ارزیابی نقادانهٔ شواهد

از مجموعه کتابهای آموزشی
تعالی خدمات بالینی

زیر نظر دکتر شهرام یزدانی

ویراستاران:

دکتر هادی بیهقی • دکتر سعید حاتمی • دکتر آرمین شیروانی

نویسندگان:

دکتر هادی بیهقی
دکتر الناز احتشامی افشار
دکتر سروش حسنی
دکتر زهرا رحمانی
دکتر مونا سید صفی‌زاده
دکتر سمیرا یزدانی‌نژاد

مرکز تعالی دانشگاه
دفتر تعالی خدمات بالینی

ارزیابی نقادانهٔ شواهد

ویراستاران:

دکتر هادی بیهقی ▪ دکتر سعید حاتمی ▪ دکتر آرمین شیروانی

نویسندگان:

دکتر هادی بیهقی

دکتر الناز احتشامی‌افشار

دکتر سروش حسنی

دکتر زهرا رحمانی

دکتر مونا سید صفی‌زاده

دکتر سمیرا یزدانی‌نژاد

از مجموعه کتابهای آموزشی

تعالی خدمات بالینی

زیر نظر دکتر شهرام یزدانی

مرکز تعالی دانشگاه

دفتر تعالی خدمات بالینی

ارزیابی نقادانهٔ شواهد

از مجموعه کتاب‌های آموزشی تعالی خدمات بالینی

طراح: آزیتا کنعانی

امور هماهنگی: معصومه روشن ضمیر

لیتوگرافی، چاپ، صحافی: سپهرنقش

ناشر: دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید بهشتی

چاپ اول: ۱۳۸۵

تیراژ: ۱۷۰۰

شابک: ۹۶۴-۸۸۵۸-۲۶-X

EAN: ۹۷۸۹۶۴۸۸۵۸۲۵۹

در طی دو دهه اخیر، جهان شاهد توجه ویژه‌ای به ارتقاء سازوکارهای سیاست‌گذاری بوده است. گسترش سریع مرزهای دانش سبب شده که کفایت اطلاعات و دانش محفوظ در اذهان سیاست‌گذاران زیر سؤال برود و دسترسی ایشان به شواهد موجود در حین فرایند سیاست‌گذاری به عنوان یک اصل تلقی گردد. امروزه بسیاری از مراکز سیاستگذاری در جهان پا را از این نیز فراتر نهاده‌اند و در جستجوی سازوکارهایی هستند که بکارگیری نظام‌مند شواهد در سیاستها را تضمین نماید. گذار از سیاست‌گذاری آگاه به شواهد به سیاست‌گذاری مبتنی بر شواهد را به جرأت می‌توان یکی از چالشهای عمده دستگاه‌های سیاست‌گذار در هزاره سوم دانست. بر این مبنا دانشگاه علوم پزشکی و خدمات بهداشتی و درمانی شهید بهشتی در راستای رسالت خود جهت تعالی پایه‌های نظام آکادمیک اقدام به تأسیس مرکز تعالی دانشگاه نموده است. این مرکز برآنست تا از طریق جمع‌آوری و تدوین مبانی نظری روا و روزآمد سیاست‌گذاران را در امر سیاست‌گذاری مبتنی بر شواهد یاری نماید.

مجموعه «تعالی خدمات بالینی» به عنوان اولین مجموعه مستندات این مرکز تدوین و منتشر گردیده است. به امید آنکه استمرار این حرکت بتواند کشور را در نیل به اهداف توسعه‌ای خود یاری نماید.

دکتر علیرضا زالی

رییس دانشگاه

پیشگفتار

ارایه خدمات بالینی کیفی از جمله تعهدات حرفه‌ای و اخلاقی هر پزشک، محسوب می‌گردد. نظام سلامت نیز در بخشی از عملکردخود به عنوان متولی سلامت مردم، موظف به اعمال سازوکارهای حاکمیت بالینی جهت تضمین کیفیت و تعالی خدمات بالینی می‌باشد. سیر شتابان پویایی دانش پزشکی و هجوم فن‌آوری‌های جدید، سبب گردیده است که عمل به این تعهدات در بعد فردی و سیستمی با دشواری‌های بسیاری مواجه گردد.

در طی دو دهه اخیر، دانش طبابت مبتنی بر شواهد، جهت پاسخگویی به این مشکلات به ویژه در بعد فردی شکل گرفته است و به موازات آن، سازوکارهای تعالی خدمات بالینی مانند ارزیابی فن‌آوری سلامت، راهکارهای طبابت بالینی و ممیزی بالینی جهت رفع این مشکلات در بعد سیستمی توسعه یافته‌اند. کشور ما نیز از این امر مستثنی نبوده است، در سالهای اخیر در اغلب دانشگاه‌های علوم پزشکی کشور، شاهد آگاهی فزاینده‌ای در رابطه با مفاهیم پایه طبابت مبتنی بر شواهد بوده‌ایم.

تمایل به عملیاتی‌سازی این مفاهیم از یک سو و فقدان یک مجموعه آموزشی مدون در رابطه با مفاهیم اساسی تعالی خدمات بالینی از سوی دیگر، موجب گردید تا کارشناسان دفتر تعالی خدمات بالینی دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید بهشتی، اقدام به تالیف مجموعه حاضر، نمایند.

این مجموعه به عنوان بسته پایه آموزشی تعالی بالینی جهت توانمندسازی دانشجویان، دستیاران و هیات علمی دانشگاه‌ها در قالب یادگیری خود محور یا آموزش کارگاهی قابل استفاده می‌باشد.

دکتر شهرام یزدانی
مدیرکل مرکز تعالی دانشگاه

فهرست

مقدمه

در طبابت مبتنی بر شواهد، پزشکان و سایر ارائه دهندگان مراقبتهای پزشکی بر مبنای شواهدی که از منابع مختلف بدست می‌آورند برای بیماران تصمیم‌گیری می‌کنند، ولی آیا هر شواهدی با هر درجه از اعتبار می‌تواند مبنای صحیحی برای تصمیم‌گیری باشد؟ طبق شواهد موجود بیش از ۳۰ درصد از توصیه‌های درمانی شواهدی که از اینترنت بدست می‌آید اشتباه و گمراه کننده است. در منابع دیگر نیز آمار تقریباً مشابهی وجود دارد. درصد بالایی از این توصیه‌های اشتباه حاصل اشکالات و کاستی‌های موجود در پژوهش‌های پزشکی است، که از ریشه‌های متعددی چون آشنا نبودن پژوهشگران با مبانی روش شناسی پژوهش، فرضیات نادرستی که در ذهن پژوهشگر وجود دارد، محدودیت‌های پژوهش، بیان غیرشفاف و ارائه نتایج غیرواقعی منشاء می‌گیرد.

برای شناخت نقاط ضعف و قوت شواهد، امروزه علمی به نام ارزیابی نقادانه شواهد[1] شکل گرفته است. دامنه این علم فراتر از پزشکی و طبابت است و روشهای آن در ارزیابی کلیه پژوهشها و متون علمی جایگاه دارد.

ارزیابی نقادانه شواهد مربوط به طبابت، با اینکه مقوله نسبتاً نوپایی است ولی در سالهای اخیر بدنبال روند رو به رشد طبابت مبتنی بر شواهد و نیاز روزافزون به تدوین مقالات مروری نظام‌مند پیشرفت قابل توجهی داشته است. این پیشرفت در حیطه‌های مختلف شواهد به یک میزان نبوده است. به عنوان مثال برای ارزیابی شواهدی که مربوط به اثربخشی مداخلات درمانی و دقت تستهای تشخیصی هستند، چارچوب کار بخوبی مشخص شده و ابزارهای متنوعی برای آن تدوین شده است. حتی مطالعات ثانویه‌ای برای تعیین اثربخشی این ابزارها صورت گرفته است، ولی در سایر حیطه‌ها مانند شواهد مربوط به غربالگری، تعیین پیش‌آگهی، شیوع و بروز بیماریها و ... هنوز جا برای کار بیشتر وجود دارد.

1- Critical Appraisal

ارزیابی نقادانه شواهد در حقیقت دانش دور ریختن شواهدی است که مبنای صحیحی برای تصمیم‌گیری در مورد بیمار فراهم نمی‌کنند. هنگامیکه شما در بالین بیمار به مشکلی برمی‌خورید، بدنبال جستجو در منابع مختلف با توجه به نحوه جستجوی خود به شواهدی دست خواهید یافت که احتمالاً بیشترین ارتباط[1] را به مشکل شما دارند. جدا کردن شواهد مربوط[2] در اکثر مواقع از طریق مطالعه عنوان یا خلاصه مقالات براحتی امکانپذیر است. قدم بعد یافتن شواهد کیفی و اثربخش از میان شواهد مربوط است که نیازمند بررسی بیشتر و ترجیحاً مطالعهٔ تمام متن مقاله می‌باشد. این بخش از کار یکی از مراحل مهم و اساسی در طبابت مبتنی بر شواهد است و در اکثر منابع مربوط به طبابت مبتنی بر شواهد، قسمت اعظم مطالب را به خود اختصاص داده است.

مطالعاتی که ممکن است در یک جستجو بدست آید، بالقوه می‌تواند طرحهای متنوعی داشته باشد ولی بسته به سوأل بالینی و نحوه جستجو احتمال دستیابی به تعداد محدودی از انواع مطالعات بیشتر است. در هر صورت از آنجاییکه نقایص و کاستی‌های مطالعات عمدتاً از مسائل مربوط به روش‌شناسی پژوهش (مانند نحوه انتخاب و حجم نمونه، محاسبه دقت آماری، نحوه آزمون فرضیه و ...) منشأ می‌گیرد، بی‌گمان ارزیابی نقادانه مطالعات بدون آشنایی با نحوه طراحی و اجرای آنها میسر نخواهد بود. با این حال، این گفته بدان معنا نیست که برای ارزیابی نقادانه شواهد نیاز به آگاهی عمیق و درک کاملی از جزئیات روش‌شناسی پژوهش وجود دارد. بر این مبنا فصل اول کتاب به مروری اجمالی بر روش‌شناسی پژوهش با تأکید بر مشکلات و خطاهایی که در هر کدام از مراحل پژوهش ممکن است پیش بیاید، اختصاص یافته است.

1- Relevancy
2- Relevant

فصل اول

روش‌شناسی پژوهش

فصل اول

روش‌شناسی پژوهش

اهداف یادگیری:

- آشنایی با تعاریف، راهبردها و طبقه‌بندی پژوهشها
- آشنایی با فرایند کلی پژوهش
- آشنایی با مراحل مختلف پژوهش
- درک مفاهیم روش‌شناسی پژوهش و به کارگیری آنها در ارزیابی نقادانه شواهد.

روش‌شناسی پژوهش[1] چیست؟

برای پاسخ به این سؤال ابتدا باید منظور خود را از پژوهش و روش بیان کنیم.

پژوهش روشی نظام‌مند و قانون‌مدار برای رسیدن به شواهد است. این شواهد در حیطه مورد نظر این کتاب برای حل مشکلاتی است که در طبابت و مراقبت‌های پزشکی بوجود می‌آید. پژوهش از آن جهت نظام‌مند شناخته می‌شود که پژوهشگر معمولاً توالی خاصی از فرایندها و اقدامات را برای رسیدن به نتیجه انجام می‌دهد. از طرف دیگر پژوهش طبق اصول و قوانین مشخص و شفافی انجام می‌شود. این قوانین و اصول پایه‌های روش علمی[2] برخورد با مشکلات را تشکیل می‌دهند.

1- Research Methodology
2- Scientific method

روش[1] یک فرایند نظام‌مند برای انجام یک فعالیت مشخص است. اختصاصاً در زمینه مورد بحث روش مجموعه‌ای از اصول است که مشخص می‌کند:

۱ چگونه دانش باید کسب شود.

۲ این دانش چگونه و به چه شکلی باید بیان شود.

۳ چگونه می‌توان واقعی یا غیرواقعی بودن دانش را ارزیابی کرد.

فرایند پژوهش

در فرآیند پژوهش مراحلی که به طور معمول در اکثر پژوهش‌ها انجام می‌شود بیان می گردد. بعضی از مراحل این فرایند در انواع مختلف پژوهش اهمیت بیشتری پیدا می‌کند و بعضی مراحل نیز اهمیت کمتری داشته و حتی ممکن است با توجه به طرح پژوهش و نوع سوأل پژوهشی از فرایند کارحذف شوند. به طور کلی فرایند پژوهش شامل مراحل زیر است:

شکل ۱

طرح ریزی پژوهش[2]

اولین مرحله در انجام یک پژوهش طرح‌ریزی دقیق پژوهش است. طرح و برنامه مراحلی که قرار است در فرایند پژوهش اتفاق بیفتد در سندی تحت عنوان پروتکل پژوهش نوشته می‌شود. پژوهشگران معمولاً زمان زیادی را پیش از شروع به جمع‌آوری داده‌ها به طرح‌ریزی پژوهش اختصاص می‌دهند.

1- Method
2- Research planning

سوأل پژوهش:

قدم اول در طرح ریزی یک پژوهش انتخاب یک سوأل پژوهشی است. روشهای رسیدن به سوأل پژوهشی بسیار متنوع است ولی با توجه به اهداف این فصل در اینجا بحث نمی‌شود. به طور کلّی معمولاً پروژه‌های پژوهشی می‌تواند به سه نوع سوأل پاسخ دهند. بدیهی است که پاسخ دادن به هر نوع از این سوأالات مستلزم نوع خاصی از طرحهای پژوهش است که در ادامه به آنها اشاره خواهد شد.

۱ سوأالات توصیفی[۱]: در این سوأالات مطالعه بطور اولیه برای توصیف یک پدیده موجود یا در حال ایجاد، طراحی می‌شود. مطالعاتی که در زمینه شیوع و وقوع بیماریها انجام می‌شود با این نوع سوأالات آغاز شده است.

۲ سوأالات ارتباطی[۲]: در این نوع سوأالات، مطالعات برای کشف رابطه بین دو یا چند متغیر طراحی می‌شوند. مثل مطالعاتی که رابطه بین عوامل خطر و بیماریها را بررسی می‌کنند.

۳ سوأالات علّی[۳]: در این نوع سوأالات، مطالعات برای تعیین اینکه آیا یک یا چند متغیر می‌تواند علت یک یا چند متغیر دیگر باشد، طراحی می‌شوند. برای مثال اگر پژوهشی برای تعیین اثر یک رژیم درمانی بر روی عوارض یک بیماری خاص انجام دهید در حقیقت تأثیر این رژیم درمانی (علت) را بر تغییرات عوارض آن بیماری خاص (اثر)[۴] مطالعه کرده‌اید.

این پرسش‌ها می‌توانند به صورت تجمعی نیز مورد توجه قرار گیرند، برای مثال مطالعات ارتباطی فرض می‌کنند که قبلاً هر کدام از متغیرهایی که می‌خواهید رابطه بین آنها را پیدا کنید از طریق مشاهده یا سنجش توصیف کرده اید. همچنین یک مطالعه علّی فرض می‌کند که شما قبلاً متغیرهای علت و معلول را توصیف کرده‌اید و وجود رابطه بین آنها را نشان داده‌اید.

1- Descriptive
2- Relational
3- Causal
4- Effect

فرضیه[1] و هدف[2]

هدف معمول در انجام پژوهش تست کردن یک فرضیه است. فرضیه‌ها بیانیه‌هایی راجع به ارتباط بین متغیرهای مختلف و تفاوت بین گروهها هستند. در حقیقت فرضیه عبارت اختصاصی است که پیش‌بینی ما را از نتیجه پژوهش نشان می‌دهد. فرضیه‌ها می‌توانند راجع به مشاهده‌ها یا متغیرها باشند. به عنوان مثال «آیا بین میزان ورزش کردن و هزینه سالانه مراقبتهای سلامتی همراهی وجود دارد؟» فرضیه‌ها همچنین می‌توانند راجع به تفاوت بین گروهها باشند مانند «آیا بیمارانی که تحت درمان با مداخله X قرار گرفته‌اند پیشرفت بهتری نسبت به افرادی که تحت درمان Y بوده‌اند نشان می‌دهند؟»

توجه داشته باشید که تمام مطالعات یک فرضیه رسمی برای تست کردن ندارند. گاهی اوقات مطالعات برای اکتشاف طراحی می‌شوند. هدف از اینگونه مطالعات کشف کامل‌تر حیطه مورد نظر است. تا از این طریق بتوان فرضیات و پیش‌بینی‌های اختصاصی‌تری پیدا کرد که در پژوهشهای آینده قابل تست کردن باشد. بنابراین نداشتن فرضیه همیشه نشانه کاستی در پژوهش نیست و می‌تواند بدین معنا باشد که این نوع مطالعات اهداف متفاوتی نسبت به مطالعات دیگر دنبال می‌کند، در این گونه موارد به جای کاربرد فرضیه از واژه هدف استفاده می‌کنیم.

شرط لازم برای بیان فرضیه یا هدف پژوهش این است که سؤال پژوهش باید یک متناظر عملی و تجربی داشته باشد. در طرح‌ریزی پژوهش، پژوهشگر باید به مفاهیم کلی و مبهم جنبه عملی و تجربی بدهد. برای این کار پژوهشگر باید متغیرهای مورد مطالعه را مشخص کرده و نحوه مشاهده و سنجش آنها را به دقّت بیان کند. به عنوان مثال مفهوم بیماریهای عروق کرونر را در نظر بگیرید متغیرهای قابل بررسی در آن می‌تواند شامل درد دوره‌ای قفسه سینه و بازوی چپ، انسداد و کاهش جریان خون عروق کرونر و یا مرگ ناگهانی باشد. حال اگر فرض کنیم یک داروی جدید (X) برای کمک به این بیماران ساخته شده است. پژوهشگر تصور

1- Hypothesis
2- Aim

می‌کند که داروی جدید X در کمک به مبتلا یان به بیماری عروق کرونر اثربخش است، با استفاده از متغیرهای فوق فرضیه پژوهش می‌تواند به صورتهای زیر بیان شود.

در افرادی که بیماری عروق کرونر در آنها تشخیص داده شده و از داروی X استفاده می‌کنند، درد قفسه سینه و بازوی چپ نسبت به بیمارانی که از داروی رایج Y استفاده می‌کنند، وقوع کمتری دارد.

این بیماران نسبت به افرادیکه از داروی رایج Y استفاده می‌کنند جریان خون بیشتری در عروق کرونر خود دارند.

افراد مبتلا به بیماری عروق کرونر که از داروی X استفاده می‌کنند نسبت به افرادیکه از داروی Y استفاده می‌کنند مرگ و میر کمتری در یک دوره پنج ساله خواهند داشت.

راهبردهای پژوهش[1]

طرح‌ریزی یک پژوهش، انتخاب و استفاده از راهبردهای پژوهش را نیز شامل می‌شود. راهبردهای پژوهش فرایندهای مشخص شده‌ای برای طرح و اجراء پژوهشها هستند.

پیش از بررسی این راهبردها با دو رویکرد کلی به پژوهش آشنا می‌شویم، رویکرد کیفی[2] و رویکرد کمّی[3] . به طور کلی دریافت‌های ما از وضعیت بیماران در غالب دو چارچوب متفاوت ولی وابسته به هم شکل می‌گیرد.

در چارچوب اول، بیمار به عنوان یک سیستم بیولوژیک در نظر گرفته می‌شود که به دلایلی عملکرد خوبی ندارد و در چارچوب دوم به بیمار به چشم یک انسان نگاه می‌شود که در جامعه زندگی می‌کند و سعی می‌کند تا تصویری از بیماری خود داشته باشد و به گونه‌ای با مشکل سلامت خود کنار بیاید.

1- Research Strategies
2- Qualitative
3- Quantitative

چارچوب اول مستلزم داشتن رویکرد کمی نسبت به پژوهش و دانش است. در این رویکرد بیمار را از دیدگاه عینی[1] نگاه می کنیم و تلاش می‌کنیم متغیرهای مهمی را که نشاندهنده علل و تظاهرات وضعیت بالینی اوست مشخص کرده و بسنجیم.

چارچوب دوم در داشتن رویکردی کیفی به پژوهش و دانش معنی پیدا می‌کند. در این رویکرد بیمار به دید یک انسان کامل نگاه می‌شود و سعی می‌شود تا درک صحیحی از تجربه[2] و دلایل بیمار برای فعالیتهایی که در شرایط خاص انجام می‌دهد حاصل گردد.

از سوی دیگر کیفیت و کمّیت در پژوهش به نوع و طبیعت داده‌ها و شواهد نیز مربوط می‌شود. اگر داده‌ها به صورت عددی باشند کمی نامیده می‌شوند و در غیر اینصورت کیفی. توجه داشته باشید که داده‌های کیفی می‌توانند موارد بیشتری به جز واژه و متن باشند بطوریکه عکس، فیلم، نوارهای صوتی و مانند آنها را نیز می‌توان داده‌های کیفی تلقی کرد.

حال که با مفهوم کیفیت و کمیت در پژوهش آشنا شدید به بررسی راهبردهای پژوهش می‌پردازیم. راهبردهایی که به طور اولیه برای انجام مطالعات قابل استفاده هستند، بنابر نوع سؤال پژوهش به دو دستهٔ راهبردهای توصیفی[3] و تحلیلی[4] تقسیم می‌شوند.

راهبردهای توصیفی

در مطالعاتی که از این راهبرد تبعیت می‌کنند، عمدتاً وضع سلامت افراد در یک جمعیت خاص توصیف می‌شود. داده‌هایی که به طور مستمر در سطح جامعه گزارش می‌شود و یا مطالعات پیمایشی خاص اطلاعات مورد نیاز برای انجام این مطالعات را تامین می‌کنند. توجه داشته باشید که در اینگونه مطالعات، هیچ تلاشی برای تحلیل روابط بین متغیرها صورت نمی‌گیرد، بنابراین قابلیت آزمودن فرضیه در راهبرد توصیفی وجود ندارد و کاربرد آن عمدتاً تولید فرضیه برای مطالعا ت بعدی است. مطالعاتی که الگوهای مرگ و میر را در سنین، جنسیت‌ها و نژادهای مختلف در یک دورهٔ زمانی مشخص مورد بررسی قرار می‌دهند، نمونه‌ای از مطالعات توصیفی هستند.

1- Objective
2- Subjective
3-Descriptive Strategy
4-Analytic

راهبرد تحلیلی [1]

راهبرد تحلیلی امکان آزمودن فرضیه را فراهم می‌کند. برای انجام مطالعات تحلیلی پژوهشگر دو راه پیش رو دارد.

۱ پژوهشگر می‌تواند از طریق مشاهده و سنجش متغیرها و تحلیل رابطهٔ آنها وجود رابطه و نوع آن را مشخص کند (طرحهای مشاهده‌ای)[2]

۲ پژوهشگر می‌تواند با انجام یک مداخلهٔ کنترل شده بر روی یک گروه خاص از جمعیت و در صورت امکان مقایسهٔ این گروه با گروه شاهد که مداخله را دریافت نکرده است، به استنباط صحیحی از اثر مداخله و رابطهٔ آن با پیامدهای مورد نظر دست یابد (طرحهای مداخله‌ای)[3]

طرحهای مداخله‌ای به نوبهٔ خود از سه راهبرد کلی تبعیت می‌کنند.

۱ راهبرد غیرتجربی

سه جزء ضروری در راهبرد غیرتجربی در شکل ۲ نشان داده شده‌است. در این راهبرد پژوهشگر جمعیت مورد نظر را تعریف کرده، افراد را برای مطالعه انتخاب نموده و بعد از انجام مداخلهٔ مورد نظر داده‌های نمونه را جمع‌آوری و مطالعه می‌کند. (شکل ۲)

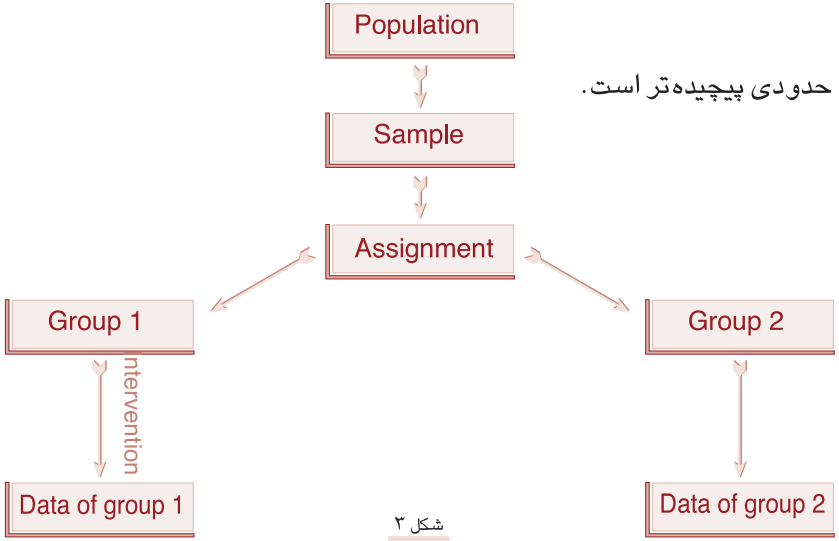
شکل ۲

1- Analytic Srategy
2- Observational Studies
3- Interventional Studies

توجه کنید که در این راهبرد، کنترلی بر روی پژوهش صورت نمی‌گیرد یا به عبارت دیگر چیزی تحت عنوان گروه کنترل یا شاهد وجود ندارد. (منظور از کنترل در یک مطالعه کاهش احتمال وجود فرضیه‌هایی غیر از فرضیه اصلی پژوهش است که می‌توانند اثر پیش‌بینی شده را توجیه کنند). این راهبرد در نشان دادن رابطه علّی از بقیه راهبردها ضعیف‌تر است.

۲. راهبرد تجربی

راهبرد تجربی معمولاً مراحل بیشتری دارد و تا حدودی پیچیده‌تر است. (شکل ۳)



شکل ۳

مشابه راهبرد غیرتجربی در راهبرد تجربی نیز افراد از یک جمعیت تعریف شده انتخاب می‌شوند، سپس از طریق انتخاب تصادفی شده یا روشهای مشابه آن به گروه کنترل و مداخله وارد می‌شوند، (مداخله در اینجا به معنی اثرگذاری یا اقدامی است که توسط

پژوهشگر به طور فعال و نظام‌مند بر روی نمونه پژوهش اعمال می‌شود). سپس گروهها مورد سنجش و مقایسه قرار می‌گیرند.

اگر تمام مراحل بدرستی انجام شود هر تفاوتی را که در داده‌های بدست آمده از دو گروه دیده می‌شود می‌توان به اثر مداخله انجام شده نسبت داد. متأسفانه غالباً عواملی به جز مداخله، وجود دارند که می‌توانند باعث تفاوت مشاهده شده شوند. تکنیکهای موجود در این راهبرد سعی دارند اثر این عوامل را از طریق اعمال کنترل حذف کنند.

۳| راهبرد شبه تجربی

این راهبرد شبیه به راهبرد تجربی است و می‌تواند شواهد مستدلی برای رابطه علّی فراهم کند. تفاوت عمده بین این دو راهبرد در نحوه تخصیص[1] نمونه گرفته شده به گروههای مداخله و کنترل است. در راهبرد تجربی از روش تخصیص تصادفی شده یا نظایر آن استفاده می‌شود ولی در راهبرد شبه تجربی اینطور نیست. با این حال در این راهبرد داشتن کنترل روی نحوه انجام مداخله و زمانهایی که مداخله باید انجام شود مشابه زاهبرد تجربی امکانپذیر است. در بعضی منابع، طرحهایی که در آنها گروههای مختلف و یا سنجش‌های متعدد (در زمانهای مختلف) وجود دارد جزء این راهبرد در نظر گرفته شده است. از آنجاییکه در این نوع راهبرد داشتن کنترل بر روی نحوه انتخاب افراد برای گروههای مختلف و نحوه انجام مداخلات دشوارتر است، مطالعه و پاسخگویی به سوأالات پژوهشی که مربوط به رابطه علّی هستند با قطعیت کمتری انجام می‌شود و نسبت به راهبرد تجربی از روایی[2] کمتری در این زمینه برخورداراست.

