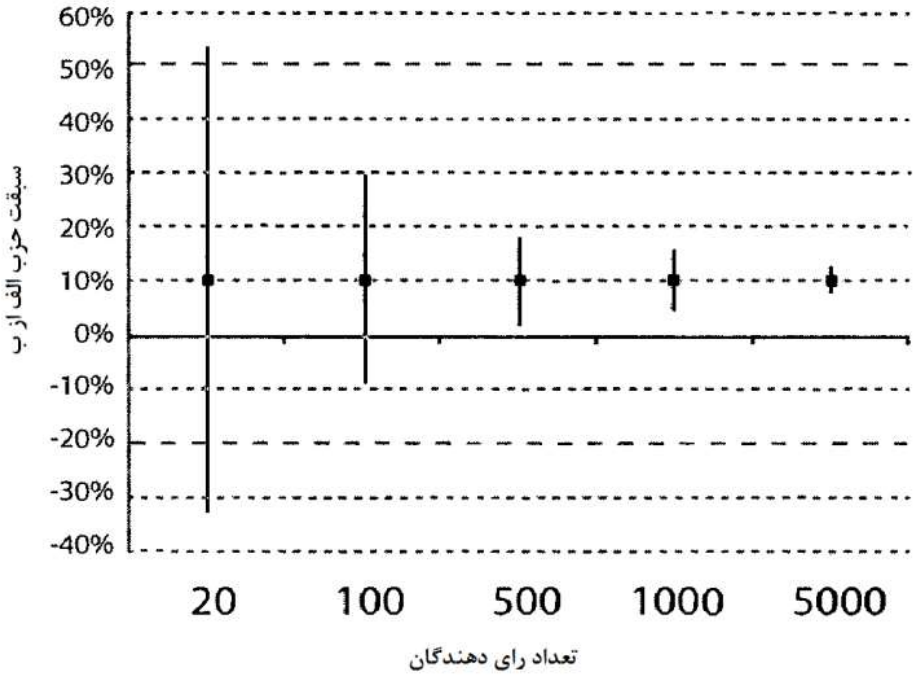
**جدول ۲. بررسی برآورد مطمئن از تفاوت درمان‌های A و B براساس تعداد نمونه‌های متوسط**

| | درمان A | درمان B | نسبت خطر (A:B =) |
|---|---|---|---|
| **تعداد فوت‌شدگان** | ۴۰ | ۶۰ | ۰،۶۷ (= ۴: ۶) |
| **مجموع** | ۱۰۰ | ۱۰۰ | |

### ارزیابی نقش شانس در آزمون‌های بی‌طرفانه

هنگام تفسیر نتایج مقایسه‌های بی‌طرفانه درمان‌ها، نقش شانس می‌تواند ما را گمراه کند. ممکن است به‌اشتباه نتیجه بگیریم که تفاوت‌های واقعی در پیامدهای درمانی وجود دارد؛ در حالی که این‌گونه نیست. یا اینکه تفاوتی وجود ندارد، در صورتی که تفاوت مشاهده می‌شود. هرچه تعداد پیامدهای درمان مورد نظر بیشتر باشد، احتمال اشتباه ما کمتر می‌شود. از آنجا که مقایسه‌های درمانی نمی‌توانند همهٔ افرادی که آن شرایط درمان را داشته‌اند یا خواهند داشت دربربگیرند، یافتن قطعی «تفاوت‌های حقیقی» درمان‌ها امکان‌پذیر نیست. در عوض، مطالعات باید بهترین حدس‌ها را دربارهٔ تفاوت‌های واقعی ارائه کنند.

اعتبار تفاوت‌های برآوردشده به‌کمک «فاصلهٔ اطمینان» (confidence interval= CI) نشان داده می‌شود. این فاصله محدوده‌ای را ارائه می‌دهد که در آن احتمالاً محدودهٔ تفاوت‌های واقعی قرار دارد. بیشتر افراد پیش از این با مفهوم فاصلهٔ اطمینان آشنا بوده‌اند، حتی اگر با این نام آشنا نباشند؛ برای مثال پیش از انتخابات، یک نظرسنجی ممکن است گزارش دهد که گروه الف، ۱۰ درصد جلوتر از گروه ب قرار دارد، اما این گزارش عنوان می‌کند که تفاوت دو گروه می‌تواند به کوچکی ۵ درصد یا بزرگی ۱۵ درصد باشد. این فاصلهٔ اطمینان نشان می‌دهد تفاوت واقعی بین دو گروه ممکن است جایی بین ۵ تا ۱۵ درصد قرار گرفته باشد. هرچه تعداد رأی‌دهندگان بیشتر باشد، عدم قطعیت نتایج کمتر خواهد شد؛ بنابراین فاصلهٔ اطمینان مربوط به تخمین تفاوت‌ها کمتر خواهد بود.