1- Assignment
2- Validity

نمونه گیری و روایی خارجی[1]

پژوهش در علوم پزشکی معمولاً از طریق جمع‌آوری داده‌ها مربوط به نمونه‌ای از جمعیت مورد نظر انجام می‌شود. چرا که مطالعه کل جمعیت معمولاً غیرممکن و یا بسیار پرهزینه است. به عنوان مثال وقتی می‌خواهیم بر روی افراد دیابتی مطالعه کنیم، حجم بسیار زیاد جمعیت دیابتیها، عدم دسترسی به تمام این بیماران و وجود افرادیکه به اشتباه دیابتی تشخیص داده شده‌اند و یا دیابت آنها تشخیص داده نشده است، مطالعه بر روی تمام دیابتیها را غیرممکن می‌سازد. بنابراین در راهبردهای مختلف پژوهش، پژوهشگر به طور معمول گروه یا نمونه‌ای از جمعیت را مطالعه می‌کند و سپس یافته‌های آنها را به جمعیت مورد نظر تعمیم می‌دهد. نمونه‌گیری فرایند انتخاب واحدهای پژوهش (افراد، سازمانها، گروهها و ...) از جمعیت مورد نظر است به طوریکه با مطالعه نمونه بتوان نتایج را بدرستی به جمعیتی که نمونه از آن گرفته شده تعمیم داد.

روایی خارجی عمده‌ترین مسأله‌ای است که باید در نمونه‌گیری مدنظر باشد. روایی خارجی بیان می‌کند تا چه میزان نتایج مطالعه در مورد افراد، مکانها و زمانهای دیگر قابل تعمیم است. به بیان دیگر هر چه نمونه گرفته شده نماینده بهتری [2] از جمعیت مورد نظر باشد روایی خارجی مطالعه بیشتر است.

در اینجا با چند تعریف مقدماتی آشنا می‌شویم:

اولین سوأالی که در نمونه‌گیری مطرح می‌شود این است که نتیجه مطالعه را به چه کسانی می‌خواهید تعمیم بدهید؟

در بیشترموارد پژوهشگر علاقه دارد تا نتایج مطالعه خود را به طور کلی بیان کرده و نتایج را صرفاً به جمعیت مورد نظر خود محدود نکند. به جمعیتی که پژوهشگر علاقه دارد نتایج پژوهش را تعمیم دهد جمعیت تئوری[3] گفته می‌شود (مانند جمعیت دیابتی‌های یک کشور)، ولی از آنجایی که نمونه‌گیری صحیح از این جمعیت غالباً غیرممکن است، پژوهشگر مجبور است از جمعیتی

1- Sampling and external validity
2- Representative
3- Theoretical Population

که در دسترس او هستند استفاده کند (مانند جمعیت دیابتیهای یک شهر). این جمعیت همان جمعیتی است که نمونه از آن انتخاب می‌شود و نتایج مطالعه مستقیماً و بدون واسطه می‌تواند به آن تعمیم داده شود. اسامی مختلفی به این جمعیت داده شده است از جمله جمعیت در دسترس[1] و جمعیت مطالعه[2].

جمعیت مطالعه باید پیش از آغاز پژوهش تعریف شود. برای این کار ابتدا ویژگیهایی که هر کدام از افراد برای ورود به مطالعه باید داشته باشند، تعریف می‌شود (معیارهای ورود و خروج).[3] همانطور که تاکنون گفته شد تعریف این ویژگیها نقش مهمی در تفسیر و تعمیم نتایج پژوهش دارد.

با مثال فرضی زیر به اهمیت معیارهای ورود و خروج بیشتر پی خواهید برد.

پژوهشگری می‌خواهد اثربخشی یک درمان را بر سرطان پستان[4] مطالعه کند او تمام بیماران مبتلا به سرطان پستان را که در دسترس هستند انتخاب کرده و مشاهده می‌کند درمان مورد مطالعه به طور میانگین منجر به هیچ پیشرفتی در پیامدها نشده است.

پژوهش‌های بعدی بهبود اساسی در پیامدها را با استفاده از درمان مشابه در خانمهای مبتلا به مرحله III سرطان پستان نشان می‌دهد. همچنین ثابت می‌شود خانمهایی که مبتلا به سرطان پستان بوده و تحت رادیوتراپی قرار گرفته‌اند هیچ سودی از این درمان نمی‌برند.

چنانچه این پژوهش برای تعریف جمعیت مورد نظر خود از معیار ورود مرحله III سرطان پستان و معیار خروج سابقه رادیوتراپی قبلی استفاده می‌کرد، نتایج آن می‌توانست بسیار متفاوت باشد. معیارهای خروج در مطالعات معمولاً شرایط خاصی را که ممکن‌است درمان را پیچیده‌تر کند و یا در تفسیر نتایج اختلال ایجاد کند، حذف می‌نماید.

این معیارها علاوه بر آنکه ویژگیهای افراد مطالعه را مشخص می‌کند جمعیتی که نتایج را می‌توان بدون واسطه به آنها تعمیم داد نیز محدود می‌نماید. برای مثال خانمهایی که تحت عنوان مرحله II سرطان پستان تشخیص داده می‌شوند نمی‌توانند وارد مطالعه

1- Accessible population
2- Study population
3- Inclusion and exclusion criteria
4- Breast cancer

شوند و در نتیجه نمی‌توان نتایج مطالعه را مستقیماً به آنها تعمیم داد.

حال که با اهمیت تعریف جمعیت مطالعه آشنا شدیم قدم بعدی تهیه لیستی از اعضای جمعیت مطالعه است که نمونه پژوهش از این لیست انتخاب می‌شود. به این لیست در اصطلاح چارچوب نمونه‌گیری[1] گفته می‌شود. به رویکردهایی که برای انتخاب نمونه از جمعیت مطالعه وجود دارد روشهای نمونه‌گیری[2] گفته می‌شود. با استفاده از این روشها نهایتاً به یک نمونه[3] می‌رسیم. توجه داشته باشید که نمونه گروهی از افراد یا واحدهاست که شما انتخاب کرده‌اید، تا در مطالعه شرکت کنند. این گروه الزاماً معادل افرادی نیست که واقعاً در مطالعه شرکت می‌کنند. مثلاً ممکن است تعدادی از افراد انتخاب شده را نتوانید پیدا کنید و یا آنها از شرکت در مطالعه اجتناب کنند. بروز چنین مواردی می‌تواند روایی خارجی مطالعه شما را تحت تأثیر قرار دهد.

روشهای نمونه‌گیری

روشهای نمونه‌گیری به دو نوع کلی احتمالی[4] و غیراحتمالی[5] تقسیم می‌شوند. در روشهای نمونه‌گیری احتمالی هر کدام از افراد جمعیت مطالعه دارای شانس انتخاب شدن در نمونه پژوهش هستند. در تعدادی از این روشها، این شانس برای هر کدام از افراد یکسان است مانند روشهایی که از انتخاب تصادفی واقعی[6] استفاده می‌کنند. در تعدادی دیگر نیز این شانس الزاماً برای تمام افراد یکسان نیست مثل روشهایی که از انتخاب شبه تصادفی[7] استفاده می‌کنند (مانند انتخاب بر اساس روز مراجعه یا حرف اول فامیل و ...). از آنجاییکه در روشهای انتخاب تصادفی هر کدام از افراد شانس برابری برای انتخاب شدن دارند نمونه بدست آمده نماینده بهتری از جمعیت مطالعه خواهد بود. بنابراین مطالعه از روایی خارجی بالاتری برخوردار می‌شود.

1- Sampling Frame
2- Sampling methods
3- sample
4- Probability sampling
5- non probability Sampling
6- true random selection
7- Quasi-random selection

روشهای نمونه گیری احتمالی

نمونه‌گیری تصادفی ساده[1]

این روش ساده‌ترین نوع نمونه‌گیری تصادفی است که در آن ابتدا لیست افراد تهیه می‌شود و بعد از طریق روشهای تصادفی مثال جدول اعداد تصادفی، تولیدکننده‌های رایانه‌ای اعداد تصادفی (نرم افزارهایی مثل اکسل EXCEL) و ... تعداد مورد نظر انتخاب می‌شود. با اینکه انجام این روش بسیار ساده است ولی ممکن است در بعضی موارد نمونه بدست آمده نماینده خوبی از جمعیت مورد مطالعه نباشد (بخش تحلیل را ببینید) و یا اساساً انجام این روش غیرممکن یا پرهزینه باشد.

نمونه‌گیری طبقه‌ای تصادفی[2]

در این روش جمعیت مورد مطالعه به گروههای هموژن مستقل تقسیم می‌شود و از هر گروه تعدادی از افراد به صورت تصادفی انتخاب می‌شوند. در این روش نمونه نه تنها نماینده تمام جمعیت است بلکه تمام زیر گروههای مهم جمعیت نیز مطمئناً در نمونه حضور خواهند داشت. در نمونه‌گیری تصادفی ساده ممکن بود به علت پدیده شانس این گروهها (خصوصاً گروههای اقلیت) نماینده‌ای در نمونه نداشته باشند.

از طرف دیگر از آنجایی که هر کدام از طبقه‌ها[3] در داخل خود نسبت به جمعیت مطالعه همگون‌تر هستند در این روش دقت آماری[4] مطالعه بالاتر می‌رود.

1- Simple random sampling
2- Proportional, Quata or stratified random sampling
3- Strata
4- Statistical Precision

نمونه‌گیری تصادفی خوشه‌ای[1]

در مواردی که جمعیت مطالعه در نواحی جغرافیایی گسترده‌ای توزیع شده است ممکن است استفاده از نمونه‌گیری تصادفی ساده یا طبقه‌ای ممکن نباشد، در این موارد منطقه جغرافیایی مورد نظر به قسمتهای کوچکتری (مانند مناطق شهرداری در شهرها) تقسیم می‌شود و تعدادی از این مناطق به طور تصادفی انتخاب می‌شوند و سپس تمام افراد مورد نظر در مناطق انتخاب شده مورد مطالعه قرار می‌گیرند.

نمونه‌گیری چند مرحله‌ای[2]

در بسیاری از پژوهشها چند روش نمونه‌گیری بصورت مرحله به مرحله مورد استفاده قرار می‌گیرد. مثلاً ابتدا مناطق جغرافیایی مورد مطالعه از طریق نمونه‌گیری تصادفی خوشه‌ای انتخاب شده و سپس در هر منطقه افراد به گروههای هموژن (مثل گروههای سنی) تقسیم می‌شوند و از هر گروه تعداد مورد نظر از طریق نمونه‌گیری تصادفی ساده برای مطالعه انتخاب می‌گردند.

روشهای نمونه‌گیری غیراحتمالی

در این روش تمام افراد برای انتخاب شدن در نمونه شانس ندارند. ولی این مطلب الزاماً بدین معنا نیست که نمونه‌ای که از روشهای غیراحتمالی حاصل می‌شود نماینده خوبی از جمعیت مطالعه نیست، با این حال مطمئناً تعیین میزان نماینده بودن چنین نمونه‌هایی مشکل‌تر از روشهای احتمالی است. بطور کلی پژوهشگران عموماً روشهای احتمالی یا تصادفی را به روشهای غیراحتمالی ترجیح می‌دهند، ولی در پژوهشهای علوم پزشکی ممکن است شرایطی وجود داشته باشد که نمونه‌گیری تصادفی در آنها میسر نباشد.

1- Cluster (area) random sampling
2- Multistage sampling

روشهای نمونه‌گیری غیراحتمالی به دو دسته کلی اتفاقی[1] و هدفمند[2] تقسیم می‌شوند.

نمونه‌گیری اتفاقی[3]

یکی از روشهای بسیار شایع برای نمونه‌گیری است. در این روش پژوهشگر از واحدهایی که در اطراف خود وجود دارد یا به آسانی در دسترس هستند، برای مطالعه استفاده می‌کند؛ مثلاً پژوهشگری که از بیماران بستری در بخش برای مطالعه استفاده می‌کند.

استفاده از افراد داوطلب برای مطالعات نیز جزء این روش نمونه‌گیری محسوب می‌شود. در این روش علیرغم سادگی اجرا هیچ شواهدی به نفع نماینده بودن نمونه گرفته شده وجود ندارد.

نمونه‌گیری هدفمند[4]

در این نوع نمونه‌گیری پژوهشگر نمونه خود را با قصد و هدف مشخص انتخاب می‌کند. در این شرایط پژوهشگر گروههای تعریف شده خاصی در ذهن دارد که در نمونه‌گیری بدنبال آنها می‌گردد.

علاوه بر روش نمونه‌گیری یکی از مسایل بسیار مهمی که در نمونه‌گیری مطرح می‌شود محاسبه حجم نمونه[5] و خطای نمونه‌گیری[6] است. حجم نمونه علاوه بر آنکه در نماینده بودن نمونه و در نتیجه در روایی خارجی مطالعه نقش دارد، نقش مهمی در توان[7] و دقت آماری[8] مطالعه نیز ایفا می‌کند، از این جهت در بخش تحلیل به این جنبه از نمونه‌گیری بیشتر پرداخته می‌شود.

نکته مهمی که لازم است به آن توجه داشته باشید این است که همیشه باید بین نمونه‌گیری تصادفی[9] و تخصیص تصادفی افراد

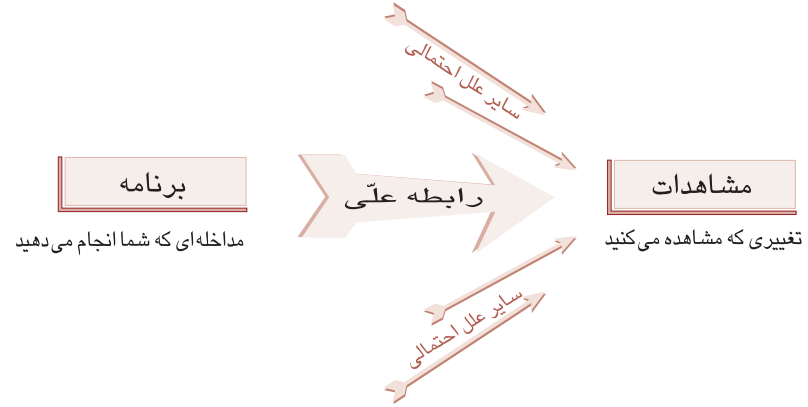
1- Accidental
2- Purposive
3- Incidental, Convenience, Haphazard or accidental
4- Purposive sampling
5- Sample size
6- Sampling error
7- Statistical power
8- Statistical precision
9- Random selection or sampling

به گروههای مقایسه[1] تمایز قائل شد. آنچه در این قسمت توضیح داده شد مربوط به نمونه‌گیری تصادفی بود. این نوع نمونه‌گیری ندرتاً در کارآزمایی‌های تصادفی شده بالینی به کار می‌رود. در تخصیص تصادفی، پس از نمونه‌گیری، واحدهای نمونه را با روشهای تصادفی مختلف به گروههای کنترل و مداخله تخصیص می‌دهیم. نحوه تخصیص افراد به گروههای مقایسه جزء مفاهیمی است که در بخش طرحهای مطالعه و روایی داخلی مطرح می‌شود و در آن قسمت مورد بررسی قرار خواهد گرفت.

طرح پژوهش[2]

طرح پژوهش در حقیقت نحوه ساختاربندی اجزاء مطالعه است. این ساختار نشان می‌دهد که چگونه اجزای مهم پژوهش (نمونه و گروهها، سنجش‌ها، مداخلات، روشهای تخصیص و ...) در کنار هم قرار می‌گیرند تا به سؤال پژوهش پاسخ داده شود. در حقیقت طرح مطالعه به منزله چسبی است که این اجزا را به هم متصل می‌کند. همانطور که گفته شد در مورد هر کدام از انواع پژوهش طرح مطالعه باید به گونه‌ای باشد که توان پاسخگویی به سؤال مورد نظر را داشته باشد. مثلاً وقتی بدنبال پیدا کردن رابطه علّی هستیم استفاده از طرحهای تجربی[3] در الویت قرار دارد. مفهومی که در اینجا مطرح می‌شود روایی داخلی[4] است. روایی داخلی از دیدگاه روش‌شناسی پژوهش، بیان می‌کند تا چه حد استنباط ما درباره رابطه علّی به واقعیت نزدیک است بنابراین روایی داخلی اصولاً در مواردی مطرح می‌شود که به دنبال یک رابطه علّی هستیم. توجه داشته باشید که رابطه علّی صرفاً به معنای «اتیولوژی» یا علت بیماریها نیست، بلکه وقتی اثر یک مداخله را بر روی یک پیامد خاص بررسی می‌کنیم نیز در حقیقت بدنبال رابطه علّی هستیم. سؤال کلیدی که در رابطه علّی مطرح می‌شود این است که آیا تغییرات مشاهده شده قابل نسبت دادن به مداخله انجام شده (اعم از رژیم‌درمانی، غربالگری، آموزش بیماران و ...) هست یا به سایر علل محتمل مربوط می‌شود؟ (شکل ۴)

1- Random assignment or allocation, randomization
2- Study design
3- Experimental
4- Internal Validity



شکل ۴

اهمیت روایی داخلی در مطالعات مربوط به حدی است که در بعضی منابع طبابت مبتنی بر شواهد، روایی کلی پژوهش معادل با روایی داخلی آن در نظر گرفته شده است و عواملی مثل دقت آماری، روایی خارجی و اهمیت بالینی نسبت به آن در درجه دوم اهمیت قرار می‌گیرند.

در این قسمت به توضیح مختصری درباره هر یک از طرحهای مربوط به راهبردهای مختلف می‌پردازیم.

طرح‌های مداخله‌ای

۱ طرح‌های تجربی[1]

طرحهایی که در طبقه‌بندی جزء راهبرد تجربی قرار می‌گیرند و به طور شایعتر برای پاسخ سؤالات بالینی در دسترس هستند، موارد زیر را شامل می‌شوند:

1- Experimental designs

- کارآزمایی‌های تصادفی شده شاهددار (RCT)
- کارآزمایی‌های شبه تصادفی شده شاهددار
- کارآزمایی‌های تصادفی شده متقاطع
- طرح عاملی

کارآزمایی تصادفی شده شاهددار[1]

در این مطالعه تجربی پژوهشگران افراد حائز شرایط را به طور تصادفی در گروههای مطالعه وارد می‌کنند.

همانطور که گفته شد انتخاب افراد حائز شرایط[2] از جمعیت مطالعه ندرتاً به طور تصادفی انجام می‌گیرد. این انتخاب معمولاً براساس معیارهای ورود و خروج خاصی است که پژوهشگر به منظور کنترل بیشتر، آنها را اعمال می‌کند. از آنجاییکه کنترل بیشتر امکان نسبت دادن تغییر ایجاد شده (اثر) در پیامدهای مورد نظر را به مداخله انجام شده (علت) بیشتر فراهم می‌کند بنابراین روایی داخلی پژوهش بالاتر خواهد بود. این در حالیست که چنین افزایشی در روایی داخلی که به قیمت اعمال کنترل بیشتر از طریق معیارهای ورود و خروج حاصل شده است، قابلیت تعمیم دادن نتایج پژوهش را به جمعیت مطالعه کاهش داده و از این طریق از روایی خارجی مطالعه می‌کاهد (معامله بین روایییها). برای مثال فرض کنید شما افراد بالای ۴۵ سال را که مبتلا به دیابت تیپ ۲ هستند، قند خون آنها صرفاً با انسولین کنترل می‌شود و سابقه بیماری دیگری ندارند، در مطالعه خود وارد کرده‌اید، با این کار ضمن اینکه کنترل شما بر عوامل تعیین کننده پیش آگهی در نمونه بیشتر شده است ولی قابلیت تعمیم دادن نتایج مطالعه‌ای با این نمونه به جمعیت افراد

1- Randomized controlled (clinical) trial(RCT)

2- Eligible

مبتلا به دیابت تیپ ۲ کاهش یافته است. در کارآزمایی‌های تصادفی شدهٔ شاهددار بعد از تخصیص تصادفی، نتایج از طریق مقایسه پیامدها در گروههای مداخله و کنترل ارزیابی می‌شود.

کارآزمایی شبه تصادفی شده شاهددار[1]

تفاوت این کارآزمایی‌ها با مورد قبلی در روش تخصیص است. در تخصیص شبه تصادفی شده شانس افراد برای ورود به هر یک از گروههای مطالعه الزاماً برابر نیست.

در این روش تخصیص افراد با سازوکارهایی مثل تاریخ تولد، روز هفته (روز مراجعه)، شماره پرونده پزشکی (زوج و فرد بودن شماره)، ماه سال و یا براساس ترتیب مراجعه انجام می‌شود. بنابراین احتمال وجود تفاوت به صورت اولیه در گروهها بیشتر خواهد بود. البته باید به این نکته توجه داشت که تخصیص تصادفی نیز الزاماً گروههای کاملاً معادل[2] بدست نمی‌دهد ولی از این طریق احتمال وجود تفاوت بین گروههای مطالعه (غیر از مداخله مورد نظر) به کمترین حد خود می‌رسد.

کارآزمایی تصادفی شده متقاطع[3]

نوعی از کارآزمایی‌های تصادفی شده بالینی است که در آن دو یا چند مداخله با هم مقایسه می‌شوند؛ بصورتیکه شرکت‌کنندگان در پژوهش پس از پایان یک دوره مداخله، مداخله بعدی را تجربه می‌کنند. برای مثال فرض کنید می‌خواهیم اثر مداخله A و B را با هم مقایسه کنیم. نیمی از افراد به طور تصادفی این مداخلات را با توالی AB دریافت می‌کنند و نیم دیگر با ترتیب BA (این کار

1- Quasi/Pseudo-Randomized Trial
2- Equivalent
3- Randomised Crossover Trial

معمولاً به صورت کور انجام می‌شود) سپس نتایج گروهها با هم مقایسه می‌گردد. مشکلی که اختصاصاً این طرح را تهدید می‌کند این است که اثر مداخله اول ممکن است تا زمان دریافت مداخله دوم به طول بیانجامد. این طرح مطالعه به پژوهشگر امکان می‌دهد تا ضمن مقایسه اثر مداخلهٔ آزمایشی با یک مداخلهٔ دیگر یا دارونما، توالی از مداخلات را که اثربخشی بیشتری دارند مشخص کند و اثر طولانی مدت مداخلات را تعیین نماید.

طرح عاملی[1]

بیشتر کارآزمایی‌ها معمولاً یک عامل را برای مقایسه در نظر می‌گیرند بدین معنا که در اکثر مطالعات اثر یک مداخله با یک یا چند مداخله جایگزین و یا دارونما مقایسه می‌شود. در حالیکه در طرح عاملی هم زمان اثر چند مداخله در سطوح مختلف با هم مقایسه می‌شود. ساده‌ترین شکل طرح عاملی فرم ۲×۲ است. در این فرم شرکت‌کنندگان به طور تصادفی وارد یکی از ۴ ترکیب ممکن می‌شوند. به عنوان مثال در یک طرح عاملی ۲×۲ از یک کارآزمایی بالینی برای سنجش اثر نیکوتین در ترک سیگار شرکت کنندگان به یکی از گروههای «جایگزینی نیکوتین» به تنهایی، مشاوره تنها، هر دو مداخله و یا هیچکدام وارد می‌شوند بدین ترتیب علاوه بر امکان ارزیابی اثر مستقل هر کدام از مداخلات روی ترک سیگار، ارزیابی اثر ترکیبی آنها نیز ممکن می‌شود. بنابراین در طرح عاملی مطالعه اثر چند مداخله به طور همزمان و نیز نتیجهٔ حاصل از اثر متقابل[2] این مداخلات امکانپذیر است. از این رو طرح عاملی یکی از طرحهای پژوهشی بسیار قوی از نظر روایی داخلی محسوب می‌شود. در طرح عاملی در حقیقت به جای انجام چند کارآزمایی بالینی با یکبار مطالعه می‌توانیم به همان نتایج برسیم. لازم به ذکر است که هر کدام از مداخلات به عنوان یک عامل[3] و نحوه ارائه مداخله به عنوان سطح[4] معروف است و مطالعه براساس اینکه چند مداخله (عامل) داشته باشد و هر مداخله چند سطح داشته باشد به فرمهای گوناگون (۲×۲، ۲×۳، ۳×۳ و ...) قابل اجراست.

1- Factorial design
2- Interaction
3- Factor
4- Level

۲| طرح‌های شبه تجربی[1]

طرحهایی که ذیل راهبرد شبه تجربی قرار می‌گیرند، مشابه طرحهای تجربی هستند با این تفاوت که گروههای کنترل و مداخله از طریق روشهای غیرتصادفی شکل می‌گیرند. در حقیقت تمام طرحهای مربوط به راهبرد تجربی می‌توانند یک طرح متناظر در راهبرد شبه تجربی داشته باشند.

طرح متقاطع غیرتصادفی شده و سری‌های زمانی منقطع دو نمونه از طرحهای شبه تجربی است.

طرح متقاطع غیرتصادفی شده[2]

مشابه فرم تصادفی شده آن است با این تفاوت که ورود افراد شرکت کننده به هر کدام از گروهها و نیز ترتیب ارائه مداخلات با روشهای غیرتصادفی تعیین می‌شود.

سری‌های زمانی منقطع[3]

در این طرح اثر یک مداخله از طریق مقایسه الگوی نتایج بدست آمده در سنجش‌های متعدد، قبل و بعد از دریافت مداخله در یک گروه از بیماران ارزیابی می‌شود. تعداد سنجش‌ها در این طرح اهمیت دارد ولی قانون کلی واحدی راجع به آن وجود ندارد. این طرح با وجود عدم استفاده از گروه کنترل، به علت استفاده از سنجش‌های متعدد جزء طرحهای شبه تجربی در نظر گرفته می‌شود. با اضافه کردن یک گروه کنترل که در زمانهای مشابه مورد سنجش قرار می‌گیرد و در ضمن مداخله مورد نظر را دریافت نمی‌کند می‌توان این طرح را از نظر روایی داخلی تقویت نمود.

1- Quasi-experimental
2- nonrandomized cross-over design
3- Interrupted time series

۳ طرح‌های غیرتجربی[1]

انجام مطالعاتی که از طرحهای غیرتجربی بهره می‌گیرند، معمولاً آسانتر و نیازمند امکانات کمتری است. به عنوان مثال یک طرح غیرتجربی می‌تواند به این صورت باشد:

گروهی از افراد تحت یک مداخله خاص قرار گرفته و سپس نتایج این مداخله، مطالعه و گزارش می‌شود. توجه داشته باشید که در این طرح گروه کنترل جداگانه‌ای وجود ندارد چنین مطالعاتی عموماً به دو گونه زیر انجام می‌شود:

- سنجش بعد از مداخله[2] : در این طرح، پیامد مورد نظر در افراد شرکت کننده صرفاً بعد از انجام مداخله ارزیابی گزارش می‌شود، بنابراین امکان هیچ گونه مقایسه‌ای در این روش وجود ندارد.

- سنجش قبل و بعد از مداخله[3] : در این روش پیامد مورد نظر قبل و بعد از انجام مداخله سنجیده می‌شود و بدین ترتیب امکان مقایسه پیامد افراد قبل و بعد از مداخله فراهم خواهد بود.

کارآزمایی تصادفی شده تک نفره[4]

این نوع مطالعه در حقیقت نوعی کارآزمایی است که فقط یک شرکت کننده دارد. کاربرد آن در طبابت، تعیین بهترین درمان برای یک بیمار خاص است. روشهای بسیاری برای انجام این مطالعه وجود دارد که یکی از آنها بدین ترتیب است:

۱ پزشک و بیمار برای آزمایش یک مداخله (یک درمان آزمایشی) از نظر توانایی آن مداخله برای رسیدن به اهداف

1- Non-experimental design
2- Post - test
3- Pre- test / post-test or before- and- after
4- N of one Randomized Trial

درمانی[1] به توافق می‌رسند.

۲ سپس بیمار تحت چند رژیم درمانی قرار می‌گیرد بدین صورت که برای هر رژیم درمانی بیمار ابتدا تحت یک دوره درمان آزمایشی قرار گرفته و سپس تحت دوره‌ای از درمانهای جایگزین و یا دارونما قرار می‌گیرد. ترتیب این دو دوره در هر بار درمان بطور کاملاً تصادفی تعیین می‌شود.

۳ در صورت امکان پزشک و بیمار نسبت به این ترتیب بی‌خبر (کور) هستند.

۴ پزشک اهداف درمانی را پایش می‌کند تا میزان تأثیر مداخله در حال انجام تعیین می‌شود.

۵ مطالعه این رژیمهای درمانی تا جایی ادامه می‌یابد که پزشک و بیمار نسبت به اثربخش بودن، مضر بودن یا بی‌تأثیر بودن درمان متقاعد شوند. چنین مطالعه‌ای معمولاً نیاز به سه بار ارزیابی در دوره‌های درمان آزمایشی و دارونما دارد.

با وجود اینکه این نوع مطالعه تحت عنوان کارآزمایی تصادفی شده معروف شده است ولی به علت اینکه صرفاً یک شرکت کننده در این مطالعه وجود دارد و عملاً گروه کنترلی شکل نمی‌گیرد نمی‌توان آنرا نوعی مطالعه تجربی به حساب آورد.

از آنجاییکه در مطالعات غیرتجربی اثر مداخلات در یک فرد یا گروه قبل و بعد از انجام مداخله مورد ارزیابی قرار گیرد، بنابراین نسبت دادن پیامد مشاهده شده به مداخله انجام شده در این نوع مطالعات با مشکلات بیشتری همراه است.

1- Treatment targets

طرحهای مشاهده‌ای

مطالعات هم‌گروهی[1]

در این نوع مطالعه، گروه تعریف شده‌ای از جمعیت برای مدت مشخصی پیگیری می‌شود. سپس پیامد افراد موجود در زیر مجموعه‌های این گروه با هم مقایسه می‌شود. به عنوان مثال برای ارزیابی تفاوت پیامدهای افرادی که در معرض یک مداخله خاص یا عامل مورد نظر بوده‌اند و یا در معرض سطوح مختلفی از آن مداخله بوده‌اند و یا بطور کلی در معرض نبوده‌اند از این روش‌استفاده می‌شود، در مطالعات همگروهی افراد برای شرکت در گروههای کنترل و مطالعه انتخاب می‌شوند. در این زمان پژوهشگران از بیمار بودن یا احتمال رفتن به سمت بیماری در شرکت‌کنندگان با خبر نیستند. یک همگروه[2] گروهی از افراد است که یک تجربه مشترک داشته‌اند. در این مطالعه یک همگروه که واجد شرایط خاصی است با همگروه دیگر که این شرایط را ندارد مقایسه می‌شود. سپس میزان شیوع پیامد مورد نظر (معمولاً ابتلاء به بیماری) در دو گروه بدست می‌آید. فرض کنید می‌خواهیم رابطه مصرف قرصهای ضدبارداری و سکته مغزی[3] را مطالعه کنیم. در یک مطالعه همگروهی پژوهشگر اینگونه عمل می‌کند:

پس از تعریف جمعیت مطالعه و انتخاب یک نمونه مناسب افراد نمونه را به دو گروه مطالعه و کنترل تقسیم می‌کند بطوریکه گروه مطالعه از خانمهایی که قرص ضدبارداری مصرف می‌کنند تشکیل می‌شود و گروه کنترل از خانمهایی که شرایط کاملاً مشابه دارند ولی از قرص ضدبارداری استفاده نمی‌کنند. توجه داشته باشید که مصرف قرصهای ضدبارداری بدون مداخله فعال پژوهشگران در حال اتفاق افتادن است. سپس پژوهشگر گروه مطالعه و گروه کنترل را برای مدتی پیگیری کرده و مشخص می‌کند از هر گروه چه تعداد به سکته مغزی مبتلا شده‌اند.

1- Longitudinal/ Prospective / Concurrent Cohort
2- COHORT
3- Stroke

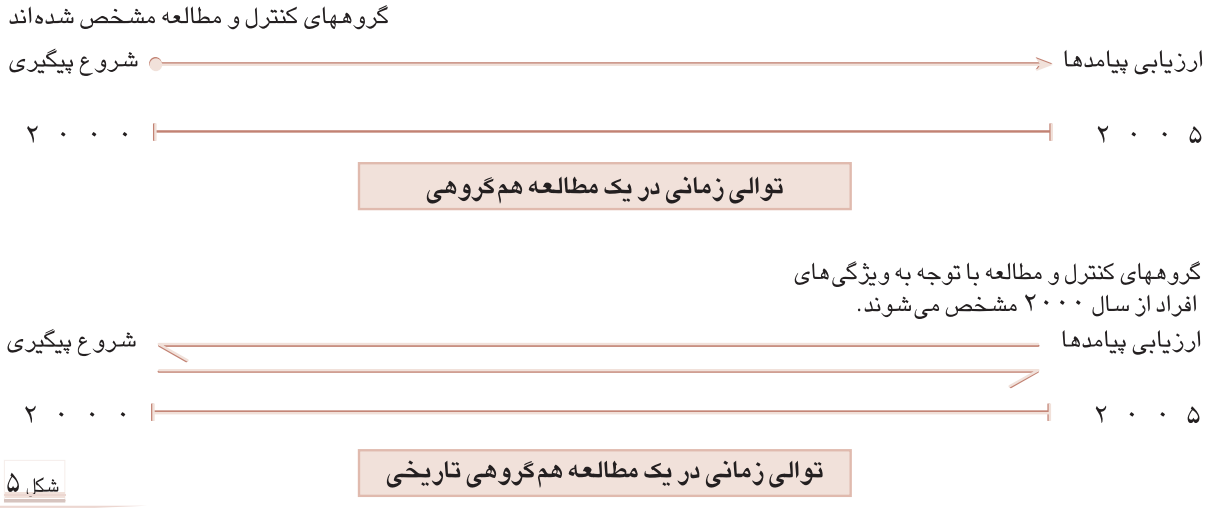
مطالعات همگروهی تاریخی[1]

همانطور که گفته شد در زمان شروع یک مطالعه همگروهی پژوهشگران از میزان وقوع پیامدهای مورد نظر در افراد تحت مطالعه آگاهی ندارند و با پیگیری آنها در طول زمان میزان بروز پیامدها تعیین می‌شود. با توجه به رشد روزافزون ثبت و ضبط دقیق داده‌های بیماران و مداخلاتی که روی آنها انجام شده است، امروزه امکان انجام نوع دومی از مطالعات همگروهی به نام مطالعات همگروهی تاریخی فراهم شده است. در این نوع مطالعه اطلاعات مربوط به مداخله‌ای که بیمار در گذشته دریافت کرده است از پایگاههای داده‌ای مراکز درمانی قابل دریافت است. بدین ترتیب در زمانیکه مطالعه انجام می‌شود نتایج و پیامدها قبلاً در پایگاههای داده‌ای ثبت شده است. انجام چنین مطالعه‌ای می‌تواند اینگونه باشد:

در فوریه ۲۰۰۵ پژوهشگران مطالعه‌ای را آغاز می‌کنند، آنها پایگاه داده‌ای[2] را برای یافتن کلیه بیمارانی که در سال ۲۰۰۰ برای درمان اوتیت میانی عود کننده تحت درمان جراحی یا درمان طبی قرار گرفته‌اند، جستجو کردند. هنگامیکه تعداد این افراد مشخص شد، پژوهشگران آنها را به دو گروه کنترل (درمان طبی) و مطالعه (درمان جراحی) تقسیم کردند. سپس پژوهشگران مجدداً پایگاه داده‌ای را برای تعیین پیامدهای مورد نظر تا ژانویه ۲۰۰۵ جستجو کردند (شکل ۵). از آنجاییکه تقسیم بندی افراد به دو گروه مطالعه و کنترل در این نوع مطالعه نیز پیش از آگاهی از پیامد افراد صورت می‌گیرد، باز هم به آن مطالعهٔ همگروهی گفته می‌شود. توجه داشته باشید پژوهشگران تنها زمانی مجاز هستند از پیامدهای افراد آگاهی یابند که فرایند تقسیم‌بندی به گروههای کنترل و مطالعه قبلاً صورت گرفته باشد، با وجود اینکه پیامدها قبل از شروع مطالعه به وقوع پیوسته‌اند. هر دو نوع مطالعه می‌توانند در بررسی علت بیماریها[3] و سود

1- Non Concurrent Cohort Studies,Historical / Database Cohort Research
2- Database
3- Etiology

و زیان اقدامات درمانی مفید باشند.(شکل ۵)



شکل ۵

گروه کنترل پیشینه‌ای[1]

در بعضی مطالعات، گروه مطالعه به صورت آینده‌نگر تشکیل شده و در معرض یک مداخله یا عامل خاص قرار می‌گیرد سپس پیامدها بدست آمده با گروه کنترلی که قبلاً در همان مرکز تحت مطالعه قرار گرفته است مقایسه می‌شود. به این گروه که داده‌های مربوط به آنها پیش از گروه مطالعه جمع‌آوری شده است در اصطلاح گروه کنترل پیشینه‌ای می‌گویند. البته به خاطر تغییراتی که در خطر بیماریها، پیش‌آگهی، مراقبتهای سلامت و ... در طول زمان بوجود می‌آید امکان ایجاد تورش[2] در مطالعاتی که از گروههای کنترل پیشینه‌ای استفاده می‌کنند بیشتر است. از طرف دیگر در این گونه مطالعات ممکن است درمان جدید به طور انتخابی به بیمارانی داده شود که احتمال موفقیت درمان در آنها بیشتر است.

1- Historical Control
2- bias

طرح مورد ـ شاهدی[1]

ویژگی منحصر به فرد در مطالعات مورد ـ شاهدی این است که چنین مطالعاتی پس از به وقوع پیوستن پیامد مورد نظر (مثلاً مبتلا شدن عده‌ای به بیماری)، شروع می‌شوند. در این مطالعات پژوهشگر با دیدی گذشته‌نگر[2] ویژگیهای افراد را پیش از شروع بیماری تعیین می‌کند. در مطالعه مورد شاهدی، موردها افرادی هستند که قبلاً به بیماری یا پیامد مورد نظر مبتلا شده‌اند و شاهدها افرادی هستند مشابه آنها که غیر مبتلا به بیماری مورد نظر هستند.

رابطه بین یک ویژگی (مداخله یا در معرض یک عامل خطر بودن) با پیامد مورد نظر از طریق مقایسه وقوع و یا میزان آن ویژگی در گروه مورد و گروه شاهد بررسی می‌شود. به عنوان مثال برای تعیین اثر تالیدومید بر روی نقایص بعد از تولد گروهی از نوزادان با نقایص بعد از تولد (موردها) را می‌توان با گروه دیگری که فاقد این نواقص هستند (شاهدها) مقایسه کرد. این گروهها از نظر میزان مصرف تالیدومید توسط مادرانشان، مورد ارزیابی قرار می‌گیرند.

این مطالعات بعضاً با واژه گذشته‌نگر نیز توصیف می‌شوند چرا که انجام آنها مستلزم بازگشت به گذشته و بررسی سابقهٔ پزشکی شرکت‌کنندگان است.

طرح مقطعی[3]

در این طرح رابطه بین یک بیماری (یا سایر ویژگیهای مشابه) با متغیرهای دیگری که به طور همزمان در جمعیت تعریف شده وجود دارند، مطالعه می‌شود. در این مطالعه افرادی که واجد پیامد مورد نظر (بیماری) هستند و یا مبتلا به آن نیستند تعیین شده و همزمان وجود ویژگی مورد نظر (مثلاً در معرض مداخله یا عامل خطری بودن) در هر گروه بررسی می‌شود. به عبارت دیگر

1- Case control design
2- Retrospective
3- Cross-sectional design

متغیرهای مستقل و وابسته در یک مقطع زمانی[1] سنجیده می‌شوند. این مطالعات زمانی سودمند هستند که پژوهشگر انتظار دارد اثر مداخله در طول زمان بدون تغییر باقی بماند و یا زمان بین مواجهه و بیماری بسیار کوتاه باشد. در بیشتر موارد این طرح نوع خاصی از طرح مورد – شاهدی محسوب می‌شود.

به طور کلی طرح‌های مشابه در مقایسه با طرحهای تجربی با مشکلات بالقوه‌ای دست به گریبان هستند.

در مورد مطالعاتی که از گروه کنترل استفاده می‌کنند، غیرقابل مقایسه بودن (معادل نبودن) گروهها از نظر عوامل موٌثر بر پیش‌آگهی که به نحوه انتخاب افراد برای شرکت در مطالعه برمی‌گردد، مهمترین و اساسی‌ترین مشکل است.

احتمال دریافت مداخلات و درمانهایی غیر از مداخله مورد نظر در چنین مطالعاتی بیشتر است.

استفاده از روشهای مختلف برای سنجش پیامد در گروههای مطالعه و کنترل در طرح‌های مشاهده‌ای بسیار شایعتر است.

علاوه بر آن استفاده از گروههای کنترل پیشینه‌ای در بعضی از این طرحها مشکلات بیشتری ایجاد کرده است.

1- Cross-Section

طرح‌های توصیفی [1]

گزارش موردی [2]

در این طرح تجربهٔ یک بیمار توصیف می‌شود. این نوع مطالعه بیشتر به بررسی موارد نادر و تظاهرات نامحتمل بیماریهامی‌پردازد.

بررسی موردها[3]

این طرح مشابه گزارش موردی است با این تفاوت که در آن تجربهٔ بیش از یک بیمار گزارش می‌شود. در این طرح تجربهٔ بیماران معمولاً در یک دورهٔ زمانی نسبتاً کوتاه گزارش می‌شود. شایعترین مقالاتی که از طریق نشریات پزشکی و پایگاه‌های داده‌ای قابل دستیابی است، گزارش موردی و بررسی موردها می‌باشد. بیماریها و یا اثرات جانبی جدید و نادر معمولاً اولین بار در قالب گزارشات موردی و بررسی موردها ارائه می‌شوند.

مطالعات پیمایشی مقطعی[4]

این مطالعات معمولاً برای سنجش شیوع بیماریها مورد استفاده قرار می‌گیرند و انجام آنها نسبتاً ساده و کم هزینه است. در این مطالعات توصیف عواملی که به طور ذاتی در جمعیت وجود دارد (مانند نژاد، گروه خونی، وضعیت اقتصادی اجتماعی و ...) و ممکن است با شیوع بیماریها رابطه داشته باشند، امکانپذیر است.یکی از کاربردهای مهم این نوع مطالعات ارزیابی نیازهای سلامت در جمعیت است.

1- Descriptive designs
2- Case history, single case report or case study
3- Case series
4- Cross-Sectional survey studies

جمع‌آوری داده‌ها [1]

اکنون که به طور خلاصه با انواع طرحهای مطالعه،کاربرد، ویژگیها و محدودیت‌های هر طرح آشنا شدید، بنابر آنچه در فرایند کلّی پژوهش گفته شد نوبت جمع‌آوری داده‌هاست. روشهای متعددی برای جمع‌آوری داده‌ها وجود دارد ولی انتخاب روش مناسب بر مبنای اهداف، طرح و منابع پژوهش صورت می‌گیرد. در این بین، پرسشنامه، مصاحبه و روشهای مشاهده‌ای از ابزارهای اصلی برای جمع‌آوری داده‌ها هستند.

پرسشنامه‌ها به طور شایع در طرحهای پیمایشی [2] استفاده می‌شوند. انواع مختلفی از پرسشنامه‌ها وجود دارد که طیفی از پرسشنامه‌های ساختارمند و مقیاس‌های استاندارد شده تا اشکال باز و بدون ساختار مشخص را شامل می‌شود.

مصاحبه [3] در حقیقت حاصل تعامل بین فرد مصاحبه شونده و مصاحبه‌گر است. از آنجایی که مصاحبه‌ها نیازمند وجود یک فرد مصاحبه‌گر و صرف زمان بیشتر است، در این روش هزینه و وقت بیشتری برای جمع‌آوری داده‌ها صرف می‌شود. از طرف دیگر خود مصاحبه‌گر ممکن است در پاسخهای مصاحبه شوندگان تأثیر بگذارد. طرحهای پژوهش کیفی معمولاً از تکنیکهای قویتری برای مصاحبه بهره می‌گیرند بطوریکه مصاحبه مثلاً در یک محیط کاملاً طبیعی (خانه، بیمارستان و ...) که پاسخ دهندگان در آن زندگی می‌کنند و یا درمان می‌شوند، انجام می‌گیرد.

روشهای مشاهده‌ای [4] نیز یکی از روش‌های شایع برای جمع‌آوری داده‌ها خصوصاً در مطالعات پزشکی و بالینی است. این روش‌ها نیز دامنه وسیعی دارند و طیفی را از پروتکل‌های مشاهده‌ای شدیداً ساختارمند [5] که در آنها رفتارها و علائم بالینی که باید ثبت شوند بدقت بیان شده است، تا ثبت بدون ساختار تجارب مشاهده‌گر مثلاً در مطالعات کیفی، در بر می‌گیرد.

بسته به نحوه ثبت و تحلیل داده‌ها، مصاحبه‌ها و مشاهدات در پژوهش‌های کمی و کیفی مورد استفاده قرار می‌گیرند. با این حال

1- Data Collection
2- Survey designs
3- Interview
4- Observational methods
5- Highly structured

استفاده از ابزارهایی که داده‌های عددی تولید می‌کنند برای پژوهش‌های کمی بسیار مناسب است. طیف وسیعی از ابزارهای سنجش استاندارد شده[1] برای سنجش متغیرهای زیست‌شناختی و بالینی در دسترس هستند.

به طور کلی باید بتوان از پایایی[2] و روایی[3] هر روشی که برای جمع‌آوری داده‌ها استفاده می‌شود اطمینان حاصل کرد. در غیر اینصورت خطاهای سنجش ناشی از راهبردهای ضعیف و ناروا[4] برای جمع‌آوری اطلاعات، مانع از رسیدن به یک استنباط صحیح در پژوهش خواهد شد. در فصل بعد در مورد تورش سنجش و انوع آن بحث خواهد شد.

پرداختن به جزئیات روش‌های جمع‌آوری داده در پژوهش با توجه به اهداف این فصل لازم به نظر نمی‌رسد برای اطلاعات بیشتر در این زمینه می‌توانید به کتابهای روش‌شناسی پژوهش مراجعه کنید.

تحلیل[5]

پس از جمع‌آوری داده‌ها ، باید بتوانیم با تحلیل آنها از پژوهش خود نتیجه‌گیری کنیم. بطور کلی تحلیل در پژوهش دو بخش را شامل می‌شود: آمار توصیفی[6] و آمار استنباطی[7] .

آمار توصیفی

بسته به پیچیدگی طرح پژوهش و تعداد سنجش‌های انجام شده برای جمع‌آوری داده‌ها در هر پژوهش نهایتاً به تعدادی عدد و رقم خواهیم رسید که اصطلاحاً به آن دادهٔ خام[8] گفته می‌شود. سازماندهی و خلاصه‌سازی این داده‌ها اولین مرحله در فراوری داده‌های خام برای استفاده در تحلیل است، مرحله بعدی جدول‌بندی و تبدیل داده‌ها به نمودارهای گرافیکی است.

1- Standardized measurement instruments
2- Reliability
3- Accuracy or Validity
4- Invalid
5- Analysis
6- Descriptive statistics
7- Inferential statistics
8- Raw data

هنگامیکه داده‌ها ساختار مناسبی پیدا کردند می‌توانید از آنها برای محاسبه آماره‌های[1] مختلف استفاده کنید. بنابراین یکی دیگر از کاربردهای آمار توصیفی جمع‌بندی داده‌ها به صورت آماره‌هایی است که به طور معمول مورد استفاده قرار می‌گیرند. این آماره‌ها شامل سنجه‌های تمایل به مرکز[2] (میانگین، میانه و مد) و سنجه‌های پراکنش[3] (دامنه تغییرات، انحراف معیار و...) است. استفاده از این آماره‌ها ما را قادر می‌سازد تا اطلاعات مربوط به یافته‌های پژوهش را به مخاطبین انتقال بدهیم. آماره‌هایی مثل میانگین و انحراف معیار برای محاسبه نمره‌های استاندارد[4] نیز بکار می‌روند. نمره‌های استاندارد برای بیان نمره‌های خاص مربوط به جمعیت استفاده می‌شوند. در ادامه بحث، توضیح بیشتری در این زمینه ارائه خواهد شد.

آمار استنباطی

همانطور در فصل‌های آینده بیشتر خواهید آموخت دقت آماری[5] در کنار سطح شواهد (طرح مطالعه) و کیفیت شواهد به عنوان یکی از موضوعات اصلی و مهم در ارزیابی نقادانه قدرت شواهد مطرح است. پی بردن به دقت آماری مطالعات مستلزم آشنایی با اصول و تعاریف پایه آمار استنباطی است. لذا به نظر می‌رسد تأمل بیشتری در این زمینه لازم باشد.

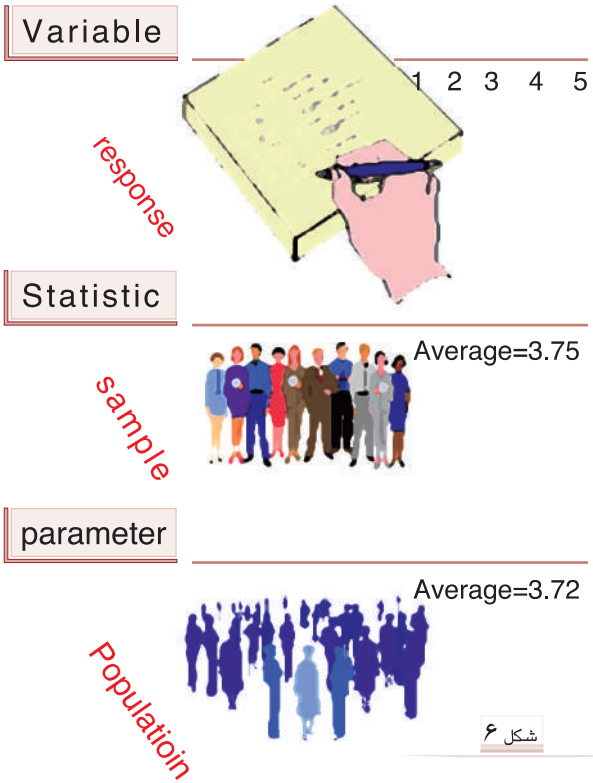
تحلیل آماری داده‌ها یک مرحله ضروری در فرایند پژوهش‌های کمّی است. برای اینکه بتوانیم تصمیم بگیریم آیا داده‌های بدست آمده از نمونه، تفاوت‌ها و الگوهای واقعی جمعیت را نشان می‌دهد، نیازمند اجرای اصول آمار استنباطی هستیم. این امر در مورد هر پژوهشی که روی یک نمونه از جمعیت انجام می‌شود صادق است. استنباط در مورد جمعیت از روی نمونه گرفته شده همیشه مستلزم در نظر گرفتن نوعی خطا به نام خطای نمونه‌گیری است.

1- Statistic
2- Measures of central tendency
3- Measures of dispersal
4- Standard scores
5- Statistical precision

همانطور که در ادامه خواهید دید خطای نمونه‌گیری مسوول تفاوت موجود بین ارزش نمونه با ارزش واقعی جمعیت است. در شرایطی که پژوهشگران معمولاً از پارامترهای جمعیت آگاه نیستند، آمار استنباطی می‌تواند تخمین خوبی از خطای نمونه‌گیری ارائه دهد.

در این قسمت ابتدا با تعاریف و نحوه استفاده از قوانین احتمالات برای رسیدن به توزیع نمونه‌گیری آشنا خواهید شد و سپس از توزیع نمونه‌گیری برای محاسبه بازه اطمینان[1] و انجام آزمون فرضیه[2] استفاده خواهید کرد.

هنگامی که از یک جمعیت نمونه‌گیری می‌کنید واحدهای موجود در نمونه (معمولاً انسانها)، هر کدام پاسخ جداگانه‌ای به پژوهش، می‌دهند. در اینجا منظور از پاسخ[3] ارزش خاصی است که هر کدام از واحدهای نمونه در سنجش ما کسب می‌کنند. در شکل ۶ می‌بینید که یک نفر از اعضای نمونه در حال پاسخگویی به پرسشنامه پژوهش است. پاسخ او عدد ۴ است. برای بیان‌پاسخهایی که از کل نمونه بدست آمده‌اند از یک آماره[4] استفاده می‌شود. انواع متعددی از آماره‌ها برای این منظور به کار می‌رود مثل میانگین، میانه[5]، مد و ..... (در این



شکل ۶

1- confidence interval
2- hypotheses testing
3- response
4- Statistic
5- median

شکل می‌بینیم که میانگین نمونه برابر ۳/۷۵ است). ولی قصد ما از نمونه بدست آوردن تخمینی از جمعیت است که نمونه از آن گرفته شده است.

اگر بتوانیم کل جمعیت را مورد بررسی قرار دهیم و یک ارزش مثل میانگین را برای آن محاسبه کنیم، دیگر به آن آماره اطلاق نمی‌شود بلکه عدد بدست آمده در اصطلاح یک پارامتر[1] از جمعیت نامیده می‌شود.

توزیع نمونه‌گیری[2]

حال ببینیم چگونه می‌توانیم از آماره نمونه خود به تخمینی از پارامتر جمعیت برسیم. یکی از مفاهیم پایه‌ای مورد نیاز مفهوم توزیع نمونه‌گیری است. برای درک بیشتر آن، نیاز به یک تجربه ذهنی دارید. فرض کنید به جای گرفتن یک نمونه واحد (مانند آنچه در تمام مطالعات انجام می‌دهید)، سه نمونه مستقل از همان جمعیت گرفته‌اید. سپس برای هر کدام از سه نمونه گرفته شده یک پاسخ خاص را جمع‌آوری کرده و از طریق آن، آماره خاص (مثلاً میانگین پاسخهای) مربوط به هر نمونه را محاسبه می کنید. با وجود اینکه هر سه نمونه از یک جمعیت واحد گرفته شده است، ولی انتظار ندارید که آماره مربوط به هر کدام دقیقاً با بقیه یکی باشد. (اختلاف اندک این آماره‌ها می‌تواند صرفاً به علت تصادف و شانس و نوسانات طبیعی انتخاب نمونه باشد) ولی به طور منطقی انتظار دارید که تمام سه نمونه بدست آمده تخمین‌های آماری مشابهی بدست دهند چرا که از یک جمعیت گرفته شده‌اند. حال فرض کنید از یک جمعیت خاص تعداد بیشماری نمونه گرفته‌اید و میانگین هر نمونه را محاسبه کرده‌اید. اگر این میانگین‌ها را به صورت یک نمودار هیستوگرام ترسیم کنید ملاحظه خواهید کرد که بیشتر مقادیر حول یک ارزش مرکزی تجمع یافته‌اند و تعداد کمتری از نمونه‌ها میانگین

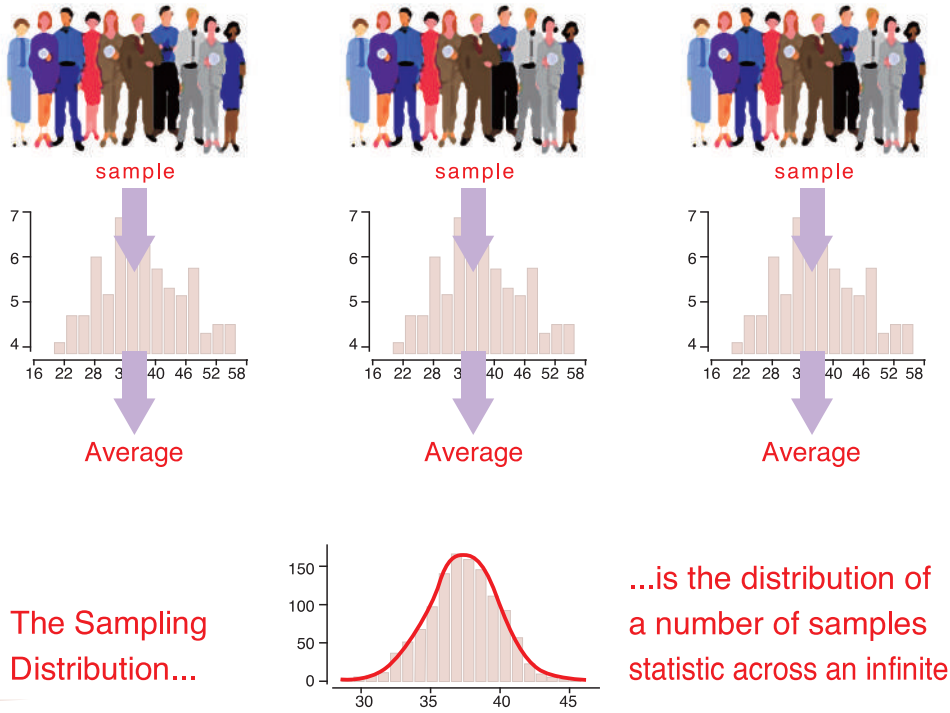
1- Parameter
2- Sampling distribution

بسیار بیشتر و یا بسیار کمتر از ارزش مرکزی دارند. به بیان دیگر نتایج بدست آمده به صورت یک نمودار زنگوله ای شکل[1] در خواهد آمد که در آمار نشانگر توزیع نرمال است. توزیع تعداد بیشماری از نمونه ها با حجم یکسان و مشابه نمونه مطالعه شما تحت عنوان "توزیع نمونه گیری" نامیده می شود. در دنیای واقعی هیچگاه بدنبال تولید یک توزیع نمونه گیری‌نیستیم. چرا که برای این کار باید تعداد بیشماری نمونه بگیریم و چنین کاری فعلاً عملی نیست! پس چرا اصلاً چنین مفهومی مطرح شده است؟ طرح چنین مفهومی از آن جهت اهمیت دارد که درک کنیم نمونه ای که ما گرفته ایم صرفاً یکی از تعداد بی شمار نمونه هاست که بالقوه می توان بدست آورد.