**کاهش فاصلهٔ اطمینان ۹۵ درصد برای تفاوت حزب الف و حزب ب با افزایش رأی‌دهندگان**

همان‌طور که می‌توان درجهٔ عدم قطعیت را پیرامون تفاوت برآوردشده در نسبت رأی‌دهندگان حمایت‌کننده از دو حزب سیاسی ارزیابی کرد، می‌توان آن را در نسبت بیمارانی که پس از دورهٔ درمانی بهبود یافته‌اند یا بدتر شده‌اند، ارزیابی کرد. در اینجا نیز هرچه تعداد پیامدهای درمان مشاهده‌شده- مثلاً بهبود پس از حملهٔ قلبی- در مقایسهٔ دو درمان بیشتر باشد، فاصلهٔ اطمینان پیرامون تخمین تفاوت درمان کمتر خواهد بود؛ از این‌رو هرچه فاصلهٔ اطمینان کمتر باشد، بهتر است.

فاصلهٔ اطمینان معمولاً نشان‌دهندهٔ این است که چقدر می‌توان مطمئن بود که مقدار واقعی، درون طیفی از تخمین‌های ارائه‌شده قرار دارد؛ برای مثال فاصلهٔ اطمینان ۹۵ درصد به این معنی است که می‌توان ۹۵ درصد مطمئن بود که مقدار واقعی در محدودهٔ فاصلهٔ اطمینان قرار دارد؛ بدین‌معنا که احتمال ۵ در ۱۰۰ (۵ درصد) وجود دارد که مقدار «درست» خارج از آن طیف قرار داشته باشد.

### «تفاوت معنی‌دار بین درمان‌ها» چه معنایی دارد؟

این پرسش گمراه‌کننده است؛ زیرا «تفاوت معنادار»، می‌تواند معانی مختلفی داشته

باشد. ممکن است به معنای تفاوت مهم برای بیمار باشد؛ با این حال زمانی که نویسندگان گزارش بیان می‌کنند «تفاوت معناداری» وجود دارد، بیشتر «معناداری آماری» مدنظر آن‌هاست. «تفاوت‌های آماری معنادار» در معنای رایج و مصطلح کلمه «معنادار» نیستند. تفاوت درمان‌ها که بسیار بعید است به‌دلیل شانس باشد – «تفاوت آماری معنادار» – ممکن است اهمیت بالینی کمی داشته یا هیچ اهمیتی نداشته باشد؛ برای مثال مروری نظام‌مند از کارآزمایی‌های تصادفی را در نظر بگیرید که تجارب ده‌ها هزار نفر از مردان سالم را که هر روز آسپرین مصرف می‌کنند، با تجارب ده‌ها هزار نفر از دیگر مردان سالمی که آسپیرین مصرف نمی‌کنند، مقایسه کرده است. این مطالعه، میزان سکتهٔ قلبی پایین‌تری میان مصرف‌کنندگان آسپیرین نشان داد و این تفاوت «از نظر آماری معنادار» بود؛ یعنی ممکن نیست که این تفاوت به‌واسطهٔ شانس ایجاد شده باشد، اما این امر نشان نمی‌دهد این نتایج اهمیت بالینی دارند. اگر در حال حاضر بروز سکتهٔ قلبی در فرد سالم بسیار پایین باشد، مصرف دارو برای پایین‌ترآوردن آن ممکن است توجیهی نداشته باشد؛ به‌ویژه که آسپرین تأثیرات جانبی دارد که برخی از آن‌ها مانند خونریزی، گاهی اوقات کشنده است.[۱] براساس شواهد مرور نظام‌مند می‌توان گفت اگر هزار مرد به مدت ده سال هر روز آسپرین مصرف کنند، پنج نفر از آن‌ها طی آن دورهٔ زمانی از سکتهٔ قلبی در امان می‌مانند، اما سه نفر آن‌ها دچار خونریزی جدی می‌شوند.