اگر این مفهوم را در ذهن داشته باشیم، می فهمیم وقتی آماره ای که از نمونه خود بدست آورده ایم محتملاً می تواند نزدیک به مرکز توزیع نمونه گیری باشد (چرا که اکثر نمونه ها در چنین جایی قرار دارند)، این احتمال نیز وجود دارد که صرفاً به علت پدیده شانس نمونه شما در یکی از دو انتهای نمودار واقع شده باشد. اگر میانگین توزیع نمونه گیری – میانگین میانگین های همه نمونه ها – را بدست آوریم این عدد به میانگین جمعیت (پارامتر مورد نظر) بسیار نزدیک خواهد بود بطوریکه میانگین توزیع نمونه گیری معادل پارامتر جمعیت در نظر گرفته می شود.

حال ببینیم انحراف معیار توزیع نمونه‌گیری چیست؟ انحراف معیار توزیع نمونه‌گیری بیان می‌کند نمونه‌ها با چه تفاوتی توزیع شده‌اند. در آمار به این مفهوم خطای معیار[2] اطلاق می‌شود. انحراف معیار[3] پراکندگی نمره‌ها پیرامون میانگین در یک نمونه خاص است در حالیکه خطای معیار بیانگر پراکندگی میانگین‌ها پیرامون میانگین میانگین‌ها در یک توزیع نمونه‌گیری است. (شکل ۷)

1- bell-shaped
2- Standard error
3- Standard deviation



شکل ۷

خطای نمونه‌گیری[1]

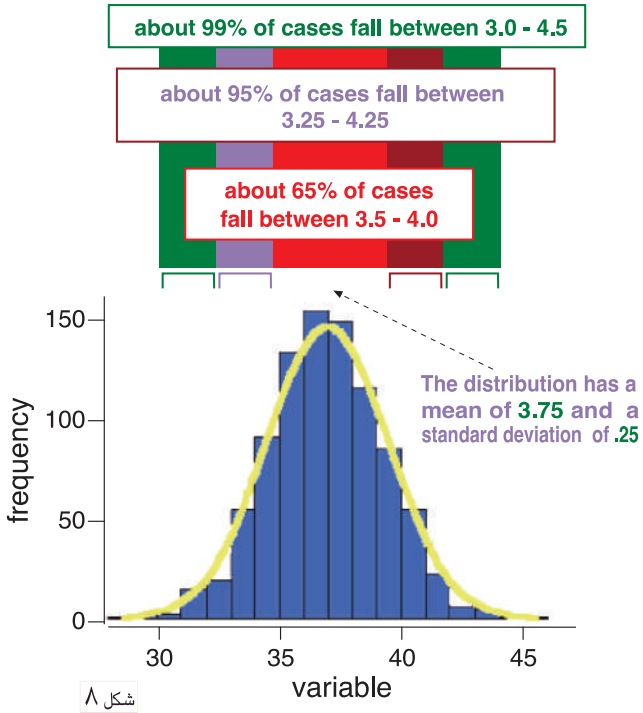
در حیطه نمونه‌گیری خطای معیار در اصطلاح خطای نمونه‌گیری نامیده می‌شود. خطای نمونه‌گیری به ما راجع به دقت تخمین آماری آگاهی می‌دهد. هر چه خطای نمونه‌گیری کمتر باشد به این معناست که دامنه[2] و تغییرپذیری[3] توزیع نمونه‌گیری به نسبت کمتر است. حال دوباره به مشکل اول برمی‌گردیم، امکان دیدن توزیع نمونه‌گیری در دنیای واقعی وجود ندارد! پس چگونه می‌توانیم خطای نمونه‌گیری را محاسبه کنیم؟ برای این کار می‌توانیم محاسبات را بر پایه انحراف معیار نمونه خود انجام دهیم. هر چه انحراف معیار نمونه ما بیشتر باشد،

1- Sampling error
2- Range
3- Variability

خطای معیار (و خطای نمونه‌گیری) بیشتر خواهد بود. خطای معیار با حجم نمونه نیز رابطه دارد بطوریکه هر چه حجم نمونه شما بزرگتر باشد، خطای معیار کوچکتر خواهد بود. زیرا هر قدر حجم نمونه بزرگتر باشد، به جمعیت واقعی نزدیکتر است. بنابراین اگر نمونه‌ای که می‌گیرید کل جمعیت را شامل شود خطای نمونه‌گیری شما به صفر می‌رسد چرا که در اینصورت دیگر نمونه ندارید بلکه کل جمعیت در اختیار شماست. بدیهی است در چنین شرایطی میانگینی که محاسبه می‌کنید همان پارامتر جمعیت است.

قانون ۶۵، ۹۵ و ۹۹ درصد

این قانون کلی هرگاه یک توزیع نرمال یا زنگوله‌ای شکل داشته باشیم قابل به کارگیری است. از میانگین به عنوان مرکز توزیع شروع کنید. اگر به اندازه یک واحد استاندارد از میانگین به سمت چپ و راست (مقادیر بیشتر و کمتر) حرکت کنید محدوده بدست آمده ۶۵ درصد موارد موجود در توزیع را شامل خواهد شد (به بیان دیگر ۶۵ درصد سطح زیر منحنی). اگر به اندازه دو واحد استاندارد به سمت چپ و راست حرکت کنید محدوده بدست آمده تقریباً ۹۵ درصد موارد را شامل خواهد شد به همین ترتیب جابجایی به اندازه سه واحد استاندارد به راست و چپ ۹۹ درصد موارد را در برمی‌گیرد. توجه کنید که در جملات قبلی صحبتی راجع به واحد انحراف معیار یا واحد



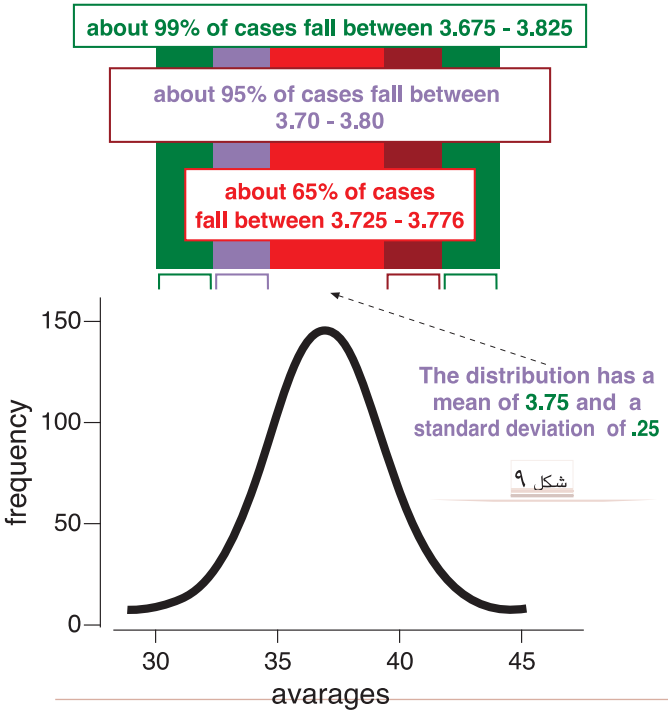
شکل ۸

خطای معیار نشد و به جای آن از واحد استاندارد استفاده شد. این امر بدین علت است که قوانین مشابهی بر هردو نوع توزیع حاکم است. برای مثال در شکل ۸ میانگین توزیع برابر ۳/۷۵ است و یک واحد استاندارد ۰/۲۵ می باشد. (اگر توزیع مربوط به داده‌های نمونه بود باید راجع به انحراف معیار صحبت می‌شد و اگر توزیع نمونه‌گیری مدنظر بود بایستی از خطای معیار استفاده می‌شد). بنابراین اگر به اندازه ۰/۲۵ از میانگین به سمت چپ و راست حرکت کنیم (بین ۳/۵ تا ۴/۰) انتظار داریم ۶۵ درصد موارد در این قسمت واقع شوند و ... (شکل ۸).

ممکن است بپرسید این محاسبات چه کاربردی دارند؟ چنانچه با داده‌های خام سروکار داشته باشیم و از میانگین و انحراف معیار نمونه خود آگاه باشیم می‌توانیم بازه‌هایی را که ۶۵، ۹۵ و ۹۹ درصد موارد، پارامتر در آن قرار خواهد گرفت پیش‌بینی کنیم. به این بازه‌ها در اصطلاح بازه‌های اطمینان ۶۵، ۹۵ و ۹۹ درصد می‌گویند. (شکل ۹)

حال باید مطالب فوق را جمع‌بندی کنیم و به نتیجه مورد نظر برسیم:

اگر یک توزیع نمونه‌گیری داشتیم می‌توانستیم بازه‌های اطمینان ۶۵، ۹۵ و ۹۹ درصد را برای پارامتر جمعیت پیش‌بینی کنیم. (این کار در واقع همان هدف اولیه ما از نمونه‌گیری است). و اینگونه می‌توانستیم پیش‌بینی کنیم مقدار متغیر مورد نظر ما در جمعیت چقدر است. فقط یک مشکل وجود دارد و آن این است که در واقعیت در اختیار



شکل ۹

داشتن توزیع نمونه‌گیری ممکن نیست. ولی ما هنوز توزیع متغیر در نمونه خود را در اختیار داریم و از طریق این توزیع می‌توانیم اندازه خطای معیار (خطای نمونه‌گیری) را محاسبه کنیم، چرا که خطای معیار برپایه انحراف معیار تعیین می‌شود و انحراف معیار در نمونه براحتی قابل محاسبه است. همچنین می توانیم از بهترین تخمین خود برای ارزش پارامتر در جمعیت یعنی آماره نمونه استفاده کنیم.

اگر تخمین مناسبی از میانگین توزیع نمونه‌گیری (با توجه به میانگین نمونه خود) و خطای معیار داشته باشیم (از نمونه خود آن را محاسبه می‌کنیم)، دو عامل کلیدی مورد نیاز را برای مشخص کردن توزیع نمونه‌گیری در اختیار داریم و از این طریق می‌توانیم بازه اطمینان پارامتر جمعیت را تخمین بزنیم.

به مثال زیر دقت کنید:

$$\text{Standard Error} = \frac{\text{Standard Deviafion}}{\sqrt{\text{Sample Size}}} = \frac{0.25}{\sqrt{100}}$$

تصور کنید برای انجام یک مطالعه نمونه‌ای از جمعیت انتخاب کرده‌اید. حال فرض کنید میانگین برای نمونه شما برابر ۳/۷۵ با انحراف معیار ۰/۲۵ بوده است. توزیع نمونه‌گیری در این مورد چگونه خواهد بود؟

ابتدا فرض می‌کنیم که میانگین توزیع نمونه‌گیری همان میانگین نمونه است (۳/۷۵)، سپس خطای معیار را محاسبه می‌کنیم. برای این کار به انحراف معیار نمونه و حجم نمونه (در این مثال N =100) نیاز داریم. خطای معیار محاسبه شده برابر ۰/۰۲۵ است.

اکنون تمام مواد لازم برای تخمین بازه اطمینان برای پارامتر جمعیت در دسترس است. با ۶۵ درصد احتمال ارزش واقعی پارامتر بین ۳/۷۲۵ و ۳/۷۷۵ (۳/۷۵ ± ۰/۰۲۵) تخمین زده می‌شود. بازه اطمینان ۹۵ درصد بین ۳/۷۰۰ تا ۳/۸۰۰ خواهد بود

و می‌توانیم با ۹۹ درصد اطمینان ادعا کنیم که ارزش واقعی پارامتر جمعیت بین ۳/۶۷۵ و ۳/۸۲۵ است. ارزش واقعی (در این مثال) ۳/۷۲ بود که توانستیم تخمین نسبتاً صحیحی از آن با توجه به نمونه خود داشته باشیم. به طور کلی باید به آماره نمونه به چشم تخمینی از پارامتر جمعیت نگاه کنیم و در نظر داشته باشیم حتی هنگامیکه از بهترین فرایندهای نمونه‌گیری استفاده می‌کنیم باز هم تضمینی وجود ندارد که آماره نمونه نماینده خوبی از پارامتر جمعیت باشد. بنابراین برای هر استنباطی در این زمینه باید خطای نمونه‌گیری را در نظر بگیریم.

آزمون معناداری آماری[1]

همانطور که گفتیم صرف وجود شانس[2] می‌تواند باعث بدست آمدن نتایج غیرواقعی شود ولی متأسفانه هیچ روش مستقیمی برای تست کردن این مسأله وجود ندارد. بنابراین پژوهشگران با یک روش غیرمستقیم و با حذف فرضیهٔ صفر[3] می‌توانند فرض خود را تأیید کنند. به این روش غیرمستقیم "آزمون معناداری آماری" می‌گویند. این روش احتمال بدست آمدن اطلاعات مشاهده شده را با فرض اینکه در واقعیت (در کل جامعه) هیچ تفاوتی بین دو گروه نباشد (فرضیهٔ صفر)، تعیین می‌کند. در حقیقت ما به سوال پژوهش در سطح نمونه خود پاسخ می‌دهیم ولی این آزمون در سطح جمعیت به آن نگاه می‌کند. با این روش می‌توانیم قضاوت کنیم آیا تفاوتی که در نمونه مشاهده کرده‌ایم از نظر آماری در جمعیت معنادار هست یا خیر؟

بنابراین هدف از این آزمون، نتیجه‌گیری[4] یا استنباط[5] درباره جمعیت بوسیله مطالعه نمونه‌هایی از آن جمعیت است.

آزمون فرضیه یا آزمون معناداری آماری بر این فرض استوار است که در جمعیت دو وضعیت در گروههای مقایسه برقرار است:

1- Statistical Significance Testing
2- Chance Selection
3- Null hypothesis
4- Conclusion
5- Inference

یا تفاوت وجود دارد یا وجود ندارد. این آزمون در بدو امر فرض می‌کند هیچ تفاوتی بین گروهها در جمعیت وجود ندارد و در ضمن صحبتی راجع به میزان اهمیت یا اندازه این تفاوت بالقوه نمی‌کند.

این روش با فرضیه مطالعه سروکار دارد، بدین صورت که در وهله اول فرض می‌کند فرضیه مطالعه صحیح نمی‌باشد. (فرضیه معمولاً بیانگر یک تفاوت بین گروههاست) در نتیجه یک فرضیه جدید به نام فرضیه صفر شکل می‌گیرد که بیان می‌کند تفاوتی در بین گروهها در کل جمعیت نیست. سپس احتمال بدست آمدن نتایج مشاهده شده در نمونه در صورتیکه هیچ تفاوتی بین گروهها در جمعیت نباشد، با استفاده از روشهای آماری محاسبه می‌شود. وقتی احتمال پیشامدن نتایج مشاهده شده در نمونه با وجود درستی فرضیه صفر[1] بسیار کوچک باشد پژوهشگران فرضیه صفر (HO) را رد کرده و به سراغ تنها جایگزین باقیمانده یعنی فرضیه اصلی خود می‌روند. از این جهت به فرضیه مطالعه، فرضیه جایگزین (Ha) [2] نیز می‌گویند.

(لازم به توضیح است که این فرایند به هیچ وجه امکان اثبات و یا نفی فرضیه را فراهم نمی‌کند.)

مراحل آزمون معنادار بودن آماری:

۱ بیان فرضیه: همانطور که گفته شد در اکثر موارد قبل از جمع‌آوری داده‌ها نیاز به یک فرضیه داریم. مثلاً دو گروه با دو خصوصیت متفاوت از جامعه را برداشته و بیان می‌کنیم از نظر خاصیت دیگری متفاوت هستند.

۲ ساختاربندی فرضیه صفر (H0)

۳ باید یک احتمال را که به اندازه کافی برای رد کردن فرضیه صفر کوچک باشد در نظر گرفت.[3] در اکثر موارد یک احتمال ۵ درصد یا کمتر کافی به نظر می‌رسد. این مطلب بیان می‌کند H0 که در واقعیت صحیح است در ۵ ٪

1- P[observed data / Ho is true]
2- Alternative hypothesis
3- Cut off level

موارد به نفع Ha قربانی می‌شود.

۴ مرحله جمع‌آوری اطلاعات: این مرحله به کمک انواع طرحهای موجود برای مطالعه انجام می‌شود.

۵ اجرا کردن آزمون: اگر تفاوتی بین گروههای کنترل و مطالعه وجود داشت محققین احتمال پیشامدن این تفاوتها را در صورت درست بودن H0 بدست می‌آورند. به این احتمال P value گفته می‌شود. برای این منظور پژوهشگران مجبورند از بین روشهای متعدد موجود نوعی از آزمون معنادار بودن آماری را انتخاب کنند که برای داده‌ها مناسب باشد.

مثال ۱ : فرض کنید یک پژوهشگر می‌خواهد به سوال زیر پاسخ دهد: "آیا در ایالات متحده به یک تعداد دختر و پسر بدنیا می‌آیند؟"

فرضیه پژوهشگر: تعداد پسرهای بدنیا آمده بیش از دخترهاست.

فرضیه صفر: در ایالات متحده دختر و پسر به یک تعداد متولد می‌شوند.

(5%) 0.05: (Cut off level)

سطح معناداری

پژوهشگر با گرفتن یک نمونه از جمعیت (در اینجا ۴ شناسنامه) در می‌یابد که همه آنها پسر بوده‌اند. حال احتمال بدست آمدن چنین نتیجه‌ای را در صورتیکه H0 درست باشد بررسی می‌کنیم.

Probability of four males in a raw: 0.5 x 0.5 x 0.5 x 0.5= 6.25%

P Value =0.0625

توانایی یا عدم توانایی در رد کردن H0: اگر P Value بدست آمده کمتر یا مساوی حد تعیین شده (سطح معناداری) باشد می‌توانیم H0 را رد کنیم، و با رد H0 خودبخود Ha تأیید می‌شود. در این مثال با P value برابر ۰/۶۲۵ نمی‌توان فرضیه صفر را رد کرد.

مثال۲) Ha: پیپ کشیدن ریسک سرطان دهان را بالا می‌برد.

H0: سرطان دهان با فرکانس برابری بین افرادی که پیپ می‌کشند و افرادیکه پیپ نمی‌کشند، اتفاق می‌افتد.

سطح معناداری (Cut off level): ۰/۵ در مرحله بعد پژوهشگر با گرفتن نمونه از جامعه پیپ کش‌ها و غیرپیپ‌کش‌ها اقدام به جمع‌آوری اطلاعات می‌کند. او با استفاده از یک آزمون معنادار بودن آماری مناسب درمی‌یابد که احتمال نتایج مشاهده شده در نمونه در صورت صحیح بودن H0 (P value) برابر ۰/۰۳ است. در نتیجه پژوهشگر H0 را رد می‌کند و نتیجه می‌گیرد دو گروه مورد مطالعه از نظر فرکانس سرطان دهان در جمعیت متفاوت هستند.

وقتی تفاوت آماری معنادار بین گروهها دیده شد اصطلاحاً می‌گوییم بین آن گروه و ویژگی خاص مورد مطالعه همراهی[1] وجود دارد. به عبارت دیگر این دو بیش از آنکه شانس به تنهایی باعث شود با هم اتفاق می‌افتند (معنی‌دار بودن آماری رابطه یا همراهی علت با معلول یکی از معیارهای اصلی رابطهٔ علّی است).

قبلاً یک شانس کمتر یا مساوی ۵٪ را برای پیشامدن نتایج مشاهده شده[2] در صورتیکه هیچ تفاوتی در جمعیت وجود نداشته باشد[3] تعریف کردیم. این ۵٪ ممکن است در مواردی که تصمیمات مهمی با توجه به این پژوهش گرفته می‌شود بسیار زیاد باشد. در حقیقت با اینکه این عدد بر بعضی ویژگیهای آماری استوار است ولی با این حال اصطلاحاً "وحی منزل" نیست و می‌توان احتمالاتی نظیر ۱ ٪، ۰/۱ ٪ یا هر احتمال دیگر را برای این منظور تعریف کرد. در هر صورت با اینکه مسأله‌ای نیست چه عددی انتخاب شود، همیشه احتمال رد H0 با وجود فقدان تفاوت در جمعیت وجود دارد. آزمون معنادار بودن آماری می‌تواند این احتمال را اندازه بگیرد ولی قادر به

1- Association
2- Observed result
3- H0 is true

حذف آن نیست. به عبارت دیگر حذف چنین خطایی حتی با استفاده‌از نمونه‌های بزرگ و بخوبی انتخاب شده ممکن نیست ولی از طریق این آزمون می‌توان احتمال خطا را اندازه گرفت و در مورد آن قضاوت کرد.

خطاهایی[1] که ممکن است در این آزمون اتفاق بیفتد عبارتست از:

۱ عدم بیان صحیح فرضیه قبل از شروع مطالعه

۲ عدم توانایی در تفسیر صحیح نتایج آزمون به علت در نظر نگرفتن خطای نوع I

۳ عدم توانایی در تفسیر صحیح نتایج آزمون به علت در نظر نگرفتن خطای نوع II

با یک مثال نتایج عدم ذکر فرضیه قبل از شروع مطالعه ملموس‌تر می‌شود. ملاحظه کنید:

یک پژوهشگر با انتخاب تصادفی صد بیمار مبتلا به فشار خون بالا و صد فرد غیرمبتلا به فشار خون بالا و مقایسه آنها از نظر صد متغیر مشخص کشف کرد که دو گروه از نظر دو متغیر با هم متفاوت هستند (این تفاوت از نظر آماری معنی‌دار بود (با P value زیر ۵٪). این دو متغیر عبارت بودند از:

(۱) افراد با فشار خون بالا در نام فامیلی خود تعداد حرف بیشتری نسبت به افراد غیرمبتلا به فشار خون بالا داشتند!

(۲) افراد با فشار خون بالا عموماً در نیمه اول هفته بدنیا آمده بودند و افراد غیرمبتلا به فشار خون بالا در نیمه دوم هفته بدنیا آمده بودند!

نتیجه‌گیری پژوهشگر: با وجود این واقعیت که این تفاوتها پیش‌بینی نشده بود ولی نام فامیلی طولانی‌تر و بدنیا آمدن در نیمه اول هفته بین دو گروه مبتلابه فشار خون بالا متفاوت است.

این مثال نشاندهنده اهمیت ذکر کردن فرضیه قبل از شروع مطالعه است. هنگامیکه تعداد زیادی از متغیرها را تست می‌کنیم

1- Errors

شانس به تنهایی می‌تواند باعث شود بعضی متغیرها از نظر آماری تفاوتهای معنادار نشان بدهند.

اگر بدون داشتن فرضیه مطالعه بدنبال تفاوتها باشیم و یا بعد از جمع‌آوری و تحلیل داده‌ها پی به تفاوتها ببریم، معیار محکمتری برای معنادار بودن تفاوت به لحاظ آماری نسبت به ۵٪ معمول مورد نیاز است.

خطای نوع I

بعضی خطاها ماهیتاً در آزمون معنادار بودن آماری وجود دارند. خطای نوع I[1] براساس مفهوم آزمون به این صورت است که فرضیه صفر (H0) به اشتباه رد شود و Ha به نادرستی مورد قبول قرار گیرد. در فرم معمول این آزمون حدود ۵ ٪ شانس ۳۵پذیرش فرضیه مطالعه (Ha) را در صورتیکه واقعاً هیچ نوع تفاوتی در جامعه‌ای که نمونه را از آن گرفتیم وجود نداشته باشد قابل قبول در نظر می‌گیرند. به این سطح از خطای نوع I اصطلاحاً سطح آلفا[2] می‌گویند.

شایان ذکر است که انجام این آزمون هیچگاه عدم قطعیت را از بین نمی‌برد، بلکه باید دید آیا این درجه از عدم قطعیت قابل تحمل است یا خیر. مثال: در جامعه‌ای که بیماریهای با منشاء آب[3] شیوع کمی دارد، یک روش جدید تصفیه آب با ۵٪ احتمال عدم توانایی در حذف کامل این بیماریها ممکن است قابل قبول نباشد ولی همین روش جدید در جامعه‌ای که آب یکی از عمده‌ترین منابع انتقال بیماریهاست حتی با احتمالهای بالاتر براحتی مورد قبول واقع می‌شود، اگر روش دیگری در دسترس نباشد.

حال ببینیم در نظر نگرفتن احتمال خطای نوع I می‌تواند به چه تفاسیر سوئی از نتایج مطالعه منجر شود. مثال: نویسنده یک مقاله مروری با ارزیابی بیست مطالعه انجام شده بر روی رابطه شیر دادن و سرطان پستان مشاهده کرد، ۱۹ مطالعه هیچ تفاوتی در شیوع سرطان پستان بین زنان شیرده و زنانی که نوزادان خود را با شیر خشک تغذیه می‌کردند، نشان نداده است ولی در یک مطالعه تفاوت معنادار در افزایش فراوانی سرطان پستان دیده شده است. (نتایج با P value=۰/۰۵ از نظر آماری معنادار بودند).

1- Type I Error
2- ∝ Level
3- Water-born

وی بخاطر وجود این مطالعه نتیجه‌گیری کرد که شیردادن مادران باید محدود شود.

این در حالیست که با در نظر گرفتن همان ۵٪ شانس از ۲۰ مطالعه براحتی یک مطالعه می‌تواند نشان دهنده تفاوت باشد، در حالیکه این نتیجه با پدیده شانس قابل توجیه است.

خطای نوع II[1]

این خطا وقتی اتفاق می‌افتد که ما نمی‌توانیم معنادار بودن تفاوت را نشان بدهیم. در حالیکه عدم توانایی در رد H0 الزاماً دلیل عدم وجود تفاوت معنادار در جمعیت نیست، به عبارت دیگر، اینکه ما نتوانیم H0 را به اندازه مورد نیاز تضعیف کنیم نشاندهنده تقویت شدن یا ثابت شدن آن نیست. اصولاً آزمون معناداری امکان رد کردن یا نکردن H0 را فراهم می‌کند ولی توانایی اثبات H0 را ندارد. بنابراین عدم توانایی در رد H0 ، صرفاً نشاندهنده این است که شواهد ما به قدر کافی برای رد این فرض قوی نبوده است.

خطای نوع I زمانی پیش می‌آید که شانس باعث می‌شود یک سری از داده‌ها را مشاهده کنیم که هیچ تفاوت معنی‌داری را نشان نمی‌دهند. این مطلب به معنای اشکال در طراحی و یا مراحل اجرایی مطالعه نیست بلکه صرفاً نشان‌دهنده شانس است که باعث نتایج نادرست می‌شود. تلاشهایی که در زمینه انجام این آزمون صورت می‌گیرد همیشه احتمال خطا را با خود بهمراه دارد.

همانطور که گفتیم هر چه تعداد افراد شرکت کننده در مطالعه کمتر باشد شانس بدست آوردن یک نمونه غیرنماینده بیشتر است. و از این طریق احتمال خطای نوع II افزایش می‌یابد. هر چه نمونه کوچکتری گرفته شود تفاوتهای مشاهده شده باید بیشتر باشد بنابراین اگر تفاوت بین دو گروه خیلی زیاد است نیاز به نمونه کوچکتری برای نشان دادن آن وجود دارد. به عبارت دیگر وقتی نمونه کوچک انتخاب می‌شود برای رد کردن H0 نیاز به تفاوت بزرگتری داریم که در اینصورت شانس شکست خوردن ما در نشان

1- Type II Error

دادن این تفاوت بزرگ افزایش می‌یابد و احتمال خطای نوع II افزون می‌گردد.

توان آماری[1]

توان آماری یک مطالعه عبارتست از احتمال نشان دادن یک تفاوت آماری معنادار توسط آن مطالعه. این کمیت توسط یک سری از تکنیکهای آماری محاسبه می‌شود. همانطور که از تعریف توان آماری مشخص است، از نظر آماری توان مکمل خطای نوع II در نظر گرفته می‌شود.[2]

هدف از یک مطالعه خوب طراحی شده داشتن خطای نوع II ۱۰ تا ۲۰ درصد یا توان آماری ۸۰ تا ۹۰ درصد است. پژوهشگرانی که از نمونه‌های کوچک‌تر استفاده می‌کنند ۳۰ تا ۴۰٪ احتمال خطای نوع II و یا حتی بیش از این احتمال را نیز قبول می‌کنند. به عبارت دیگر احتمال آنکه مطالعه بتواند یک تفاوت آماری معنادار راکه واقعاً در جامعه وجود دارد نشان بدهد (توان آماری) در این مطالعات ۶۰ تا ۷۰ درصد در نظر گرفته می‌شود.

به طور کلی باید در نظر داشته باشید که در آزمونهای آماری سه کمیت حجم نمونه، سطح آلفا و توان آماری اجزایی هستند که میزان دقت آماری را تعیین می‌کنند.

تغییر دادن هر کدام از این اجزاء روی دو جزء دیگر موثر است مثلاً همانطور که گفته شد هر چه حجم نمونه بیشتر باشد توان آماری مطالعه بیشتر خواهد بود و یا هر چه سطح آلفا کوچکتر در نظر گرفته شود از آنجاییکه رد فرضیه صفر دشوارتر می‌شود بنابراین توان مطالعه در نشان دادن تفاوتهای کوچک ولی معنادار کمتر خواهد شد. بنابراین در انجام هر پژوهش می‌توان با تغییرهر کدام از این سه جزء به ترکیب دلخواه رسید.

1- Statistical Power
2- Statistical power=1-Type II error

همانطور که در فصل بعد نیز خواهید دید، قضاوت راجع به دقت آماری در ارزیابی نقادانه مطالعات بیشتر به دو جزء P value و بازه اطمینان مربوط می‌شود.

در مورد P value ضمن آنکه کمتر بودن اندازه از یک حد قراردادی[1] مثلاً ۵٪ اهمیت دارد، اندازه آن نیز به تنهایی حائز اهمیت است و هر چه کوچکتر باشد تفاوت مشاهده شده معنادارتر است. مثلاً P value ۰/۰۰۱ از ۰/۰۱ صرف نظر از بازهٔ اطمینان دقت آماری بیشتری دارد.

قضاوت در مورد بازه اطمینان کمی پیچیده‌تر است. به طور کلی با یک درجه اطمینان ثابت مثلاً ۹۵ درصد هر چه بازه محاسبه شده برای اندازهٔ پارامتر در جمعیت پهنای کمتری داشته باشد، دقت آماری بیشتر است. بازه اطمینان علاوه بر دقت آماری در بررسی حجم اثر (خصوصاً در مورد پیامدهای دوتایی) نیز اهمیت دارد که در فصل بعد در بارهٔ آن توضیح داده شده است.

1- ∝ level/cut off level

12- How to use the Evidence: assessment and application of scientific evidence. National Health and Medical Research Council; 2000.

Chapter 1

References

1- Trochim WM. Research Methods Knowledge Base. New York. Cornell University; 2003.

2- Altman DG. Practical Statistics For Medical Research. 2nd ed. Chapman & Hall 1997.

3- Polgar S, Thomas SA. Introduction to Research in the Health Science. 4th ed. Churchill Livingstone 2000.

4- Moye LA. Statistical Reasoning in Medicine. Springer, 2000.

5- Guyatt G, Rennie D, Hayward R. User's Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice.American Medical Association; 2002.

6- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine, How to Practice and Teach EBM, 2nd ed. Churchill Livingstone 2000.

7- Egger M, Smith GD, Altman DG, Systematic Reviews in Health Care: Meta-analysis in context. 2nd ed. BMJ Publishing Group 2001.

8- Cochrane Reviwer's Handbook. The Cochrane Collaboration; 2004.

9- Cochrane Rviewers' Handbook Glossary. The Cochrane Collaboration; 2001.

10- Riegelman RK. Studying a Study and Testing a Test. 4th ed. Lippincott Williams and Wilkins; 2000.

11- How to Review the Evidence: Systematic identification and review of the scientific literature. National Health and Medical Research Council; 1999.

فصل دوم

مبانی ارزیابی نقادانهٔ شواهد

فصل دوم
مبانی ارزیابی نقادانه شواهد

پیش از پرداختن به ارزیابی نقادانه شواهد مربوط به سوالات بالینی مختلف و ارائه ابزارهای کاربردی لازم است با مفاهیم کلی و رویکردهای مربوط به ارزیابی شواهد آشنا شویم.

ارزیابی یک مقاله علمی از جنبه‌های متعددی امکان‌پذیر است، در این فصل سعی شده است با پاسخ به چند سوال کلی مفاهیم مورد نیاز برای ارزیابی شواهد روشن شود.

آیا کیفیت مطالعه باید مورد ارزیابی قرار بگیرد یا کیفیت گزارش آن؟

قبل از شروع به ارزیابی مطالعات باید تعیین کنید که از چه زاویه‌ای می‌خواهید مطالعه را بررسی کنید.

کیفیت گزارش مقاله (آیا در گزارش این مطالعه تمام موارد ضروری ذکر شده است؟)

ارزیابی گزارش مقاله و ارزیابی کیفیت مطالعه دو فرآیند مستقل و جداگانه نیستند. در حقیقت در ارزیابی کیفیت گزارش مقاله بررسی می‌کنید تا چه میزان گزارش مقاله اطلاعات لازم را در مورد طرح، نحوهٔ اجرا و تحلیل پژوهش در اختیار شما قرار می‌دهد. برای مثال شواهد نشان می‌دهد گزارش بسیاری از مطالعات جزئیات روش شناختی مهم را حذف می‌کنند. مثلاً در مروری بر ۱۲۲

کارآزمایی تصادفی شده برای بررسی اثر مهار کننده‌های اختصاصی باز جذب سروتونین مشخص شد که تنها یک مورد روش تصادفی کردن را ذکر کرده است. رویکردی که به طور شایع برای برخورد با این مشکل توصیه می‌شود این است که کیفیت مطالعه را برای ویژگی مورد نظر که در گزارش مقاله به خوبی ذکر نشده است پایین در نظر می‌گیرند مگر اینکه خلاف آن ثابت شود. از آن جایی که گزارش ناکافی معمولاً بازتابی از روش‌های ناصحیح است می‌توان این رویکرد را توجیه کرد. با این وجود مطالعاتی که به خوبی اجرا شده‌اند ولی گزارش آنها ناکافی و مشکل‌دار است در این رویکرد در جایگاه نادرستی قرار می‌گیرند. رویکرد جایگزینی که می‌توانید پیش بگیرید ارزیابی دقیق و کامل کیفیت گزارش مقاله بدون بررسی کفایت روش‌های مطالعه است. این رویکرد نیز خالی از اشکال نیست زیرا مطالعه‌ای که از نظر کیفیت روشها دچار مشکل است ولی به خوبی گزارش شده است با این رویکرد می‌تواند بیشترین اعتبار را کسب کند. به عنوان راه حل نهایی می‌توانید داده‌های اضافی را از طریق برقراری تماس با پژوهشگران و نویسندگان مقاله به دست آورید ولی این روش نیز همیشه عملی نیست. امروزه بهره‌گیری از راهکارهای مشخص برای نحوهٔ گزارش مطالعات (خصوصاً در مورد کارآزمایی‌های بالینی) در تعدادی از نشریات وضعیت را تا حدودی بهبود بخشیده است ولی هنوز هم نارسایی‌ها و کاستی‌هایی که در گزارش مطالعات وجود دارد ممکن است با کاستی‌های طراحی، اجرا و تحلیل پژوهش اشتباه شود. بنابراین ارزیابی گزارش مقاله در کنار بررسی نقادانهٔ خود مطالعه معنی پیدا می‌کند.

تعدادی از ابزارهای ارزیابی نقادانهٔ شواهد صرفاً مربوط به بررسی کیفیت گزارش مقالات هستند ولی اکثر این ابزارها هم کیفیت گزارش مقاله و هم ارزیابی خود مطالعه را مد نظر قرار داده‌اند. به هر حال در نظر داشته باشید انتخاب ابزار مناسب برای ارزیابی نقادانه‌ی گزارش مطالعات، همیشه کار ساده‌ای نیست. در یک مقاله‌ی مروری نظام‌مند ابزارهایی که بدین منظور طراحی شده بودند مورد مطالعه قرار گرفتند.[1] این مقاله نشان داد ابزارهای بررسی گزارش مقالات از جهات مختلف از جمله تعریف

1- Moher 1995, Moher 1996

کیفیت، میزان مطابقت با روش‌های علمی، تعداد موارد موجود و زمان لازم برای استفاده از آنها، با هم متفاوتند. از طرف دیگر وقتی کیفیت گزارش مطالعات مشابه با ابزارهای مختلف سنجیده شد، میزان کیفیت گزارش شده در هر یک از ابزارها متفاوت بود.

به طور کلی در گزارش مقالات باید جنبه‌هایی که از نظر ارزیابی مطالعات اهمیت دارند به طور شفاف و کامل توصیف شوند. به عنوان مثال در مقالاتی که دقت تشخیصی یک تست خاص را بررسی می‌کنند ارائه توصیف شفاف و کاملی از تست آزمایشی و تست مرجع و تعریف شفاف پیامدهای منفی و مثبت برای هر دوی این تستها، در گزارش پژوهش ضروری است.

چه ویژگی‌هایی از مطالعه باید در ارزیابی نقادانهٔ شواهد مورد بررسی قرار گیرند؟

برای پاسخ به این سوال ابتدا باید با ابعاد شواهد[1] آشنا شوید. به طور کلی برای شواهد می‌توان سه بعد قائل شد: قدرت شواهد[2]، اندازهٔ اثر[3] و میزان مربوط بودن شواهد[4].

ارزیابی نقادانهٔ یک مطالعه بدون در نظر گرفتن هر یک از این ابعاد و اجزاء آنها ناکامل خواهد بود. حال کمی بیشتر با این ابعاد آشنا شویم:

قدرت شواهد

آیا اثر مشاهده شده در مطالعه مربوط به تورش[5] و یا شانس است؟

به طور کلی خطاهایی که می‌تواند در بروز پیامد مشاهده شده در مطالعه مداخله کنند به دو نوع کلی نظام‌مند (تورش)[6] و غیرنظام‌مند (خطای تصادفی)[7] تقسیم می‌شوند.

1- Dimensions of evidence
2- Strength of evidence
3- Size of effect
4- Relevance of Evidence
5- Bias
6- Systematic
7- Random error

درک مفهوم تورش و خطای تصادفی اهمیت زیادی در فهم قدرت شواهد دارد چرا که در حقیقت قدرت شواهد چیزی جز توانایی مطالعه در کاهش اثر تورش و خطای تصادفی نیست. به این مثال توجه کنید. فرض کنید با یک اسلحه در حال شلیک به مرکز یک سیبل هستید. پس از انجام چند شلیک سیبل را نزدیک آورده و به آن نگاه می‌کنید. تمام گلوله‌ها در یک جهت منحرف شده‌اند. (ممکن است این انحراف یک طرفه از هدف به علت نقص در مگسک اسلحه‌ی شما باشد). به چنین تمایل نظام‌مندی برای انحراف از ارزش واقعی در اصطلاح روش شناختی «تورش» گفته می‌شود. عدم وجود تورش در یک مطالعه به معنی روایی[1] نتایج آن مطالعه است.

در مثال فوق گلوله‌ها می‌توانند در جهات مختلف و بدون نظم خاصی پراکنده شده باشند. خطاهایی که بدین‌سان از انحراف پراکنده و بدون نظم از ارزش واقعی منشاء می‌گیرند تحت عنوان خطاهای تصادفی نام برده می‌شوند. عدم وجود خطای تصادفی در یک مطالعه به معنی دقت[2] آن مطالعه است. خطای تصادفی از شانس منشاء می‌گیرد بنابراین نمی‌تواند جهت خاصی داشته باشد. در مورد خطای تصادفی و روشهای مقابلهٔ با آن در بخش روش‌شناسی پژوهش مطالبی بیان شد. بر خلاف خطای تصادفی، تورش همان طور که گفته شد به اثر نظام‌مندی اشاره دارد که در یک جهت خاص و قابل پیش‌بینی بر روی پیامد مطالعه تأثیر می‌گذارد. با این وجود تورش یک پیش‌داوری، قبل از آگاهی از واقعیت نیست و زمانی اتفاق می‌افتد که پژوهشگر به طور غیرارادی عواملی را وارد پژوهش می‌کند که ناخواسته بر روی نتایج و پیامدهای مطالعه تأثیر می‌گذارند. در مورد انواع مختلف تورش و منشاء آن در مطالعات به تفصیل در بخشهای مربوطه توضیح داده شده است.

قدرت شواهد را می‌توان به سه جزء تقسیم نمود: سطح شواهد، کیفیت شواهد و دقت آماری.

1- Validity
2- Precision

سطح شواهد

سطح شواهد[1] اولین جزء قدرت شواهد است و در برگیرندهٔ توانمندی طرحی است که پژوهشگر با توجه به نوع سوال خود برای مطالعه انتخاب می‌کند. در حقیقت سطحی که به یک مطالعه نسبت داده می‌شود نمودی است از میزانی که طرح مطالعه توانسته است تورش را از بین ببرد.

در مورد هر نوع سوال بالینی که برای پزشک مطرح می‌شود انواع خاصی از طرح‌های مطالعه وجود دارد که قدرت بیشتری برای پاسخ‌گویی به آن سوالات دارند. در فصلهای دیگر کتاب طرحهای مناسب برای هر کدام از انواع سوالات بالینی بیان شده است. تنها نکته‌ای که در این جا می‌توان به آن اشاره کرد اهمیت مقالات مروری نظام‌مند[2] به عنوان بالاترین سطح شواهد است. در این نوع مطالعه پژوهشگر قاعدتاً تمام شواهد مربوط به یک سوال بالینی را یافته و ارزیابی می‌کند. بنابراین چنانچه مقالات مروری نظام‌مند به درستی تهیه شده باشند نتایج آنها دقیق‌ترین و بهترین پاسخ را در اختیار ما قرار می‌دهند با این حال مقالات مروری همیشه در دسترس نیستند. اکثر این مقالات در مورد اثربخشی مداخلات و عملکرد تستهای تشخیصی تهیه شده است و در سایر موارد تعداد مقالات مروری نظام‌مند نسبتاً محدود است.

برای پی بردن به اهمیت سطح شواهد به مثال زیر توجه کنید:

فرض کنیم شما می‌خواهید میزان اثربخشی آنتی‌اکسیدان‌ها از جمله بتاکاروتن را در پیشگیری از سرطانهای مختلف بررسی کنید. مطالعات غیرتصادفی شده از جمله مطالعات همگروهی نشان دهنده‌ی اثرات محافظتی قدرتمندی برای این مواد است[3] در حالی که کارآزمایی‌های تصادفی شدهٔ بالینی(RCT) اثر خاصی برای مواد آنتی‌اکسیدانت در کاهش خطر سرطان قائل نیستند[4]

در این مثال با توجه به محدودیت‌های طرح‌های غیرتصادفی شده (شبه تجربی) از جمله نبود کنترل مناسب بر روی عوامل

1- Level of evidence
2- Systematic Reviews
3- Byers and Perry 1992
4- TCCPSG 1994, Oxman et al 1996, Halley et al 1998

مداخله‌گر احتمال وجود تورش در نتایج این مطالعات بیشتر از مطالعات تصادفی شده است.

بنابراین پذیرش نتایج کارآزمایی‌های تصادفی شده که تخمین دقیق‌تری از اثر واقعی مداخله را در اختیار می‌گذارند اکیداً توصیه می‌شود.

چرا به طرحهای مختلف برای پژوهش نیاز داریم؟

همانطور که گفته شد برای پاسخگویی به سوٴالات بالینی متفاوت استفاده از طرح‌های پژوهشی مختلف ضروری است. به عنوان مثال در بررسی عوارض و اثرات سوء احتمالی درمانهای انجام شده ممکن است کارآزمایی‌های تصادفی شده به علت محدودیت در طول دوره پیگیری، حجم نمونه و مسائل اخلاقی، چندان کارا نباشند. در این موارد استفاده از مطالعات شبه تجربی از جمله مطالعات مورد ـ شاهدی و مطالعات همگروهی امکان بررسی عوارض مداخلات را در مقیاس وسیعتر فراهم می‌آورد. مثلا بهترین شواهد در دسترس در مورد اثرات جانبی داروی ضدانعقادی وارفارین (افزایش خطر خونریزی‌های شدید) از مطالعات همگروهی بدست آمده است. در ۵ کارآزمایی تصادفی شده که از وارفارین برای پیشگیری از سکته مغزی در افرادی که دچار فیبریلاسیون دهلیزی غیرروماتیسمی بودند، استفاده شده بود، افزایش خطر مطلق[1] برای خونریزی‌های شدید ۰/۶ درصد گزارش شده بود ولی در مطالعات همگروهی مبتنی بر جمعیت[2] این میزان ۳ درصد تخمین زده شده است.[3] همچنین مطالعاتی مورد ـ شاهدی به عنوان منبع اولیه اطلاعات مورد نیاز برای تعیین خطر خونریزی‌های گوارشی در اثر مصرف داروهای ضدالتهابی غیراستروئیدی[4] مطرح هستند.[5]

1- Absolute risk
2- Population - based
3- Glasziou et al 1998
4- NSAIDs
5- Henry et al 1996

کیفیت شواهد

ارائه تعریف مشخصی از کیفیت کار ساده‌ای نیست. در اینجا منظور از کیفیت شواهد[1] روشهایی است که پژوهشگر در قالب یک طرح پژوهشی مشخص در طول پژوهش به کار می‌گیرد تا میزان تورش و اثر عوامل مداخله‌گر را به حداقل برساند. بطور مثال دو مطالعه‌ای که هر دو بصورت کارآزمایی بالینی انجام شده اند (از سطح شواهد یکسانی برخوردارند) ممکن است از نظر انتخاب نمونه، کور بودن[2] و نحوهٔ تخصیص تصادفی با هم متفاوت باشند و لذا کیفیت نتایج دو مطالعه نیز با هم متفاوت خواهد بود. بنابراین بخش عمدهٔ کیفیت مطالعه به روایی نتایج حاصل از آن مربوط می‌شود. اهمیت ارزیابی کیفیت شواهد به اندازه‌ای است که بسیاری از صاحبنظران ارزیابی نقادانه شواهد را معادل با ارزیابی کیفیت شواهد در نظر می‌گیرند.

تقسیم‌بندیهای متعددی برای روایی مطالعات وجود دارد ولی مطابق آنچه که در منابع طبابت مبتنی بر شواهد آمده است، برای سادگی کار می‌توان روایی را به دو قسمت روایی داخلی و خارجی تقسیم کرد. روایی داخلی به واقعی بودن نتایج مطالعه با توجه به شرایط درون مطالعه می‌پردازد ولی بر خلاف آن روایی خارجی بر میزان تعمیم‌پذیری واقعی نتایج مطالعه به شرایط دیگر اشاره دارد. بنابراین روایی خارجی برخلاف روایی داخلی تنها زمانی معنی‌دار خواهد بود که شرایطی مانند جمعیت، رژیمهای درمانی، مکانهای طبابت و پیامدهااز قبل مشخص شده باشد.

به طور کلی همانطور که در ادامه خواهید دید روایی داخلی[3] در متون طبابت مبتنی بر شواهد معادل روایی در نظر گرفته می‌شود و اینگونه تعریف می‌شود: میزانی که طرح و روشهای استفاده شده در مطالعه می‌تواند از خطاهای نظام‌مند پیشگیری کند.[4] دوباره تأکید می‌کنیم موضوع مهمی که نباید با روایی اشتباه شود دقت مطالعه است.

در مطالعاتی که در زمینهٔ پزشکی و طبابت انجام می‌شود و بویژه‌کارآزمایی‌های بالینی چهار منشاء عمده برای تورش وجود

1- Quality of the evidence
2- Blinding
3- Internal Vlidity
4- Moher 1995

دارد که شامل تورش انتخاب، تورش عملکرد، تورش پیگیری و تورش سنجش است.

با وجود اینکه در بسیاری مواقع شواهد تجربی قدرتمندی در زمینه ارتباط بین نتایج مطالعات و میزان خطر تجمعی تورش‌ها وجود ندارد ولی دلایل منطقی فراوانی برای در نظر گرفتن تورش‌های بالقوه در ارزیابی نقادانه مطالعات موجود است.

۱ تورش انتخاب[1]

احتمال ایجاد این تورش در همهٔ مطالعاتی که به نحوی از گروههای کنترل استفاده می‌کنند وجود دارد. این نوع تورش یکی از عوامل مهمی است که می‌تواند نتایج مطالعه را تغییر دهد. تورش انتخاب همانطور که از نامش پیداست از نحوهٔ تخصیص و شکل‌گیری گروههای مقایسه در مطالعات ناشی می‌شود.[2]

به مطالعه فرضی زیر دقت کنید:

یک مطالعه مورد ـ شاهدی از بیماران مبتلا به سرطان پستان قبل از یائسگی سابقه استفاده از قرصهای ضدبارداری را بین ۵۰۰ خانمی که سرطان پستان داشتند و ۵۰۰ خانمی که از نظر سنی با آنها مشابه بودند و به علت فشار خون بالا یا دیابت بستری شده بودند بررسی کرده است. پژوهشگران در این مطالعه مشاهده کردند که ۴۰ در صد خانمهایی که سرطان پستان داشتند در طی ۵ سال اخیر از قرصهای ضدبارداری استفاده کرده‌اند. در حالی که فقط ۵ درصد از افرادی که دیابت یا فشار خون بالا داشتند (گروه کنترل) از این قرصها استفاده کرده بودند. بنابراین نتیجه گرفتند که رابطه قوی بین مصرف قرصهای ضدبارداری و سرطان پستان قبل از یائسگی وجود دارد؟

به نظر شما آیا می‌توان گفت در این مطالعه تورش انتخاب باعث چنین نتیجه‌گیری شده است؟

1- Selection Bias
2- Kunz 1998

برای پاسخ به این سوال ابتدا باید تعیین کنیم آیا خانمهایی که عضو گروه کنترل هستند به جز ابتلا به سرطان مشابه گروه مطالعه بوده‌اند یا خیر. پاسخ این سوال منفی است چرا که افراد گروه کنترل به علت دیابت یا فشار خون بالا در بیمارستان بستری شده بودند. حال باید از خود بپرسید آیا این ویژگی خاص (دیابت یا فشار خون بالا) می‌تواند پیامد مطالعه (میزان مصرف قرصهای ضدبارداری) را تحت تأثیر قرار دهد. پاسخ این سوال مثبت است. از آنجائیکه عده زیادی معتقدند که قرصهای ضدبارداری می‌تواند فشار خون و قند خون را افزایش بدهد، پزشکان معمولاً از تجویز قرصهای ضدبارداری به افراد مبتلا به دیابت و فشار خون بالا خودداری می‌کنند. و این ویژگی گروه کنترل، باعث می‌شود که مواجهه با قرصهای ضدبارداری در گروه کنترل کمتر از حد مورد انتظار باشد. بنابراین این روش انتخاب افراد برای گروههای مطالعه و کنترل بالقوه احتمال تورش انتخاب را بالا می‌برد.

تورش انتخاب همانطور که گفته شد می‌تواند در مطالعات همگروهی نیز اتفاق بیفتد مثال زیر را در نظر بگیرید:

اثر مصرف سیگار بر روی ایجاد انفارکتوس قلبی، از طریق انتخاب ۱۰۰۰۰ مصرف کنندهٔ سیگار میانسال و ۱۰۰۰۰ مصرف کنندهٔ پیپ میانسال که از سیگار استفاده نمی‌کنند، مورد بررسی قرار گرفته است. هر دو گروه برای ده سال مورد پیگیری قرار گرفتند پژوهشگران دریافتند که بروز موارد جدید سکته قلبی[1] در کسانی که سیگار می‌کشیدند ۴٪ در ده سال بوده است در حالی که این عدد در کسانی که پیپ می‌کشیدند ولی از سیگار استفاده نمی‌کردند ۷٪ بوده است. از آنجائیکه تفاوت بین دو گروه از نظر آماری معنادار بود پژوهشگران نتیجه گرفتند که افراد سیگاری از خطر کمتری برای ابتلا به MI نسبت به افراد غیرسیگاری که پیپ می‌کشند برخوردار هستند.

همانطور که مشاهده می‌کنید با وجود اینکه این اختلاف از نظر آماری معنادار است ولی با نتایج بسیاری از مطالعات در تعارض است حال ببینیم آیا اختلاف مشاهده شده می‌تواند مربوط به تورش انتخاب باشد.

1- Myocardial infarction

اینکه آقایان اکثریت افرادی که پیپ می‌کشند را تشکیل می‌دهند و آقایان میانسال نسبت به خانمهای میانسال بیشتر به MI مبتلا می‌شوند دو واقعیت مورد قبول است. با در نظر داشتن این دو واقعیت اولین سوال این است که آیا گروه مطالعه و کنترل از این نظر با هم متفاوت هستند؟ پاسخ به این سوال مثبت است چرا که آقایان اکثریت افرادی که پیپ می‌کشند را تشکیل میدهند در حالی که خانمهای سیگاری نسبت به خانمهایی که پیپ مصرف می‌کنند بسیار بیشتر هستند. سوال دومی که مطرح می‌شود اینست که آیا این تفاوت در گروه کنترل و مطالعه می‌تواند بر روی پیامدهای اندازه‌گیری شده تأثیر بگذارد؟ پاسخ این سوال نیز مثبت است چرا که آقایان در معرض خطر بیشتری برای ابتلا به MI هستند. بنابراین مشاهده می‌کنید که تورش انتخاب در نتایج این مطالعه موثر بوده است.

بهترین روش برای جلوگیری از تورش انتخاب، تخصیص تصادفی واقعی است که در کارآزمایی‌های تصادفی شده از آن استفاده می‌شود. هدف از تخصیص تصادفی، بوجود آوردن گروههایی است که از نظر وجود هر عامل مداخله‌گر شناخته شده یا شناخته نشده‌ای قابل مقایسه و معادل باشند. موفقیت در رسیدن به این هدف به دو مانور مرتبط با هم بستگی دارد.

اول اینکه توالی تخصیص افراد به گروههای مختلف باید غیرقابل پیش‌بینی باشد. این کار از طرق مختلف از جمله استفاده از الگوریتم‌های رایانه‌ای و استفاده از جداول تولید کنندهٔ اعداد تصادفی انجام می‌شود.

دوم اینکه این توالی غیرقابل پیش‌بینی باید از دید پژوهشگرانی که افراد را وارد مطالعه می‌کنند، مخفی باشد. به این فرایند در اصطلاح[1] مخفی بودن تخصیص[2] گفته می‌شود.

نکته بسیار مهمی که باید به آن توجه کنید تفاوت بین کور بودن یا مخفی بودن تخصیص با کور بودن[3] بیماران، ارائه دهندگان مراقبت، افرادی که پیامدها را ارزیابی می‌کنند و تحلیلگران پژوهش است. این تفاوت عمده سه علت دارد:

1- Randomization blinding
2- Allocation concealment
3- Blinding

۱ مخفی کردن تخصیص افراد به منظور کاهش تورش انتخاب صورت می‌گیرد ولی کور بودن (که بعد از تخصیص مداخله معنی پیدا می‌کند)، تورش عملکرد[1] و تورش سنجش[2] را کاهش می‌دهد. (در ادامه به آن پرداخت شده است).

۲ از دیدگاه عملی، مخفی نگاه داشتن تخصیص افراد به گروههای مطالعه تا لحظهٔ ارائه مداخله همیشه امکانپذیر است ولی امکان کور بودن بعد از ارائه مداخله در بعضی موارد مثلاً در مواردی که یک درمان جراحی با یک درمان دارویی مقایسه می‌شود، وجود ندارد.

۳ کنترل تورش انتخاب به کل کارآزمایی و در نتیجه تمام پیامدهای مقایسه شده مربوط می‌شود در حالی که کنترل تورش سنجش معمولاً مربوط به یک پیامد خاص است و ممکن است در مورد همهٔ پیامدها موفقیت‌آمیز نباشد.

بنابراین کور بودن تا لحظهٔ تخصیص مداخله و کور بودن از لحظه تخصیص مداخله به بعد تورش‌های مختلفی را تحت تأثیر قرار می‌دهند، به اجزاء مختلفی از مطالعه مربوط می‌شوند و در اجرا کاملاً متفاوت هستند.

دانستن توالی تخصیص می‌تواند باعث ورود انتخابی بیماران براساس عوامل موثر در پیش‌آگهی آنها شود.

از این طریق بیمارانی که طبق توالی باید به گروه مداخله وارد شوند به علت اینکه ممکن است برای این گروه نامناسب پنداشته شوند (به علت عوامل موثر در پیش‌آگهی) از گروه مداخله حذف می‌شوند و بیمارانی که برای گروه مداخله مناسب تشخیص داده می‌شوند با آگاهی قبلی وارد این گروه می‌شوند.

مطلوبترین شکل کار این است که فرایند فراخوانی و انتخاب افراد برای گروههای مورد مطالعه نسبت به هر اثری که افراد مسئول فراخوانی شرکت‌کنندگان می‌توانند داشته باشند غیرقابل نفوذ باشد. این کار از راههای مختلف قابل انجام است که در بخش ارزیابی نقادانه شواهد مربوط به درمان توضیح داده شده است.

1- Performance bias
2- Detection bias

شواهد تجربی نشان می‌دهد فقدان مخفی بودن کافی انتخاب تصادفی با وجود تورش ارتباط دارد.[1] در حقیقت مخفی بودن تخصیص از سایر جنبه‌های تخصیص تصادفی از جمله تولید یک توالی غیرقابل پیش‌بینی اهمیت بیشتری در کاهش تورش انتخاب دارد، در نتیجه ارزیابی کارآزمایی‌ها براساس این معیار اهمیت ویژه‌ای پیدا می‌کند.