**«معنی‌داری آماری» یعنی چه؟**

«صادقانه بگویم این ایده‌ای گمراه‌کننده است. براساس این مفهوم می‌توان گفت مثلاً تفاوت یک دارو و دارونما یا امید به زندگی در دو گروه از افراد می‌تواند تنها ناشی از شانس باشد... بعید است دلیل چنین تفاوت بزرگی شانس باشد. متخصصان آمار از سطوح استاندارد «غیرمحتمل» استفاده می‌کنند. معمولاً آن‌ها از کلمهٔ «معنادار» در سطح ۵ درصد استفاده می‌کنند (گاهی به‌صورت P=۰/۰۵ نوشته می‌شود). در این حالت، تفاوتی معنادار تلقی می‌شود که اگر همه‌چیز بر مبنای شانس اتفاق بیفتد، احتمال وقوع کمتر از ۱ در ۲۰ خواهد بود».

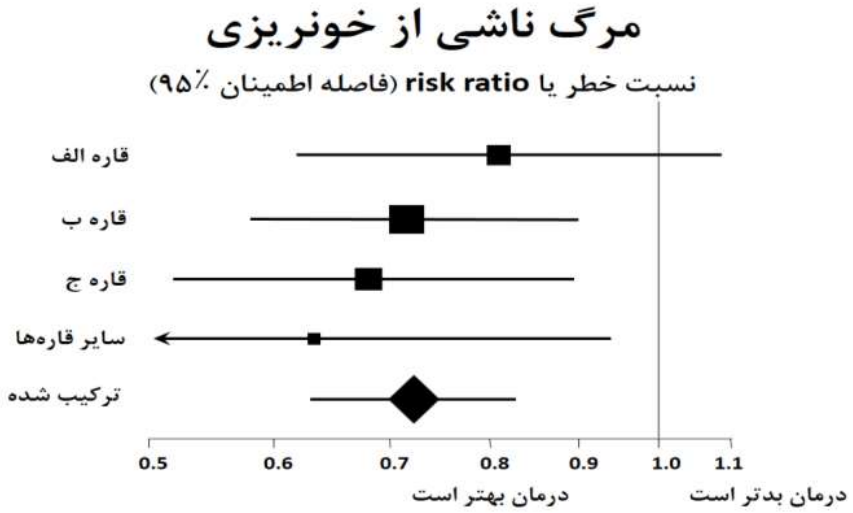
Spiegelhalter D, Quoted In: *Making Sense of Statistics*. 2010.
www.senseaboutscience.org

## دستیابی به اعداد بزرگ در آزمون‌های درمانی بی‌طرفانه

گاهی اوقات ممکن است در آزمون‌های درمانی اعداد بزرگ را از پژوهش‌هایی به دست بیاوریم که در یک یا دو مرکز انجام شده‌اند؛ با این حال به‌منظور بررسی تأثیر درمان بر پیامدهای نادر مانند مرگ، باید در بسیاری از مراکز و در بسیاری از کشورها از بیماران دعوت کنیم در پژوهش‌ها شرکت کنند تا شواهد قابل‌اعتمادی به دست بیاید؛ برای مثال مشارکت ۱۰ هزار نفر در سیزده کشور نشان داد داروهای استروئیدی تجویزشده برای افراد مبتلا به آسیب‌های مغزی جدی- درمانی که بیش از سه دهه استفاده می‌شد- کشنده بوده است.[۲] در یکی دیگر از آزمون‌های بی‌طرفانه که همان تیم پژوهشی انجام داده بود، مشارکت ۲۰ هزار نفر در چهل کشور نشان داد دارویی ارزان به نام ترانکسامیک اسید (Tranexamic Acid) مرگ ناشی از خونریزی پس از آسیب را کاهش می‌دهد.[۳] از آنجا که این مطالعات به‌منظور کاهش سوگیری و عدم قطعیت ناشی از نقش شانس طراحی شده‌اند، نمونه‌ای بارز از آزمون‌های بی‌طرفانه هستند و شواهدی با کیفیت خوب از ارتباط بسیار مهم با مراقبت‌های سلامت در سراسر جهان ارائه می‌دهند. درواقع، در یک نظرسنجی چاپ‌شده در نشریهٔ پزشکی بریتانیا (BMJ)،

دومین کارآزمایی از این کارآزمایی‌های تصادفی‌سازی‌شده به‌عنوان مهم‌ترین مطالعهٔ سال ۲۰۱۰ انتخاب شد.