۲/ تورش عملکرد

تورش عملکرد به تفاوتهای نظام‌مندی که در مراقبتهای ارائه شده به گروههای مورد مقایسه به جز مداخلهٔ تحت مطالعه، وجود دارد مربوط می‌شود. برای پیشگیری از این تفاوتهای ناخواسته در مراقبت و از بین بردن اثر دارونما بهتر است افرادی که مراقبت را ارائه می‌کنند و بیماران، از اینکه کدام گروه مداخلهٔ آزمایشی را دریافت می‌کند آگاه نباشد[2] . بعضی شواهد نشان می‌دهند که این نوع ناآگاهی یا کور بودن در محافظت در مقابل تورش اهمیت دارد.[3] مطالعات نشان می‌دهند آلودگی[4] (ارائه مداخله به گروه کنترل) و هم درمانی[5] (ارائه مراقبتهای ناخواسته و اضافی به هر کدام از گروههای مقایسه) می‌تواند بر روی نتایج موثر باشد[6] به علاوه شواهدی موجود است که نشان می‌دهد شرکت کنندگانی که از گروه خود (کنترل یا مداخله) آگاهی دارند، علائم بیشتری را گزارش می‌کنند که منجر به تورش در نتایج می‌شود[7].

ارزیابی تورش عملکرد در مطالعات شبه تجربی باز هم مشکل‌تر است چرا که علاوه بر ضروری بودن سنجش مواجهه با مداخله مورد نظر باید اطمینان حاصل کنیم هیچ تفاوتی بین گروههای مقایسه از نظر مواجهه با سایر عواملی که می‌توانند بر روی پیامدها موثر باشند، وجود ندارد. از طرف دیگرکور بودن معمولاً در مطالعات شبه تجربی عملی نیست، مثلاً در مطالعات مورد ـ شاهدی بیماران قبلاً پیامد را تجربه کرده‌اند. در این گونه مطالعات باید در نظر بگیریم آیا آگاهی از نوع گروه

1- Chalmers 1983, Schulz 1995, Moher 1998
2- masking, blinding
3- Karlowski 1975, Colditz 1989, Schulz 1995
4- contamination
5- Cointervention
6- Sackett 1979, CCSG 1978
7- Karlowski 1975

مقایسه می‌تواند (مثلاً از طریق گزارش علائم بیشتر) روی نتایج موثر باشد یا خیر.

۳ تورش پیگیری[1]

این تورش به تفاوت نظام‌مند بین گروههای مقایسه در ریزش شرکت‌کنندگان در مطالعه اشاره دارد. کاستی‌های مربوط به پیگیری دو بعد دارد.

– بعد زمانی: آیا طول دورهٔ پیگیری از نظر پاتوفیزیولوژی و تجربی برای ایجاد اثر مداخلات و ظهور پیامدهای مورد نظر کافی بوده است؟

– بعد تعدادی: فقدان پیگیری کامل می‌تواند منجر به حذف تعدادی از بیماران بعد از تخصیص به گروههای مقایسه شود. که این امر به نوبهٔ خود به تورش پیگیری می‌انجامد. ریزش‌های که در دورهٔ پیگیری بوجود می‌آید به بیمارانی اشاره دارد که در مرحله‌ای از طول دورهٔ پیگیری از دسترس پژوهشگران خارج می‌شوند، این امر می‌تواند به علت امتناع آگاهانهٔ بیماران از ادامهٔ مشارکت در پژوهش[2]، تصمیم بالینی پژوهشگران برای قطع مداخلهٔ تخصیص یافته و نهایتاً به خاطر عدم دسترسی به بیماران مثلاً به علت مسافرت و ترک محل پژوهش بدون اطلاع قبلی باشد.

بیمارانی که بعد از مداخله از مطالعه حذف می‌شوند به ندرت نمایندهٔ واقعی کلیهٔ شرکت کنندگان در مطالعه هستند. برای مثال ممکن است بیماران به خاطر وخامت حاد بیماری و یا ایجاد اثرات جانبی[3] شدید، از دسترس پژوهشگران برای پیگیری خارج شده باشد. از طرف دیگر بیمارانی که التزام کافی برای ادامهٔ درمان ندارند عمدتاً از نظر عوامل موٴثردر پیش‌آگهی با سایر بیماران متفاوت هستند. بنابراین اصولاً تمام بیمارانی که وارد گروههای مطالعه شده‌اند باید در تحلیل داده‌ها وارد شوند و در گروههایی که بدواً به آنها

1- Attrition bias or Exclusion bias
2- Drop out
3- Side effects

وارد شده بودند مورد تحلیل قرار بگیرند این امر مستقل از التزام بیماران به رعایت پروتکل مطالعه است بنابراین علاوه بر طول دورهٔ پیگیری، نسبتی از بیماران که در تحلیل پژوهش وارد نشده‌اند نیز باید در گزارش مطالعه ذکر شود (ارزیابی گزارش مقاله) تا بتوان در مورد احتمال تورش پیگیری قضاوت کرد. با این حال شواهد نشان می‌دهد آنچه در مورد تورش پیگیری درگزارش مطالعه ذکر می‌شود و یا به طور شایعتر صرفاً به آن اشاره می‌شود رابطه ثابت و پایداری با احتمال واقعی تورش مورد نظر ندارد[1] . لذا بهتر است در استفاده از شرایط پیگیری به عنوان معیار روایی در مطالعات، کمی محطاتانه‌تر عمل کنید.

۴ | تورش سنجش

این تورش به تفاوت نظام‌مند بین گروههای مقایسه، در ارزیابی پیامدها برمی‌گردد.

اطلاعاتی که برای سنجش پیامدها استفاده می‌شوند معمولاً از دو منبع بدست می‌آیند.

۱ حافظهٔ شرکت‌کنندگان در مطالعه و یا استفاده از داده‌ها موجود در پروندهٔ قبلی آنها

۲ سنجش‌هایی که پژوهشگران انجام می‌دهند.

اطلاعاتی که از حافظهٔ افراد شرکت‌کننده در مطالعه بدست می‌آید در معرض دو گونهٔ خاص از تورش سنجش قرار دارد: تورش یادآوری و تورش گزارش‌دهی

تورش یادآوری[2] این تورش به نقایصی که ممکن است در حافظه شرکت کنندگان وجود

1- Schulz 1995
2- Recall bias

داشته باشد اشاره دارد، خصوصاً نقایصی که باعث می‌شود یک گروه کمتر از گروه دیگر اتفاقات قبلی را به خاطرآورد. مثال فرضی زیر را در نظر بگیرید:

در یک مطالعه مورد ـ شاهدی برای کشف علل اسپینابیفیدا[1]، ۱۰۰ مادری که نوزادان آنها با بیماری متولد شده بودند و ۱۰۰ مادری که نوزاد آنها بدون این بیماری بودند مورد بررسی قرار گرفتند. از بین مادرانی که فرزند مبتلا به اسپینابیفیدا داشتند ۵۰٪ سابقه گلو درد در طی حاملگی را گزارش کردند که این میزان در بین مادرانی که فرزندان غیرمبتلا به اسپینابیفیدا داشتند ۵٪ بود.

بنابراین پژوهشگران نتیجه گرفتند که بین گلودرد در دورهٔ حاملگی و اسپینابیفیدا رابطه وجود دارد.

قبل از اینکه نتیجه این مطالعه را قبول کنید، ممکن است از خود بپرسید آیا تورش یادآوری می‌تواند این نتایج را توجیه کند؟ ممکن است بگویید مادرانی که ضربهٔ روحی ناشی از داشتن یک نوزاد مبتلا به اسپینابیفیدا را تجربه کرده‌اند حافظه خود را با شدت بیشتری جستجو می‌کنند و از این طریق ممکن است وقایعی را به خاطر آورند که مادران دیگر معمولاً به خاطر نمی‌آورند.

البته نمی‌توان مطمئن بود که این تورش نتایج را تحت تأثیر قرار می‌دهد بلکه فقط می‌توان گفت شرایط ایجاد چنین تورشی وجود داشته است. بنابراین حداقل قسمتی از نتایج این مطالعه را می‌توان به تورش یادآوری نسبت داد.

تورش گزارش‌دهی[2] زمانی در مطالعه اتفاق می‌افتد که یک گروه نسبت به گروه دیگر تمایل بیشتری برای گزارش کردن آنچه به خاطر می‌آورند، داشته باشند. این تورش نیز مانند تورش یادآوری می‌تواند روایی سنجش پیامدها را مخدوش کند.

به مثال زیر دقت کنید:

1- Spina bifida
2- Reporting bias

یک مطالعه مورد ـ شاهدی برای تعیین رابطهٔ بین گنوره و شرکای جنسی متعدد انجام شده است در این مطالعه صد خانمی که تشخیص گنوره در آنها به تازگی داده شده است با صد خانم دیگر که گنوره نداشتند در همان کلینیک مورد مقایسه قرار گرفتند. به خانمهایی که تشخیص گنوره داشتند اطلاع داده شده بود که پیشگیری از عواقب جدی بیماری صرفا از طریق یافتن و درمان شریک جنسی آنها امکانپذیر بوده است. هر دو گروه از خانمها دربارهٔ تعداد شرکای جنسی که در ۲ ماه اخیر داشتند مورد سوال قرار گرفتند گروهی از خانمها که مبتلا به گنوره بودند تعداد شرکای جنسی خود را به طور متوسط ۴ برابر تعداد شرکای جنسی در خانمهای غیرمبتلا به گنوره گزارش کردند. بنابراین پژوهشگران نتیجه گرفتند که به طور متوسط تعداد شرکای جنسی در خانمهای مبتلا به گنوره ۴ برابر خانمهای غیرمبتلا به گنوره است. در مورد این مطالعه می‌توان گفت خانمهایی که به گنوره مبتلا شده‌اند اجبار بیشتر و در نتیجه شک و دودلی کمتری برای گزارش کردن تعداد شرکای جنسی خود احساس می‌کردند تا خانمهایی که به بیماری مبتلا نشده‌اند. تورش گزارش‌دهی در مواقعی که اطلاعات مورد نیاز شخصی یا خصوصی هستنداحتمال بروز بیشتری دارد. بعلاوه اگر گروهی نسبت به گروه دیگر تحت فشار مضاعفی برای گزارش دقیق اتفاقات قبلی باشد احتمال این تورش بیشتر است.

همانطور که متوجه شده‌اید احتمال ایجاد این دو نوع خاص از تورش سنجش در مطالعات شبه تجربی و خصوصاً مطالعات مورد ـ شاهدی بیشتر است چرا که در این نوع مطالعات شرکت‌کنندگان قبلاً از اتفاق افتادن یا نیفتادن پیامد مورد نظر آگاهی دارند.

هرگاه سنجش پیامدها[1] به تفسیر ذهنی[2] داده‌ها وابسته باشد احتمال نوع خاصی از تورش سنجش تحت عنوان تورش پژوهشگر[3] بیشتر می‌شود. تورش پژوهشگر از این ویژگی انسانها منشاء می‌گیرد که انسان معمولاً چیزی که انتظار دیدن آن را

1- Measurement of outcome
2- Subjective
3- Investigator bias

دارد می‌بیند. راههایی برای بازشناخت و تصحیح اثرات این ویژگی روانشناختی پژوهشگران وجود دارد. یکی از این راهها بی‌خبر نگه داشتن افرادی که پیامدها را ارزیابی می‌کنند نسبت به نوع گروههای مورد مقایسه است.

کور بودن افرادی که پیامد را ارزیابی می‌کنند با کور بودن بیماران و ارائه‌دهندگان مداخله از این جهت متفاوت است که مورد اول تورش سنجش را کاهش می‌دهد ولی کور بودن بیمار و ارائه‌دهندهٔ مداخله باعث کاهش تورش عملکرد می‌شود:

به مثال زیر دقت کنید: در یک مطالعه برای سنجش عوارض مصرف داروهای ضدالتهابی غیراستروئیدی[1] بر روی دستگاه گوارش، پژوهشگران که خود، پزشک بیماران مورد مطالعه نیز بودند از بیماران پرسیدند آیا یکی از NSAID های مصرف شده همراهی بیشتری با علائم گاستریت داشته است؟ بعد از اینکه از تمام بیماران راجع به علائمشان سوال شد. پژوهشگران دریافتند که بین دو گروه تفاوت معناداری در ایجاد گاستریت وجود دارد و دو NSAID مصرف شده فرکانس متفاوتی از علائم گاستریت ایجاد می‌کنند.

در این مطالعه پژوهشگرانی که پیامدها را ارزیابی می‌کردند از اینکه هر بیمار چه دارویی دریافت می‌کرده است آگاهی داشتند به علاوه آنها علائم ذهنی (از جمله حالت تهوع، درد معده و سوء هاضمه) را برای تعیین وجود گاستریت مورد ارزیابی قرار داده بودند. در چنین شرایطی کور بودن فرد ارزیابی کنندهٔ پیامد اهمیت حیاتی می‌یابد.

کور بودن افراد ارزیابی کنندهٔ پیامدها نسبت به گروههای مقایسه در کارآزماییهای بالینی نیز منطقاً اهمیت زیادی دارد. و همانطور که گفته شد، این کور بودن در مطالعاتی که از سنجهٔ پیامد ذهنی[2] استفاده می‌کنند اهمیت بیشتری پیدا می‌کند[3] با این حال بعضی شواهد تجربی در نشان دادن رابطهٔ کور بودن ارزیابی پیامدها و نتایج مطالعه موفق نبوده‌اند.[4]

1- NSAIDs
2- Subjective outcome measure
3- Colditz 1989, Schulz 1995, Karlowski 1975
4- Reitman 1988

محدودیت‌های ارزیابی کیفیت

حال که با اجزاء کیفیت آشنا شدید باید بدانید که ارزیابی کیفیت مطالعات همیشه به طور کامل امکانپذیر نیست. دو مشکل اساسی برای ارزیابی روایی مطالعات وجود دارد. اولین مشکل ناکافی بودن گزارش مطالعه راجع به هر یک از اجزاء کیفیت در مطالعه است.[1]

محدودیت دوم، عدم وجود شواهد تجربی کافی برای نشان دادن ارتباط بین معیارهایی که تصور می‌شود در سنجش روایی مطالعه نقش دارند با نتایج نهایی مطالعه است. قطعاً نیاز به انجام پژوهشهای متعددی وجود دارد تا تعیین شود چه معیارهایی برای سنجش روایی مطالعات به عنوان عوامل تعیین کننده نتایج مطالعات اهمیت واقعی دارند.

به طور خلاصه ارزیابی روایی مطالعات تا حدود زیادی به قضاوت فرد ارزیابی کننده بستگی دارد. نیاز به قضاوت خواننده در ارزیابی کیفیت مطالعه شبه تجربی و غیرتجربی (رجوع کنید به فصل روش‌شناسی پژوهش) با توجه به طبیعت این مطالعات باز هم بیشتر است. بنابراین اتخاذ رویکردی محتاطانه‌تر در این شرایط توصیه می‌شود.

دقت آماری مطالعه

اندازهٔ P value و طول بازهٔ اطمینان و یا به طور کلی دقت تخمینی که از اثر مشاهده شده در مطالعه به دست می‌آید نیز در سنجش قدرت شواهد اهمیت دارد. با اینکه P value و فاصلهٔ اطمینان دو مفهوم به هم مرتبط هستند، ارزش واقعی P (نه صرفاً اینکه P value از حدی قراردادی کمتر باشد) اطلاعات بیشتری نسبت به بازهٔ اطمینان درباره دقت آماری بدست می‌دهد. (درباره تعریف P value، فاصلهٔ اطمینان و نحوهٔ محاسبه آنها در بخش روش‌شناسی پژوهش توضیح داده شده است).

1- Begg 1996, Sort 1994, Schulz 1994, WGRR 1994

به طور کلی P value نشانگر این است که آیا وجود شانس می‌تواند تفاوت مشاهده شده در مطالعه را توجیه کند. در حالی که هیچ ارزش جادویی برای قضاوت در مورد P value وجود ندارد، هر چه مقدار P value کمتر باشد با قطعیت بیشتری می‌توان گفت که اثر مشاهده شده از شانس منشاء نگرفته است.

اندازهٔ اثر [1]

دومین بعد از ابعاد شواهد، میزان بزرگی اثر مداخله یا عامل مورد بررسی در مطالعه است. اندازهٔ اثر بیان می‌کند تا چه حد نتایجی که در مطالعه بدست آمده‌اند از نظر بالینی اهمیت دارند. فرض کنید مطالعه‌ای در دست داریم که در آن اثر یک داروی ضدفشار خون با دارونما مقایسه شده است. این مطالعه از نظر معیارهای سنجش قدرت شواهد (سطح مطالعه، کیفیت و دقت) بالاترین اعتبار را کسب کرده است و نشان می‌دهد تجویز این داروی ضدفشار خون با P value ۰/۰۰۱ فشار خون را به اندازه ۳ میلی‌متر جیوه کاهش می‌دهد (P value ۰/۰۰۱ نشان می‌دهد نتایج مطالعه از نظر آماری معنی‌دار است). با دقت در این مثال می‌بینید که برای بکارگیری نتیجهٔ مطالعات علاوه بر ارزیابی قدرت مطالعه توجه به اندازهٔ اثر آن نیز اهمیت زیادی دارد. زیرا اندازهٔ اثر مطالعه از نظر بالینی برای ما اهمیتی نداشته باشد هر چقدر هم کیفیت آن بالا باشد ارزشی نخواهد داشت. در این مثال کاهش فشار خون به اندازه ۳mm/Hg از نظر بالینی ارزش ندارد.

سنجه‌های اندازه‌گیری اندازهٔ اثر

انتخاب سنجهٔ مناسب برای اندازه‌گیری اندازهٔ اثر به عوامل متعدد بستگی دارد. مهمترین این عوامل مقیاس سنجش[2] است. اثر

1- Size of the effect
2- Scale of measurement

بعضی از پیامدها از طریق مقیاس پیوسته[1] اندازه‌گیری می‌شود (مثل فشار خون، کلسترول سرم، کیفیت زندگی[2] در حالی که سنجش سایر پیامدها از طریق مقیاس دوتایی[3] انجام می‌شود (مثل بهبود یا عدم بهبود، مرگ یا حیات، عود یا عدم عود). به چند نمونه از سنجه‌هایی که بطور شایع برای اندازه‌گیری پیامدهای دوتایی و پیوسته استفاده می‌شوند توجه کنید.

پیامدهای پیوسته

- تفاوت بین میانگین‌های دو گروه: تفاوت بین گروه مطالعه و گروه کنترل (گروههای مقایسه) در ارزش میانگین‌های متغیر پیامد.
- تفاوت استاندارد شده: تفاوت بین میانگین‌های گروه مطالعه و گروه کنترل برای هر مطالعه که از طریق تخمین انحراف معیار سنجش‌ها استاندارد شده است. این کار اثر واحد سنجش را از بین می‌برد ولی در عوض تفسیر نتایج را پیچیده‌تر می‌کند.

پیامدهای دوتایی

- تفاوت خطر[4] (RD) تفاوت (مطلق) نسبت پیامد بین دو گروه کنترل و مطالعه.

اگر پیامد مورد نظر نمایندهٔ یک واقعهٔ نامطلوب (مثلاً مرگ) باشد و تفاوت خطر منفی شود (زیر صفر) نشان دهندهٔ آن است که درمان خطر را کاهش داده است. در این شرایط تفاوت خطر بدون در نظر گرفتن علامت، کاهش خطر مطلق[5] (ARR) نامیده می‌شود.

- خطر نسبی[6] یا نسبت خطر[7](RR) : نسبت وقوع پیامد در گروههای مطالعه و کنترل.

این سنجه نشان دهندهٔ خطر وقوع پیامد در گروه مطالعه نسبت به میزان آن در گروه کنترل است.

1- Continuous scale
2- Quality of life
3- dichotomous or binary scale
4- Risk difference
5- Absolute Risk Reduction
6- Relative risk
7- Risk ratio

برای یک پیامد نامطلوب اگر خطر نسبی کمتر از ۱ باشد نشان دهندهٔ آن است که مداخلهٔ ما خطر را کاهش می‌دهد. مکمل آن (خطر نسبی - ۱) کاهش خطر نسبی[۱](RRR) نیز به طور شایع مورد استفاده قرار می‌گیرد.

- نسبت شانس[۲](OR) : نسبت شانس پیامد در گروه مطالعه به شانس متناظر در گروه کنترل.

برای پیامدهای نامطلوب نسبت شانس کمتر از ۱ نیز نشان دهندهٔ کاهش خطر توسط مداخلهٔ انجام شده است.

در برخی مطالعات (مثل مطالعات مورد شاهدی مبتنی بر جمعیت[۳] نسبت شانس تخمین مناسب و خوبی از خطر نسبی است. (در صورتی که پیامد مورد نظر شایع باشد نسبت شانس دیگر تخمین خوبی از خطر نسبی نخواهد بود).

- نسبت مخاطره[۴](HR): نسبت مخاطره در گروه مطالعه و گروه کنترل (وقتی زمان وقوع پیامد مورد نظر مشخص باشد)

مخاطره[۵] در این تعریف احتمال وقوع پیامد در زمان مشخص t است با این فرض که پیامد تا قبل از زمان t اتفاق نیفتاده باشد. برای یک پیامد نامطلوب، مقادیرکمتر از ۱ نشان می‌دهد که مداخلهٔ ما خطر وقوع آن پیامد را کاهش می‌دهد.

- تعداد لازم برای درمان [۶] (NNT): تعدادی از بیماران که باید تحت درمان قرار بگیرند تا از یک مورد وقوع پیامد مورد نظر جلوگیری شود. این سنجهٔ پیامد از معکوس کردن تفاوت خطر (RD) بدون در نظر گرفتن علامت منفی بدست می‌آید.

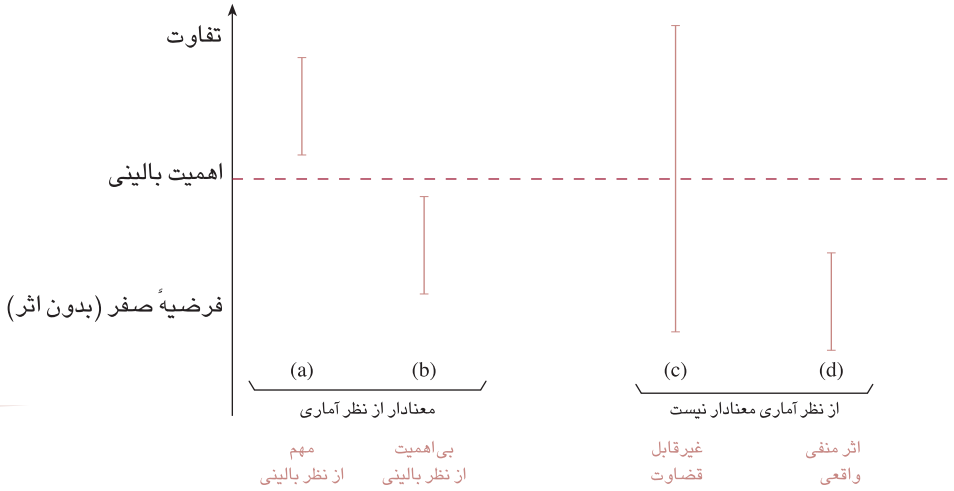
چنانچه مداخله ما باعث افزایش خطر وقوع یک پیامد مضر شود، آنگاه به معکوس تفاوت خطر (RD) تعداد لازم برای آسیب[۷] گفته می‌شود.

رابطهٔ معناداری آماری و اندازهٔ اثر

همانطور که گفته شد در بعضی مطالعات ممکن است اثراتی که از نظر آماری معنی‌دار هستند آنقدر کوچک باشند که از

1- Relative Risk Reduction)
2- Odds Ratio
3- Population-based case-control studies
4- Hazard Ratio
5- Hazard
6- Number Needed to Treat
7- Number Needed to Hram

نظر بالینی اهمیت نداشته باشند. بنابراین ارزیابی حجم اثر نیز مانند معنادار بودن آماری اهمیت دارد. در مورد پیامدهای دوتایی اگر بازهٔ اطمینان ۹۵٪ فقط شامل اثرات درمانی باشد که از نظر بالینی اهمیت دارند آنگاه می‌توان گفت مداخله مفید بوده است. (شکل ۱)



شکل ۱

خطوط عمودی نشاندهندهٔ فاصله اطمینان ۹۵ درصد برای تفاوت بین گروههای کنترل و مطالعه است.

(a) تفاوت مشاهده شده هم از نظر آماری معنی‌دار است و هم از نظر بالینی اهمیت دارد.

(b) تفاوت مشاهده شده از نظر آماری معنی‌دار است ولی از نظر بالینی اهمیت ندارد.

(c) تفاوت مشاهده شده از نظر آماری معنی‌دار نیست و از اهمیت بالینی نامشخصی برخوردار است.

(d) تفاوت مشاهده شده نه تنها از نظر آماری معنی‌دار نیست بلکه از نظر بالینی نیز بی‌اهمیت است.

شواهدی که مانند مطالعهٔ (a) هم از نظر آماری معنی‌دار هستند و هم از نظر بالینی اهمیت دارند (بازهٔ اطمینان ۹۵ درصد آنها فقط شامل اثرات مهم از نظر بالینی است) برای ما ایده‌آل هستند.

مواردی مثل مطالعات (b) و (d) نیز نتایج آگاهی دهنده‌ای دارند (نشان می‌دهند که مداخلهٔ انجام شده نمی‌تواند اثری را که از نظر بالینی اهمیت دارد تولید کند)، ولی باید از مواردی مثل مطالعه (c) اجتناب کرد چرا که در آنها بازهٔ اطمینان ۹۵ درصد آنقدر وسیع است که هم با اهمیت بالینی تطابق دارد و هم فقدان اثر درمانی را شامل می‌شود. بازهٔ اطمینان خیلی وسیع معمولاً فاقد ارزش است مثلاً می‌توانیم بگوییم به احتمال ۱۰۰ درصد میانگین نمرات دانش‌آموزان یک کلاس در درس زیست‌شناسی بین ۰ تا ۲۰ است. درحالی که دقت این اطلاع ۱۰۰ درصد است ولی در عمل هیچ ارزشی ندارد.

مربوط بودن شواهد

آخرین بعد شواهد مربوط بودن شواهد[1] است.

آیا در مطالعه پیامدهایی که برای بیمار حائز اهمیت است یا مورد علاقهٔ بیماران می‌باشد انتخاب شده است؟ آیا سایر انواع پیامدها بخوبی با پیامدهایی که از نظر بیمار حائز اهمیت هستند رابطه (علّی) دارند؟ آیا پیامدها نشان دهندهٔ اثرات کوتاه مدت هستند یا بلند مدت؟

توجه داشته باشید مربوط بودن شواهد علاوه بر تناسب پیامدها مقولاتی چون میزان تکرارپذیری مداخلات در شرایط دیگر و تعمیم‌پذیری یافته‌های مطالعه را نیز در بر می‌گیرد که در ارزیابی نقادانه شواهد بحث نمی‌شود. در اینجا منظور از مربوط بودن شواهد بیش از همه مربوط بودن و تناسب پیامدهای مطالعه است.

1- Relevance of the evidence

تناسب پیامدها:

پیامدهایی که در سنجش اثرات مداخلات استفاده می‌شود به سه گروه تقسیم می‌شود.

- پیامد جایگزین[1]
- پیامد مربوط به بالین[2]
- پیامد مربوط به بیمار[3]

معمولاً انواعی از پیامدها در مطالعات و کارآزمایی‌های بالینی استفاده می‌شود که سنجش آنها راحت‌تر باشد و بتواند تغییرات و تفاوتها را در دوره زمانی نسبتاً اندکی نشان بدهد. این پیامدها می‌توانند از نظر بیماران مهم یا بی‌اهمیت باشند. به تعریف انواع مختلف پیامد توجه کنید:

- **پیامد جایگزین:** سنجش‌های آزمایشگاهی یا علایم فیزیکی[4] که به عنوان جایگزینی برای پیامدهای معنادار بالینی (که به طور مستقیم نحوه زندگی، عملکرد و احساسات بیمار را می‌سنجند) به کار می‌رود. تغییراتی که توسط مداخله، در این پیامدها اعمال می‌شود به عنوان بازتابی از تغییر در پیامدهای معنادار بالینی تلقی می‌گردد. به عنوان مثال سطح کلسترول خون ممکن است بعنوان پیامد جایگزین برای وجود بیماریهای عروق کرونر در نظر گرفته شود.

- **پیامد بالینی:** پیامدهایی که بر مبنای بیماری تحت مطالعه تعریف می‌شوند. برای مثال: بقا در سرطان، وقوع شکستگی‌های مهره‌ای در درمان استئوپروز، بهبود زخم، مسافت راه رفتن و ...

1- Surrogate outcome
2- Clinical relevant outcome
3- Patient relevant outcome
4- Physical sign

- **پیامد مربوط به بیمار:** پیامدهایی که برای بیماران و خانواده آنها اهمیت دارد. این پیامدها باید به گونه‌ای باشد که بیمار بتواند آنها را تجربه کند. (مثل کیفیت زندگی، بازگشت به ظرفیت عملکردی نرمال)

مثالهایی از هر کدام از پیامدها در جدول یک ارائه شده است.

- چه هنگام استفاده از پیامدهای جایگزین سودمند است؟

شواهد نشان می‌دهد تعداد محدودی از پیامدهای جایگزین حائز شرایط لازم برای به کارگیری در مطالعات هستند.

این شرایط عبارتند از:

۱ پیامد جایگزین باید یک متغیر فیزیولوژیک باشد.

۲ پیامد جایگزین باید بتواند با پیامدهای بالینی یا پیامدهای مربوط به بیمار همراهی نشان دهد.

۳ پیامد جایگزین نیاز به یک مبنای بیولوژیک یا پاتوفیزیولوژیک دارد تا بتوانیم بپذیریم که به عنوان یکی از علل تعیین کننده پیامد بالینی در بیماری تحت مطالعه مطرح است.

مثال:

در یک مطالعه بین فشار خون بالا (به عنوان متغیّر جایگزین) با میزان شکستگی استخوان همراهی و ارتباط دیده شده است، ولی این ارتباط مبنای بیولوژیک و منطقی ندارد، درحالیکه میزان ابتلا به پرفشاری خون و پوکی استخوان (که منجر به افزایش میزان شکستگی می‌شود) هر دو با افزایش سن افزایش می‌یابند.

ترکیب دو مورد آخر بیان می‌کند که پیامد جایگزین باید پیش‌بینی کننده بسیار خوبی از متغیرهای بالینی باشد. این معیار برای بسیاری از پیامدهای جایگزین که رابطه آماری (همبستگی) بدون رابطه علّی با متغیرهای

بالینی دارند مفروض است. با این حال برخی از پیامدهای جایگزین ارزش مداخله ندارند و صرفاً به عنوان یک عامل تعیین خطر به کار می‌روند.

فرضاً پرفشاری خون، علاوه بر اینکه خطر MI را بالا می‌برد، عاملی است که با مداخله بر روی آن می‌توان بر میزان وقوع MI تأثیر گذاشت، ولی پیامدی مانند گزانتوما[1] که ممکن است در اثر سطوح بالای کلسترول خون بوجود آید، صرفاً به عنوان زنگ خطری برای وقوع MI، برای پزشک کاربرد دارد و برداشتن آن به روشهای جراحی، خطر MI را پایین نمی‌آورد.

ممکن است این رابطه آنقدر اختصاصی نباشد که بتوان اطمینان حاصل کرد هر تغییری در پیامد جایگزین در متغیرهای بالینی نیز تغییر ایجاد می‌کند. مثلاً تغییر در فشار خون نمی‌تواند خطر بیماری عروق کرونر را بدقت پیش‌بینی کند چرا که اتیولوژی آن چند عاملی است.

بنابراین هنگام بررسی پیامدهای جایگزین باید جنبه‌های دیگر پیامد اصلی را نیز مدنظر داشته باشید. به عنوان مثال در مطالعه‌ای نشان داده شده است که داروی «A» به عنوان دارویی مناسب برای کاهش کلسترول مورد استفاده قرار می‌گیرد؛ ولی توجه داشته باشید که صرف کاهش کلسترول خون برای ما ارزشمند نیست و اثر آن در کاهش مرگ و میر در اثر MI، ارزش دارد.

حال در مطالعات دیگر دیده شده است که داروی «A» مرگ و میر را با مکانیسمهای دیگر افزایش می‌دهد؛ آیا باز هم این دارو، برای کاهش خطر MI داروی مناسبی است؟

1- Xanthomas

جدول یک ـ مثالهایی از پیامدهای جایگزین و پیامدهای بالینی و پیامدهای مربوط به بیمار

جدول ۱

بیماری	پیامد جایگزین	پیامد بالینی	پیامد مربوط به بیمار
بیماریهای قلبی عروقی	فشار خون	سکته مغزی، MI، بقا	کیفیت زندگی
HIV/AIDS	شمارش لنفوسیت‌های TC D4	وقایع AIDS ، بقا	کیفیت زندگی
شکستگی	سنجش تراکم استخوان	شکستگی استخوان	شکستگی‌های علامتدار، وضعیت عملکردی
بیماری عروق کرونر	سطح کلسترول خون	MI، بقا	درد، کیفیت زندگی
اوتیت میانی	درمان ضدمیکروبی	درمان بالینی	کری

Source: entracted from Fleming and DeMets 1996

پیامدهایی که در مطالعات بالینی سنجیده می‌شود عمدتاً پیامدهایی هستند که از نظر پژوهشگران و پزشکان اهمیت دارند. علاوه بر آن در مطالعات تمایل زیادی وجود دارد تا از پیامدهایی استفاده شود که عمدتاً در کوتاه مدت وقوع آنها قابل ارزیابی است و سنجش آنها براحتی امکانپذیر است. با این تفاسیر، واضح است که این رویکرد تمام پیامدهایی را که از منظر بیماران مهم هستند شامل نمی‌شود. بنابراین مزایا و آسیب‌های بالقوه مداخلات که برای بیماران اهمیت دارند، یا اندازه‌گیری نمی‌شوند و یا

گزارش نمی‌کردند. پیامدهایی که به طور معمول سنجیده می‌شوند بازتابی از احساسات واقعی بیماران نیستند و میزان رضایت آنها را از وضعیت موجود خود نشان نمی‌دهند. عواملی که بر بهبود کیفیت زندگی و کمیت آن (مثل مرگ و میر) دلالت دارند بیشتر به بیماران مربوط می‌شوند[1] این نوع سنجه‌ها می‌توانند بیانگر تعادل بین مزایا و مضرات مداخلات باشند. برای مثال، یک دارو ممکن است سطح کلسترول خون را بخوبی کاهش دهد (و در نتیجه ریسک بیماریهای کرونری قلب را در آینده کم کند) ولی نتیجه استفاده از آن برای بیمار کاهش کیفیت زندگی بخاطر عوارض جانبی دارو باشد. خصوصاً، برای افرادی که در خطر کمتری برای وقوع پیامد نامطلوبی که مداخله قصد دارد از آن پیشگیری کند، هستند کاهش خطر مطلق پیش‌بینی شده ممکن است آنقدر کوچک باشد که نتواند اثرات سوء درمان را جبران کند.

رتبه‌بندی پیشنهاد شده برای شواهد از نظر میزان مربوط بودن پیامدها به بیماران در ذیل بیان شده است.

۱ شواهدی که در آنها مداخلات از طریق پیامدهای بالینی مربوط به بیماران سنجیده شود. این پیامدها شامل فواید[2] و آسیب‌ها[3]، کیفیت زندگی و میزان بقا است.

۲ شواهدی که در آنها اثر مداخلات از طریق پیامدهای جایگزین سنجیده شود به شرطی که قبلاً ثابت شده باشد که این پیامدها پیش‌بینی کننده خوبی از پیامدهای مربوط به بیماران برای مداخله مشابه هستند.

۳ شواهدی که در آنها از پیامدهای جایگزین ثابت شده استفاده شده است در حالیکه ارزش این پیامدها در مورد مداخله متفاوتی ثابت شده باشد.

۴ شواهدی که در آن از پیامدهای جایگزین اثبات شده استفاده شده است ولی ارزش این پیامدها در مورد مداخله و جمعیت متفاوت به اثبات رسیده است.

1- Hadorm and Baker 1994
2- Benefits
3- Harms

۵ شواهدی که در آن از پیامدهای جایگزین اثبات نشده استفاده شده است.

اگر بخواهیم تناسب سنجه‌های پیامد جایگزین را در نظر بگیریم باید به شرایطی دقت کنیم که این سنجه‌ها قبلاً مورد تأیید قرار گرفته‌اند. برای مثال اگر افزایش تراکم استخوان به عنوان یک پیامد جایگزین مناسب و معقول برای کاهش ریسک شکستگی‌های استئوپروتیک در کارآزمایی یک بیسفوسفونات[1] تأیید شده باشد می‌تواند به عنوان یک پیامد جایگزین برای یک داروی دیگر از همین گروه مطرح باشد. با این حال استفاده از سنجش تراکم استخوان برای یک نوع داروی دیگر با مکانیسم اثر متفاوت مثل درمان جایگزینی هورمون[2] چندان متناسب به نظر نمی‌رسد.

ارزیابی نقادانه شواهد با چه هدفی صورت می‌گیرد؟

حال که با مفهوم ارزیابی نقادانه شواهد آشنا شدید و ابعاد و معیارهای کلی ارزیابی مطالعات را شناختید نوبت آن رسیده است که هدف خود را از انجام فرایند ارزیابی نقادانه تعیین کنید.

همانطور که اشاره شد ارزیابی نقادانه شواهد یکی از مراحل پایه‌ای و اساسی در طبابت مبتنی بر شواهد است و فصل مشترک بسیاری از فعالیتها در این زمینه می‌باشد. با این توضیح سناریوهای زیر را در نظر بگیرید:

فرض کنید به عنوان یک پزشک در مطب خود نشسته‌اید که یک بیمار به شما مراجعه می‌کند. در حین گرفتن شرح حال مختصری از بیمار متوجه می‌شوید که تاکنون با مورد مشابه آن مواجه نشده‌اید و اطلاعات شما برای تصمیم‌گیری درباره بیمار کافی نیست. در حالیکه هنوز مصاحبه شما با بیمار ادامه دارد، بدون اینکه جلب توجه کرده باشید یک جستجوی سریع در پایگاههای داده‌ای الکترونیکی که در رایانه شخصی خود ذخیره کرده‌اید انجام می‌دهید. از بیمار

1- Bisphosphonate
2- Hormone Replacement Therapy (HRT)

می‌خواهید که برای معاینه آماده شود. در همین حین به سرعت عنوان مقالات را بررسی کرده و یکی از آنها را که دقیقاً به سوال شما مربوط است انتخاب می‌کنید با بررسی خلاصه مقاله و ارزیابی نقادانه آن به این نتیجه می‌رسید که کیفیت مطالعه در حد قابول قبولی است. نتیجه مطالعه را می‌خوانید و به سراغ بیمارتان که هم اکنون برای معاینه آماده شده است می‌روید.

حال سناریوی زیر را در نظر بگیرید.

در یک جلسه گزارش صبحگاهی[1] مشغول بحث درباره یک مشکل بالینی خاص هستید. هر کدام از همکاران شما بر اساس جستجوها و مطالعات قبلی خود شواهدی را در تأیید نظرشان بیان می‌کنند. در حالیکه شواهد زیادی راجع به سوال بالینی مورد بحث وجود دارد ولی نتایج آنها بعضاً ضد و نقیض است. از طرف دیگر این شواهد از قدرت و کیفیت‌های متفاوتی برخوردارند. بحث ناتمام می‌ماند و شما به نتیجه خاصی نمی‌رسید. تصمیم می‌گیرید با همکاری دوستان خود یک مقاله مروری نظام‌مند راجع به سوال بالینی مورد نظر تدوین کنید. (مراحل این کار به تفسیر در کتاب چهارم این مجموعه آورده شده است.)

پس از انجام مراحل اولیه و جمع‌آوری مطالعات حال نوبت ارزیابی نقادانه این شواهد است.

سناریوهای فوق دو مثال از شرایطی است که نیاز به ارزیابی نقادانه شواهد در آنها وجود دارد. ولی همانطور که متوجه شده‌اید این شرایط کاملاً با هم متفاوت هستند. در سناریوی اول شما با استفاده از مفاهیم گفته شده و با استفاده از ابزارهایی که برای این منظور طراحی شده‌اند با توجه به سوال بالینی و نوع مطالعه بدست آمده، کیفیت مطالعه را ارزیابی می کنید ولی در سناریوی دوم کار قدری پیچیده‌تر است.

1- Morning report

رویکرد مرور نظام‌مند به ارزیابی نقادانه شواهد:

خوانندگان مقالات مروری نظام‌مند بطور طبیعی علاقه دارند بدانند کیفیت مطالعات وارد شده در مرور نظام‌مند چگونه بوده است و چرا بعضی از مطالعات با اینکه به سوال بالینی مورد نظر پاسخ می‌دادند از مقاله مروری نظام‌مند حذف شدند. در این شرایط شما باید بتوانید قضاوت خود را که معمولاً مبتنی بر ارزیابی شما از کیفیت مطالعات است توجیه کنید.

استانداردسازی ارزیابی نقادانه شواهد

ارزیابی دقیق، مشخص و استاندارد شده کیفیت مطالعات از دو جنبه اهمیت دارد:

اول اینکه در یک مقاله مروری نظام‌مند باید سعی شود نتایج مبتنی بر باکیفیت‌ترین شواهد در دسترس باشد. برای نیل به چنین هدفی نیازمند فرایندهای روا و استاندارد شده هستیم، تا صرفاً مطالعات مربوط و با کیفیت قابل قبول وارد مرور نظام‌مند شوند.

دوم اینکه منتقل کردن میزان کیفیت مطالعات ورودی به خواننده اهمیت دارد و بیانگر قدرت شواهد ارائه شده برای هر توصیه درمانی است.

آیا ساختارمند بودن ارزیابی نقادانه اهمیت دارد؟

اگر ارزیابی نقادانه بدون ساختار درستی انجام شود، ممکن است ارزیاب‌ها تمایل داشته باشند مطالعاتی را که نتیجه نامطلوب دارند با سخت‌گیری بیشتری ارزیابی کنند. برای مثال از ۲۸ نفر مرور کننده خواسته شد تا یک مطالعه ساختگی را ارزیابی کنند با این توضیح که این ۲۸ نفر به طور تصادفی نسخه منفی یا مثبت مطالعه را دریافت کردند[1] . قسمت روشها که در هر دو نسخه کاملاً مشابه بود، در نسخه منفی مطالعه نسبت به نسخه مثبت آن نمره بسیار کمتری را بدست آورد. بنابراین اینکه تمام مقالات را به یک میزان و با یک درجه از سخت‌گیری ارزیابی کنیم، اهمیت زیادی در انجام یک مرور نظام‌مند دارد.

1- Mahoney 1977

چه تعداد از افراد باید در ارزیابی نقادانه شواهد شرکت کنند؟

بخاطر اهمیت انتخاب صحیح مطالعات در یک مقاله مروری نظام‌مند حداقل باید از دو نفر برای ارزیابی مطالعات استفاده شود. هر کدام از این مرور کنندگان به طور جداگانه و مستقل از هم مقالات را مطالعه کرده و رتبه‌بندی کنند. شایان ذکر است که این افراد باید قبلاً مهارتها و تجارب کافی را درباره انواع مطالعات و ارزیابی آنها کسب کرده باشند. استفاده از تعداد مرورکنندگان بیشتر در انجام مرور نظام‌مند دو جنبه دارد.

از یک جهت با افزایش تعداد مرورکنندگان احتمال خطاها و تورش‌ها کاهش می‌یابد و پایایی[1] یافته‌ها بیشتر می‌شود از طرف دیگر استفاده از بیش از یک نفر مرور کننده بالقوه می‌تواند باعث بروز عدم توافق شود.

آیا صرف آشنایی با روش‌شناسی پژوهش و ارزیابی نقادانه برای مرورکنندگان مطالعات کافی است یا داشتن تخصص در زمینه محتوای مطالعه نیز اهمیت دارد؟

با وجود اینکه متخصصانی که در زمینه محتوای مقاله صاحب نظر هستند ممکن است نظریات از قبل شکل گرفته‌ای داشته باشند که می‌تواند باعث ایجاد تورش شود[2] ولی در عوض آنها می‌توانند ارزیابی پایاتری نسبت به کیفیت مطالعات داشته باشند.[3] بنابراین مطلوب‌ترین حالت این است که هم از افرادیکه در زمینه محتوای مطالعه تخصص دارند و هم افرادیکه در این زمینه تخصص ندارند استفاده شود ولی هر دو گروه باید به اندازه کافی از مسائل روش شناختی مربوط آگاه باشند.

آیا کور بودن ارزیابی نقادانه نسبت به شرایط مطالعه مهم است؟

این کور بودن می‌تواند نسبت به اسم نویسندگان، سازمان، نشریه و نتایج مطالعه باشد.

شواهد تجربی نشان می‌دهد که ارزیابی کور مطالعات نسبت به ارزیابی با آگاهی و باز، نمره‌های کمتر ولی دقیق‌تر و پایاتری بدست

1- Reliability
2- Oxman 1993
3- Jadad 1996

می‌دهد[1] سایر شواهد تجربی اثرات مفید کمتری برای کور بودن ارزیابی نشان داده‌اند[2]. با این حال ارزیابی کور بسیار زمانبر است و مرورکنندگان باید بر اساس مزایای بالقوه و هزینه‌های کور بودن ارزیابی در مورد انجام آن تصمیم‌گیری کنند.

بنابراین ارزیابی نقادانهٔ شواهد به منظور تدوین یک مقالهٔ مروری نظام‌مند مستلزم رعایت یک سری نکات خاص است. به علاوه در چنین شرایطی ارزیابی نقادانهٔ شواهد با محدودیت‌هایی مانند تعداد زیاد مطالعات، محدودیت زمان و محدودیت تعداد مرورکنندگان روبرو می‌شود، در حالیکه این محدودیت‌ها در حالت عادی (طبابت روزمره) وجود ندارد.

چه رویکردهایی برای انجام ارزیابی نقادانه شواهد مطرح است؟

ارزیابی نقادانه شواهد بر اساس دو رویکرد وابسته به هم امکانپذیر است.

رویکرد اول ارزیابی شواهد براساس نوع طرح پژوهش[3] است. در این رویکرد از ابزارهای ارزیابی شواهد که برای هر طرح پژوهشی خاص تهیه شده‌اند استفاده می‌شود و ارزیابی در مورد هر طرح پژوهش مستقل از نوع سوال بالینی صورت می‌گیرد.

رویکرد دوم ارزیابی شواهد براساس نوع سوال بالینی است در این رویکرد ابتدا طرحهای پژوهش که در مورد هر کدام از سوالات بالینی قابل قبول هستند انتخاب شده و سپس این مطالعات مورد ارزیابی قرار می‌گیرند. رویکردی که در اکثر متون طبابت مبتنی بر شواهد و از جمله این کتاب انتخاب شده است، ارزیابی بر مبنای نوع سوال بالینی است.

1- Jadad 1996
2- Berlin 1997
3- Study design

چه ابزارهایی در ارزیابی نقادانه شواهد استفاده می‌شود؟

ابزارهایی که برای ارزیابی نقادانه شواهد استفاده می‌شوند معمولاً به دو دسته مقیاس‌ها[1] و چک لیست‌ها[2] تقسیم می‌شوند. در هر چک لیست چند معیارکه مربوط به جنبه‌های مختلف کیفیت شواهد است وجود دارد و خوانندگان‌می‌توانند پس از خواندن مقاله رعایت شدن یا نشدن هر کدام از این معیارها را تعیین کنند و سپس با توجه به تعداد و اهمیت موارد رعایت شده و رعایت نشده راجع به کیفیت مطالعه قضاوت نمایند.

چک لیست‌ها کاربرد عمومی‌تری نسبت به مقیاس‌ها دارند و می‌توانند در بسیاری از موقعیتها مثل طبابت روزمره مورد استفاده قرار گیرند. مقیاس‌ها نیز مشابه چک لیست‌ها از چند معیار تشکیل شده‌اند. با این تفاوت که در مقیاس هر معیار دارای یک نمره خاص است و مطالعه‌ای که این معیار را رعایت کرده باشد نمرهٔ آن را اخذ می‌کند. قضاوت نهایی خواننده مقاله بر اساس این نمره‌ها انجام می‌شود. استفاده از مقیاس‌ها در ارزیابی مطالعات برای تهیه مقالات مروری نظام‌مند شایعتر است. مطالعات زیادی برای تعیین میزان روایی و پایایی این ابزارها انجام شده است. در یکی از این مطالعات، دیوید موهر[3] و همکارانش ۲۵ مقیاس و ۹ چک لیست را که برای ارزیابی کیفیت و روایی کارآزمایی‌های تصادفی شده بالینی استفاده می‌شدند، مورد بررسی قرار دادند.[4]

این مقیاس‌ها و چک لیست‌ها بین ۳ تا ۵۷ معیار داشتند و پر کردن آنها بین ۱۰ تا ۴۵ دقیقه زمان می‌برد. تقریباً تمام معیارهای ذکر شده در این ابزارها بر اساس معیارهای مورد پذیرش که در کتابهای مربوط ذکر شده‌اند، انتخاب شده بود.

بسیاری از این ابزارها نتوانسته بودند کیفیت گزارش مقاله را از روایی طرح و اجرای مطالعه تمیز بدهند بعلاوه در بسیاری از این ابزارها شرط رعایت یک معیار گزارش شدن آن در مطالعه (مثلاً بیان نحوه تخصیص افراد به گروههای درمان و کنترل) بود، نه اینکه آیا این معیار به طور مناسب در مطالعه لحاظ شده است یا خیر.

1- Scales
2- Check lists
3- David Moher
4- Moher 1995, Moher 1996

بسیاری دیگر از این ابزارها معیارهایی را شامل می‌شوند که مستقیماً به روایی مربوط نمی‌شد. مثلاً اینکه آیا محاسبه توان مطالعه انجام شده است یا خیر (این معیار بیشتر مربوط به دقت آماری مطالعه است) یا اینکه آیا معیارهای ورود و خروج[1] به خوبی توصیف شده است (این معیار بیشتر به تعمیم‌پذیری و کاربرد مطالعه مربوط می‌شود تا روایی آن). علاوه بر مشکلات مذکور مقیاس‌ها مشکلات دیگری نیز دارند. از آنجاییکه هیچ استاندارد طلایی[2] برای سنجش روایی واقعی مطالعات وجود ندارد، امکان اعتبار سنجی هر سیستم نمره دهی پیشنهاد شده‌ای نیز محدود است. با وجود اینکه می‌توان مقیاسها را بر اساس اصول علمی سنجش تهیه کرد ولی باز هم ارتباط بین نمره‌ای که از این مقیاس‌ها بدست می‌آید، با درجه‌ای که این مطالعات، تورش‌ها را حذف کرده‌اند چندان واضح نیست.

بسیاری از مقیاسهایی که برای ارزیابی روایی کارآزمایی‌های تصادفی شده بالینی استفاده می‌شوند نمره نهایی را از طریق جمع کردن نمرات مربوط به هر معیار (با یا بدون وزن دهی به هر کدام از آنها) تعیین می‌کنند. با اینکه این رویکرد به سادگی قابل انجام است ولی شواهد تجربی در جهت حمایت از آن وجود ندارد[3]. مقیاس‌هایی که تعداد معیارهای بیشتری دارند و از سیستم امتیازدهی پیچیده‌تری برخوردار هستند نسبت به رویکردهای دیگر وقت بیشتری از خوانندگان می‌گیرند. در حالیکه پایایی آنها در ارزیابی روایی مطالعات با شواهد تجربی تأیید نشده است[4].

بنابراین به طور کلی ابزارهایی که از رویکردهای ساده‌تر برای ارزیابی روایی استفاده می‌کنند و میزان رعایت هر کدام از معیارها و اهمیت آنها را جداگانه بیان کنند (نه به صورت یک نمره کلی) توصیه می‌شوند.

در ادامهٔ کتاب در هر فصل معیارهای ارزیابی و بررسی روایی شواهد مربوط به سوٌالات بالینی اصلی و شایع شامل شواهد مربوط به تشخیص، غربالگری، درمان، اتیولوژی و آسیب[5]، پیش‌آگهی[6] و شیوع و بروز[7] بحث شده است.

هر فصل با یک سناریوی بالینی استاندارد آغاز شده و در ادامه به ارائهٔ چک لیست ارزیابی همراه با توضیح معیارهای چک لیست و ارزیابی مقالهٔ نمونه پرداخته شده است.

1- Inclusion & Exclusion criteria
2- Gold standard
3- Emerson 1990, Schulz 1995
4- Juni 1999
5- Etioligy and harm
6- Prognasis
7- Prevalence and Incidence

References

1- How to Use the Evidence: Assesssment of scientific evidence, National Health and Medical Research Council; 2000.

2- How to Review the Evidence: Systematic identification and review of the scientific literature. NHMRC; 1999.

3- Cochrane Reviewer's Handbook; The Cochrane Collaboration; 2004.

4- Cochrane Reviewers' Handbook Glossary. The Cochrane Collaboration; 2001.

5- Cochrane Neuromascular Disease Group. Writing a Cochrane Systematic Review. The Cochrane Collaboration.

6- Riegelman RK. Studying a Study and Testing a Test. 4th ed. Lippincott Williams and Wilkins; 2000.

7- Greenhalgh T. How to Read a Paper. 2nd ed. BMJ publishing Group 2001.

8- Egger M, Smith GD, Altman DG, Systematic Reviews in Health care: Meta-analysis in context. 2nd ed. BMJ Publishing Group 2001.

9- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine, How to Practice and Teach EBM, 2nd ed. Churchill Livingstone 2000.

10- Guyatt G, Rennie D, Hayward R. User's Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice.American Medical Association; 2002.

11- A Guide to the Developement, Implemention and Evaluation of Clinical Practice Guideline, NHMRC; 1999.

فصل سوم

شواهد مربوط به تشخیص

فصل سوم
شواهد مربوط به تشخیص

خانمی ۴۵ ساله به مطب شما مراجعه می‌کند. او از سردردی که به تازگی برایش پیش آمده شکایت دارد و ذکر می‌کند که سر دردش در نواحی گیجگاهی شدیدتر است ولی تاکنون اختلال بینایی نداشته است. در معاینه فیزیکی متوجه تندرنس شریان تمپورال می‌شوید که در سمت راست تا حدودی بیشتر است. بیمار اظهار می‌کند که پیش از این به دو مختصص دیگر نیز مراجعه کرده و هر دوی آنها پیشنهاد بیوپسی از شریان تمپورال را برای وی مطرح کرده‌اند. بیمارکه در یکی از کتابخانه‌های محلی مشغول به کار است، درباره بیوپسی از شریان تمپورال در اینترنت جستجو کرده است و حاضر به پذیرش عوارض احتمالی آن از جمله نکروز پوست، ریزش ابرو و فلج قسمتی از صورت بخاطر صدمه به عصب فاشیال نیست.

بخاطر می‌آورید که ماه قبل در یکی از کنفرانس‌ها مطلبی راجع به تشخیص آرتریت تمپورال[1] بوسیله اولتراسونوگرافی داپلکس[2] شنیده‌اید ولی از کافی بودن دقت تشخیصی آن برای آرتریت تمپورال مطمئن نیستید. شروع به جستجو می‌کنید:

در یک خانم ۴۵ ساله که با سردرد و تندرنس شریان تمپورال مراجعه کرده است دقت تشخیصی اولتراسونوگرافی داپلکس برای تشخیص آرتریت تمپورال در مقایسه با نمونه‌برداری از شریان چه میزان است؟

پس از جستجو در پایگاه داده‌ای "پاب مد"[3] به مقاله‌ای از نشریهٔ پزشکی نیوانگلند[4] (NEJM) بر می‌خورید که به نظر کاملاً به این سؤال بالینی مرتبط است:

1- Temporal Arteritis
2- Duplex Ultrasonograghy
3- PubMed Database
4- New England Journal of Medicine

عنوان مطالعه "اولتراسونوگرافی داپلکس رنگی در تشخیص آرتریت تمپورال"[1] است (ضمیمه ۱). اکنون نوبت ارزیابی نقادانه مقاله با توجه به چک لیست زیر می‌باشد.

آیا مطالعه انجام شده برای تعیین دقت تست تشخیصی از کیفیت مناسب برخوردار است؟

- آیا تست تشخیصی در طیف مناسبی از بیماران مانند بیمارانی که در عمل به آنها برخورد می‌کنیم مورد ارزیابی قرار گرفته است؟
- آیا تست مرجع[2] به کار رفته، از لحاظ بالینی مناسب است؟
- آیا مقایسه نتایج تست آزمایشی[3] و تست مرجع به طور کامل و مستقل انجام شده است؟
- آیا مقایسه و تفسیر نتایج تست آزمایشی و تست مرجع بدون آگاهی از نتایج هر کدام بوده است؟
- آیا انجام هر دو تست پیش از شروع هر درمانی برای بیماری مورد نظر بوده است؟
- آیا نتایج تستهای همه شرکت‌کنندگان در تحلیل مطالعه در نظر گرفته شده است؟

آیا این مطالعه از نظر بالینی اهمیت دارد؟

- آیا مطالعه نشان دهنده توانایی کافی تست تشخیصی برای تعیین دقیق افراد مبتلا و غیرمبتلا به بیمای مورد نظر است؟

ارزشیابی کامل عملکرد تستهای تشخیصی[4] مستلزم بررسی پایایی[5]، دقت تشخیصی[6]، اثرات تشخیصی و درمانی و تأثیر خالص تست بر پیامدهای بیمار است. در مورد هر یک از جنبه‌های ارزیابی تست‌های تشخیصی بالقوه امکان دستیابی به مطالعات متنوعی وجود دارد.