نمودار زیر براساس داده‌هایی تنظیم‌شده که تیم برندهٔ جایزه BMJ ارائه داده است تا نشان دهد برای کاهش خطرات ناشی از گمراهی به‌دلیل نقش شانس مهم این است که تخمین‌های درمان بر مبنای اطلاعات بیشتر باشند. شکل لوزی پایین این نمودار نشان‌دهندهٔ نتیجهٔ کلی آزمایش ترانکسامیک اسید است. این شکل نشان می‌دهد داروی مذکور، مرگ ناشی از خونریزی را تا ۳۰ درصد کاهش می‌دهد (نسبت خطر دقیقاً بیش از ۰٫۷). این نتیجهٔ کلی مطمئن‌ترین تخمین اثر این دارو را ارائه می‌دهد؛ با این حال برآوردهای به‌دست‌آمده از مراکز قارهٔ الف نشان‌دهندهٔ اثر کمتر تأثیرگذار این داروست (که از نظر آماری معنادار نیست و به‌احتمال زیاد تخمینی کمتر از اثر واقعی است) و برآوردهای به‌دست‌آمده از مراکز در «قاره‌های دیگر»، نشان‌دهندهٔ اثری مهم است (که به‌احتمال زیاد تخمینی بیشتر از اثر واقعی است).




**تأثیر ترانکسامیک اسید بر مرگ‌ومیر در بیماران ترومایی با خونریزی قابل توجه بر اساس قاره (اطلاعات منتشرنشده از** CRASH-2 2010; 376:23-32 :*Lancet*)

همان‌گونه که می‌توان در یک آزمایش چندملیتی با ترکیب داده‌های به‌دست‌آمده از مراکز متعدد، نقش شانس را کاهش داد، گاهی اوقات نیز می‌توان نتایج مطالعات مشابه،

اما مجزا را از نظر آماری ترکیب کرد – فرایند شناخته‌شده به‌عنوان «متاآنالیز» (ن. ک: فصل ۸). اگرچه متخصصان آمار طی سال‌ها روش‌هایی برای متاآنالیز توسعه دادند، تا دهۀ ۱۹۷۰ میلادی (دهۀ ۱۳۵۰ شمسی) این روش مورد توجه نبود تا زمانی که دانشمندان علوم اجتماعی در ایالات‌متحده آمریکا، سپس پژوهشگران پزشکی از آن استفاده کردند. در پایان قرن ۲۰، متاآنالیز به‌عنوان جزئی مهم از آزمون‌های درمانی بی‌طرفانه پذیرفته شد؛ برای مثال پنج مطالعه در پنج کشور مختلف به‌طور جداگانه برای رسیدگی به درخواستی ۶۰ ساله و پاسخ داده نشده، سازمان‌دهی شدند و مورد حمایت مالی قرار گرفتند؛ چه سطحی از اکسیژن خون بیشترین احتمال زنده‌ماندن نوزادان نارس را بدون معلولیت‌های جدی تضمین می‌کند؟ اگر سطوح اکسیژن خون بیش‌ازحد بالا باشد، نوزادان ممکن است کور شوند، اگر هم بیش‌ازحد کم باشد، ممکن است بمیرند یا دچار فلج مغزی شوند؛ از آنجا که حتی در این نوزادان ضعیف، تفاوت ناشی از سطوح مختلف اکسیژن نسبتاً کم است، باید تعداد زیادی نمونه فراهم شود تا این تفاوت مشخص شود؛ بنابراین تیم‌های پژوهشی مسئول هریک از این پنج مطالعه توافق کردند تا شواهد مطالعات خود را برای ارائۀ برآورد قابل‌اعتمادتر نسبت به هریک از مطالعات مجزای خود با هم ترکیب کنند.[۴]

**نکتۀ کلیدی**

«نقش بازی شانس» را باید با ارزیابی «اطمینان» کیفیت و کمیت شواهد قابل‌دسترس مدنظر قرار داد.