1- Color Duplex Ultrasonography in the Diagnosis of Temporal Arteritis N Engl J Med 1997, 337: 1336-42
2- Reference test
3- Experimental test
4- Diagnostic test performance
5- Reliability
6- Diagnostic accuracy

تستها در پزشکی بطور شایع برای انجام غربالگری، تشخیص، تعیین درجه و مرحله و پایش پیشرفت بیماریها استفاده می‌شوند. اطلاعات تشخیصی از منابع متعددی بدست می‌آیند. بعضی از این منابع شامل مطالعات تصویربرداری، آزمایشات بیوشیمیایی، بررسی‌های روانپزشکی، مطالعات آسیب شناختی و علائم و نشانه‌هایی[1] است که در طی شرح حال گیری و معاینات بدست می‌آید.

هر کدام از اطلاعات بدست آمده از این منابع می‌تواند واجد ارزش تشخیصی باشد و به عنوان نتیجه یک تست تشخیصی یا غربالگری در نظر گرفته شود. این اطلاعات می‌تواند به منظور یافتن نوع بیماری در افراد بیمار (تشخیص) و یا پیدا کردن زودرس بیماری در افراد بدون علامت (غربالگری) جمع‌آوری شود.

استفاده از مطالعاتی که به منظور ارزیابی پایایی، دقت و میزان اثر این گونه تست‌ها انجام شده‌اند برای انتخاب بهترین تست و تفسیر صحیح نتایج آن ضروری است.

برای داشتن درک صحیحی از بررسی‌های تشخیصی، پزشک باید قادر باشد با توجه به نتایج بدست آمده از تست به استنباط درستی از احتمال وجود بیماری مورد نظر در بیمار خود برسد.

تستها به ندرت یک تشخیص را صددرصد قطعی می‌کنند ولی در عمل معمولاً می‌توانند اطلاعات کافی را برای پذیرفتن[2] و یا رد کردن[3] یک تشخیص در اختیار ما بگذارند. به عبارت دیگر تستها می‌توانند تشخیص را آنقدر قطعی کنند که بر پایه آن بتوان بین مزایای درمان و عواقب مورد انتظار در صورت عدم درمان به قضاوت صحیحی رسید. این فصل بر ارزیابی نقادانه مطالعات مربوط به "دقت تستهای تشخیصی" تمرکز دارد که بیانگر ارتباط احتمالی بین نتایج مثبت و منفی تستها و وجود یا عدم وجود بیماری است. دقت تست تشخیصی نشان می‌دهد تست تا چه اندازه در تمایز بین افراد مبتلا به یک بیماری خاص و افراد غیرمبتلا موفق است. مطالعات مربوط به دقت تستهای

1- Signs and Symptoms
2- Rule in
3- Rule out

تشخیصی از جهات مختلف با سایر مطالعات بالینی متفاوت هستند. در مطالعاتی که برای بررسی دقت یا عملکرد تستهای تشخیصی انجام می‌شود، نتایج تست بین گروههای مختلف مبتلا، یا غیر مبتلا به بیماری مورد نظر مقایسه می‌شود. هر یک از افراد شرکت کننده در مطالعه علاوه بر تست آزمایشی مورد نظر با تست مرجع به عنوان یک استاندارد طلایی[1] نیز مورد آزمایش قرار می‌گیرد. در این گونه مطالعات رابطه‌ای که بین نتایج تست و وضعیت بیماری وجود دارد، با استفاده از سنجه‌های احتمال[2] مانند حساسیت[3]، ویژگی[4]، نسبت درست نمایی[5] و ... توصیف می‌شود. نزدیک بودن نتایج تست مرجع به واقعیت اهمیت بسیار زیادی دارد. چرا که در غیر اینصورت رسیدن به تخمین صحیحی از عملکرد تست آزمایشی غیرممکن خواهد بود. رسیدن به این مهم گاهی اوقات مستلزم ترکیب کردن اطلاعات مختلف، انجام آزمایشات تهاجمی و یا پیگیری طولانی مدت بیمار به عنوان تست مرجع است.

سطح مطالعه

اولین قدم در ارزیابی نقادانه شواهد مربوط به دقت تستهای تشخیصی بررسی طرح مطالعه است. طرحهای پژوهشی که برای پاسخ به سوالات بالینی مربوط به دقت تستهای تشخیصی مناسب می‌باشند، علاوه بر مقالات مروری نظام‌مند[7]، عمدتاً شامل طرحهای همگروهی و مورد ـ شاهدی هستند (همانطور که پیش از این گفته شد مقالات مروری نظام‌مند بالاترین سطح شواهد را تشکیل می‌دهند). بنا بر شواهد موجود استفاده از طرح مورد ـ شاهدی علیرغم فراوانی کمتر نسبت به طرح همگروهی به عنوان بزرگترین منبع تورش[8] در مطالعات مربوط به دقت تست‌های تشخیصی مطرح است. مطالعاتی که از این طرح استفاده می‌کنند معمولاً تمایل دارند دقت تستهای تشخیصی را بیش از آنچه هست تخمین بزنند (تقریباً سه برابر بیش از مطالعات همگروهی). بنابراین مطالعاتی که از طرح همگروهی بهره می‌برند نسبت به سایر طرحها واجد سطح بالاتری از شواهد هستند.

1- Gold Standard
2- Probabilistic measures
3- Sensitivity
4- Specificity
5- Likelihood ratio
6- Invasive procedures
7- Systematic Reviews
8- Bias

کیفیت مطالعه

- آیا تست تشخیصی در طیف مناسبی از بیماران (مانند بیمارانی که در عمل با آنها برخورد می‌کنیم) مورد ارزیابی قرار گرفته است؟

در اینجا نحوه انتخاب نمونه مطالعه اهمیت پیدا می‌کند. همانطور که قبلاً هم گفته شد نحوه نمونه‌گیری عمدتاً بر تعمیم‌پذیری و در نتیجه روایی خارجی مطالعه تأثیر می‌گذارد ولی بویژه در مطالعات مربوط به تعیین دقت تستهای تشخیصی، نمونه‌گیری در روایی و کیفیت مطالعه نیز تأثیر بسزایی دارد. نمونه‌گیری مطلوب در این نوع مطالعات از طریق فراخوانی افراد به صورت تصادفی و یا به صورت پی در پی[1] است (در نمونه‌گیری پی در پی تمام بیمارانی که در یک بازه زمانی مراجعه می‌کنند و از نظر بالینی متناسب با سوال پژوهشی ما هستند برای مطالعه انتخاب می‌شوند. هرگاه جمعیت مطالعه نماینده[2] طیف بیماریهایی که تست مورد نظر در طبابت باید بین آنها تمایز دهد، نباشد تورشی تحت عنوان تورش طیف[3] پدید می‌آید. توجه داشته باشید این امر که نمونه گرفته شده باید نماینده مناسبی از جمعیت مطالعه باشد، مسأله‌ای جداگانه و مربوط به روایی خارجی مطالعه است! (اگر تفاوت این دو موضوع هنوز برایتان کاملاً مشخص نشده است لطفاً به فصل اول کتاب، مبحث نمونه‌گیری مراجعه کنید).

در عمل فراخوانی افراد فاقد و واجد بیماری مورد نظر در گروههای جداگانه (مانند آنچه در مطالعات مورد ـ شاهدی صورت می‌گیرد) برای مطالعه آسانتر است، ولی از آنجاییکه نسبت بیماریابی[4] بسته به شدت بیماری و وجود بیماریهای همراه متفاوت است، این نوع نمونه‌گیری در طرح مورد ـ شاهدی می‌تواند باعث ایجاد تورش در مطالعه شود. به بیان دیگر انتخاب موردهایی که از قبل به عنوان مبتلا به بیماری، شناخته شده‌اند همچنین انتخاب شاهدهایی که از هر جهت سالم هستند، نیز تخمین ویژگی تست را دچار تورش خواهد کرد. آنچه معمولاً در مطالعات مورد ـ شاهدی اتفاق می‌افتد انتخاب موردها از بین افراد مبتلا به بیماری متوسط تا شدید و استفاده از گروههای شاهد سالم است. این کار تخمین بسیار بالایی از حساسیت و ویژگی بدست می‌دهد که

1- Consecutive
2- Representative
3- Spectrum bias
4- Detection rate

نسبت به آنچه در طبابت روزمره و نمونه‌های بیمارستانی دیده می‌شود بسیار بیشتر است.

علاوه بر نحوه انتخاب نمونه اینکه نمونه از چه سطحی از مراقبت‌های سلامت انتخاب می‌شود نیز اهمیت خاصی دارد. چرا که طیف بیماری و تشخیصهای افتراقی جایگزین در سطوح مختلف از مراقبت سلامت (مانند خدمات اولیه و ثانویه) متفاوت است. در نمونه‌هایی که از طبابت عمومی به عنوان سطح اول مراقبتهای سلامت گرفته می‌شود، اکثریت افراد مبتلا در مراحل اولیه بیماری هستند. در حالیکه در نمونه‌های بیمارستانی که نماینده سطح دوم مراقبتهای سلامت هستند، تعداد بیشتری از بیماران به نوع متوسط تا شدید بیماری مبتلا می‌باشند و بیماریهای جایگزین متفاوتی نیز دارند، زیرا بعضی از این بیماریها در سطوح قبلی درمان شده و یا به مراکز دیگر ارجاع شده‌اند.

اثر تورش طیف بر حساسیت تست با دقت در جدول ۱ مشخص می‌شود:

جدول ۱

طیف بیماری				
نمونه بیمارستانی (n=100)	نمونه طبابت عمومی (n=100)	موردها در مطالعه مورد ـ شاهدی (n=100)	حساسیت	درجه بیماری
۲۰	۸۰	۰	۰/۵۰	ابتدایی
۳۰	۱۵	۲۰	۰/۷۵	متوسط
۵۰	۵	۸۰	۱/۰۰	پیشرفته
۰/۸۳	۰/۵۶	۰/۹۵	—	حساسیت اندازه‌گیری شده

همانطور که مشاهده می‌کنید فراوانی نوع پیشرفته بیماری در مورد مطالعه مورد ـ شاهدی بیشتر از نمونه بیمارستان و در نمونه بیمارستانی بیشتر از نمونه طبابت عمومی می‌باشد و حساسیت‌های گزارش شده در هر نمونه نیز به ترتیب برابر ۰/۹۵، ۰/۸۳ و ۰/۵۶ می‌باشد. این تفاوت بیانگر اثر تورش طیف بر حساسیت تست است.

اثر تورش طیف بر ویژگی تست در جدول ۲ نشان داده شده است.

		طیف تشخیص‌های جایگزین			
بیماری	ویژگی	شاهدها در مطالعه مورد ـ شاهدی (n=100)	نمونه طبابت عمومی (n=100)	نمونه بیمارستانی (n=100)	
تشخیص جایگزین "X"	۰/۳۰	۰	۳۰	۷۵	
تشخیص جایگزین "Y"	۰/۹۵	۰	۶۵	۲۵	
بالغین سالم	۰/۹۹	۱۰۰	۵	۰	
ویژگی اندازه‌گیری شده	—	۰/۹۹	۰/۷۶	۰/۴۶	

جدول ۲

در گروه شاهدها (در مطالعه مورد ـ شاهدی) از آنجاییکه تمام گروه از بالغین سالم تشکیل شده‌است و این افراد فاقد بیماریهای

جایگزین "X" و "Y" هستند، ویژگی گزارش شده برای تست بیشینه مقدار یعنی ۰/۹۹ است ولی در نمونه طبابت عمومی و بیمارستانی که افراد به طیف وسیعتری از بیماریهای جایگزین مبتلا هستند، ویژگی تست پایین‌تر گزارش شده است (۰/۷۶ و ۰/۴۶). این تفاوت نمایانگر اثر تورش طیف بر ویژگی تست است.

- آیا تست مرجع استفاده شده از لحاظ بالینی مناسب است؟

انتخاب استاندارد مرجع مناسب در مطالعه دقت تستهای تشخیصی اهمیت حیاتی دارد. استاندارد مرجع در حقیقت به منزله استاندارد طلایی است و مقایسه آن با تست مورد آزمایش همیشه یک طرفه می‌باشد، یعنی اگر بین نتایج حاصل از تست مرجع و تست آزمایش منافاتی وجود داشت، بنا به فرض همیشه نتیجه تست آزمایشی نادرست در نظر گرفته می‌شود.

- آیا مقایسه نتایج تست آزمایشی و تست مرجع به طور کامل و مستقل انجام شده است؟

مستقل بودن انجام دو تست مرجع و آزمایشی نیز به نوبه خود حائز اهمیت است. در بعضی شرایط تشخیص مرجع بر اساس نتایج طیفی از تستها و سایر شواهد بالینی در دسترس داده می‌شود، در چنین شرایطی نتیجه تست آزمایشی نباید مشمول این طیف از نتایج قرار بگیرد، در غیر اینصورت دقت تشخیصی به احتمال زیاد بیش از حد واقعی تخمین زده خواهد شد. این اثر به تورش مشارکت[1] معروف است.

تورش تأیید[2]: این تورش زمانی اتفاق می‌افتد که تصمیم‌گیری در مورد انجام تست مرجع تحت تأثیر نتیجه تست آزمایشی و یا سایر عواملی که وجود بیماری را نامحتمل جلوه می‌دهد قرار بگیرد، چرا که در چنین شرایطی اکثر پژوهشگران در مورد انجام تستهای تهاجمی دچار تردید می‌شوند.

1- Incorporation bias
2- Verification bias

عدم تأیید کامل تست آزمایشی در دو سطح امکان‌پذیر است:

۱ تأیید ناکامل[1] که در آن برای تعدادی از شرکت کنندگان که غالباً نتیجه تست آزمایشی در آنها منفی شده است، تست مرجع انجام نمی‌شود.

۲ تأیید افتراقی[2] که در آن تستهای مرجع متفاوتی بر اساس نتایج تست آزمایشی انجام می‌شود.

تورش تأیید ناکامل معمولاً منجر به کاهش تعداد منفی کاذب و منفی واقعی می‌شود و بدین ترتیب تمایل دارد حساسیت را بیشتر و ویژگی را کمتر از مقدار واقعی گزارش کند. برخلاف آن تورش تأیید افتراقی هر دو تخمین را به سمت بالا دچار تورش خواهد کرد.

شواهد تجربی نشان می‌دهد استفاده از تستهای متفاوت در مطالعه (تورش تأیید افتراقی) در مجموع عملکرد تست تشخیصی را بیش از حد واقعی تخمین می‌زند ولی تورش تأیید ناکامل اثر ثابت شده و یکنواختی بر تخمین میزان دقت تشخیصی تستها نداشته است.

- آیا مقایسه و تفسیر تست آزمایشی و تست مرجع بدون آگاهی از نتایج هر کدام بوده است؟

کور بودن[3] در مطالعات مربوط به دقت تستهای تشخیصی بدین معناست که هر یک از تستها بدون اطلاع از نتایج تست دیگر انجام شده و تفسیر گردد. این موضوع در تستهایی که مستلزم قضاوت ذهنی پژوهشگر است، مانند تستهایی که مبتنی بر ادراک پژوهشگر از تفسیر تصاویر و صداهای تشخیصی هستند، اهمیت ویژه‌ای می‌یابد. براساس شواهد مطالعات غیرکور به طور متوسط دقت تشخیصی را بیش از آنچه هست تخمین می‌زنند.

- آیا انجام هر دو تست پیش از شروع هر درمانی برای بیماری مورد نظر بوده است؟

1- Partial verification
2- Differential verification
3- Blinding

اگر پیش از انجام هر کدام از تست‌ها درمان بیماری شروع شده باشد ممکن است تورشی به نام تناقض درمان[1] ایجاد شود؛ بیمارانی که تشخیص بیماری مورد نظر با تست اول برای آنها گذاشته شده است ممکن است تا پیش از انجام تست دوم درمان شوند و بدین ترتیب بسته به اینکه کدام تست زودتر انجام شده به عنوان منفی کاذب یا مثبت کاذب طبقه‌بندی گردند.

- آیا نتایج تستهای همه شرکت کنندگان در تحلیل مطالعه در نظر گرفته شده است؟

بسیاری از تست‌ها نتایج بعضی افراد را به عنوان غیرقابل تشخیص و یا نارسایی تست[2] گزارش می‌کنند. اگر چه لحاظ کردن این پیامدها در تحلیل مطالعه همیشه کار ساده‌ای نیست ولی بی‌توجهی به آنها معمولاً دقت تست را از آنچه هست مطلوبتر نشان می‌دهد.

دقت آماری

دقت آماری[3] بر خلاف مطالعات مربوط به اثربخشی مداخلات درمانی مانند کارآزماییهای تصادفی شدهٔ شاهددار، در مطالعات مربوط به دقت تست‌های تشخیصی کمتر مورد بحث قرار می‌گیرد، چرا که معمولاً در اینگونه مطالعات فرضیه‌ای به طور رسمی تست نمی‌شود. از طرف دیگر در منابع طبابت مبتنی بر شواهد دقت آماری در مطالعات تستهای تشخیصی کمتر مورد تأکید قرار گرفته است. بنابراین در چک لیست ارزیابی نقادانه شواهد مربوط به تستهای تشخیصی، موردی تحت عنوان دقت آماری آورده نشده است.

ارزیابی نقادانه کیفیت گزارش مقالات مربوط به دقت تستهای تشخیصی

ارزیابی کیفیت گزارش در این مقالات از جایگاه خاصی برخوردار است ولی به لحاظ تفکیک موضوعی در چک لیست مربوط به ارزیابی مطالعه ذکر نمی‌شود.

1- Treatment paradox
2- Test failure
3- Statistical precision

گزارش مطالعه در صورتی مطلوب خواهد بود که توصیف شفافی از تست مرجع و تست آزمایشی ارائه دهد. بیان تعریف پیامدهای مثبت و منفی برای هر دو تست و توصیف ویژگیهای جمعیت شناختی[1]، بیماریهای همراه و سابقه ارجاع بیماران نیز در گزارش چنین مقالاتی لازم است. طبق شواهد تجربی در مطالعاتی که گزارش آنها فاقد توصیف دقیقی از تست آزمایشی و ویژگیهای جمعیت بوده است، دقت تشخیصی در حدود ۱/۵ برابر بیش از مقدار واقعی تخمین زده شده است و در مطالعاتی که هیچ توصیفی از تست مرجع در گزارش آنها وجود ندارد، دقت تشخیصی بطور متوسط ۰/۷ مقدار واقعی گزارش شده است.[2]

ارزیابی اهمیت بالینی مطالعه

- آیا مطالعه نشان‌دهنده توانایی کافی تست تشخیصی برای تمایز بین افراد مبتلا و غیرمبتلا به بیماری مورد نظر است؟

برای پاسخ به این سوال ابتدا باید با سنجه‌های دقت تشخیصی آشنا شوید. سنجه‌هایی که به طور شایع و سنتی برای بیان دقت تستهای تشخیصی به کار می‌روند شامل حساسیت و ویژگی هستند. علاوه بر آنها سنجه‌هایی چون نسبت درست نمایی و ارزش اخباری[3] نیز وجود دارد که در طبابت مبتنی بر شواهد دارای کاربرد بیشتری هستند.

حساسیت و ویژگی

تستهای مرجع و تست‌های آزمایشی به طور طبیعی نتایج خود را به صورت یک تقسیم‌بندی دوتایی (وجود بیماری، عدم وجود بیماری) ارائه می‌دهند. بنابراین به طور کلی تست‌ها نتایج بعضی از افراد را نرمال و بعضی را غیرنرمال گزارش می‌کنند. تعریف نرمال بودن در طبابت از زوایای متعددی امکانپذیر است ولی تعریفی که در تعیین دقت تست‌های تشخیصی کاربرد بیشتری دارد،

1- Demographic characteristics
2- Lijmer et al
3- Predictive value

عبارتست از: دامنه‌ای از نتایج که خارج از این دامنه، بیماری مورد نظر بسیار محتمل خواهد بود. تقسیم‌بندی نتایج تست براساس وجود یا عدم وجود بیماری به ما امکان می‌دهد نتایج هر تست را در قالب یک جدول ۲×۲ نشان دهیم، که در آن افراد به گروههای مثبت حقیقی[1] (TP) منفی حقیقی[2] (TN) یعنی نتایج صحیح تست و مثبت کاذب[3] (FP) و منفی کاذب[4] (FN) یعنی نتایج ناصحیح تست تقسیم می‌شوند. محاسبه حساسیت، ویژگی و نسبت درست نمایی از روی تعداد افرادی که در هر کدام از گروههای فوق طبقه‌بندی شده‌اند امکانپذیر است. (جدول ۳)

نتایج تست	مبتلا به بیماری	غیر مبتلا به بیماری	
تست مثبت	مثبت حقیقی (TP)	مثبت کاذب (FP)	مجموع مثبت‌ها
تست منفی	منفی کاذب (FN)	منفی حقیقی (TN)	مجموع منفی‌ها
	مجموع مبتلایان	مجموع غیرمبتلایان	

جدول ۳

حساسیت: نسبت افراد مبتلا به بیماری که تست آنها مثبت شده است.

$$\text{حساسیت} = \frac{\text{تعداد مثبت حقیقی (TP)}}{\text{مجموع مبتلایان}}$$

1- True positive
2- True negative
3- False positive
4- False negative

ویژگی: نسبت افراد غیرمبتلا به بیماری که تست آنها منفی شده است.

$$\text{ویژگی} = \frac{\text{تعداد منفی حقیقی (TN)}}{\text{مجموع غیرمبتلایان}}$$

سنجه دیگری که امروزه به طور شایع در گزارش مطالعات دقت تستهای تشخیصی به کار می‌رود نسبت مثبت کاذب[1] است که به صورت زیر محاسبه می‌شود.

$$\text{ویژگی} - ۱ = \frac{\text{تعداد مثبت کاذب (FP)}}{\text{مجموع غیرمبتلایان}}$$

▪ نسبت درست نمایی

محاسبه نسبت درست نمایی اندکی پیچیده‌تر است. از نظر مفهومی نسبت درست نمایی در مطالعه بیان می‌کند، یک فرد مبتلا به بیماری نسبت به یک فرد غیرمبتلا چند برابر احتمال دارد که نتیجه خاصی (معمولاً مثبت یا منفی) از تست ما کسب کند.

در تستهایی که نتیجه آنها به صورت دوگانه[2] گزارش می‌شود، دو نسبت احتمال محاسبه می‌گردد.

نسبت درست نمایی مثبت (LR+ve) که اندازه آن معمولاً بزرگتر از ۱ است با توجه به جدول ۳ اینگونه محاسبه می‌شود:

$$LR + ve = \frac{(\text{تعداد مبتلایان} \div \text{تعداد مثبت حقیقی})}{(\text{تعداد غیرمبتلایان} \div \text{تعداد مثبت حقیقی})}$$

به عبارت دیگر

$$LR + ve = \frac{\text{حساسیت}}{\text{ویژگی} - ۱}$$

1- False Positive rate
2- Binary

نسبت درست نمایی منفی (LR-ve) که معمولاً عددی بین ۰ و ۱ است به شرح زیر محاسبه می‌شود.

$$LR - ve = \frac{(\text{تعداد مبتلایان} \div \text{تعداد منفی کاذب})}{(\text{تعداد غیرمبتلایان} \div \text{تعداد منفی کاذب})}$$

به عبارت دیگر

$$LR - ve = \frac{\text{حساسیت} - ۱}{\text{ویژگی}}$$

نکته قابل توجه در محاسبه سنجه‌های ویژگی، حساسیت و نسبت درست نمایی این است که در تمام این محاسبات از ستونهای عمودی جدول استفاده می‌شود. بنابراین اعداد بدست آمده برای این سنجه‌ها مستقل از فراوانی مبتلایان و غیرمبتلایان به بیماری در نمونه مطالعه است. پس ارزش این سنجه‌ها مستقیماً تحت تأثیر تغییرات شیوع بیماری در نمونه گرفته شده قرار نمی‌گیرد .

در بعضی شرایط تست مورد بررسی نتایجی بدست می‌دهد که به صورت توالی‌های تقسیم شدهٔ مرتبی است که از سنجش‌های پیوسته بدست آمده‌اند. برای مثال ممکن است نتایج یک تست تصویربرداری مثل MRI برای پیدا کردن یک ضایعه آناتومیک به اشکال زیرگزارش شود: قطعاً مثبت، احتمالاً مثبت، نامشخص، احتمالاً منفی و یا قطعاً منفی. حساسیت و ویژگی به طور طبیعی در شرایطی کاربرد دارند که فقط دو دسته از نتایج وجود دارد ولی در این شرایط نیز می‌توان با قرار دادن یک حد آستانه جدید و ترکیب کردن دسته‌های بالا و پایین آستانه، نتایج را به دو دسته پیامدهای مثبت و منفی تقسیم کرد و بدین ترتیب حساسیت و ویژگی را برای تست محاسبه نمود. انجام این کار اگر چه نسبتاً ساده است ولی باعث از دست رفتن حجم

زیادی از اطلاعات مفید می‌شود. نسبت درست نمایی، جایگزین بسیار بهتری در این شرایط است، چرا که برای هر دسته از نتایج تست می‌توان یک نسبت درست نمایی مجزا محاسبه کرد و بدین ترتیب از تمام طیف نتایج و اطلاعات تشخیصی بدست آمده استفاده نمود.

ـ کاربرد نسبت درست نمایی

همانطور که گفته شد نسبت درست نمایی یکی از سنجه‌های بسیار پرکاربرد در بیان دقت تستهای تشخیصی است. نسبت درست نمایی در طبابت بالینی به منظور تخمین احتمال وجود بیماری در یک فرد با توجه به نتیجه تست وی به کار می‌رود. انجام این عملیات بر پایه نظریه بیز[1] شکل گرفته است. در این نظریه بیان می‌شود که شانس مبتلا بودن به بیماری پس از انجام تست[2] در یک فرد برابر است با شانس پیش آزمون[3] ضرب در نسبت درست نمایی. توجه کنید که برای استفاده از این نظریه باید ابتدا احتمال پیش آزمون[4] را به شانس[5] تبدیل کنید و از آن در محاسبه احتمال پس آزمون[6] استفاده نمایید. نحوه محاسبه احتمال پیش آزمون (احتمالی که پزشک برای بیماری پیش از انجام تست در نظر می‌گیرد) بسته به داده‌های مختلف از جمله شیوع بیماری، قضاوت بالینی پزشک و شواهد در دسترس تعیین می‌شود.

محاسبه احتمال پس آزمون (احتمال جدیدی که با توجه به نتیجه تست و احتمال پیش آزمون برای بیمار بودن بدست می‌آید) با استفاده از جدول ۳ و نظریه فوق بدین ترتیب است. (برای آسانی کار احتمال پیش آزمون برابر شیوع بیماری در نظر گرفته شده است)

Pre-test probability = Prevalence = (TP + FN / TP + FN + FP + TN)

Pre-test odds = Prevalence/ (1 - Prevalence)

Post-test odds = Pre- test odds x Likelihood ratio

1- Bay'es theorem
2- Post-test odds
3- Pre-test odds
4- Pre-test probability
5- Odds
6- Post-test probability

(LR+ve در صورتی که نتیجهٔ تست مثبت باشد و LR-ve در صورتیکه نتیجه تست منفی باشد)

Post - test Probability = Post- test odds / (Post- test odds + 1)

توضیح بیشتر راجع به این کاربرد بسیار مهم نسبت درست‌نمایی با توجه به هدف این کتاب لازم به نظر نمی‌رسد در این زمینه توصیه می‌شود به منابع دیگر طبابت مبتنی بر شواهد مراجعه کنید.

▪ ارزش اخباری

دستهٔ دیگر از سنجه‌های دقت تشخیصی، ارزشهای اخباری هستند. این ارزش‌ها بیان می‌کنند نتایج مثبت و منفی تستها با چه احتمالی صحیح هستند. محاسبه این سنجه نیز با توجه به جدول ممکن می‌باشد:

ارزش اخباری مثبت (PPV) نسبت افرادی است که تست آنها مثبت شده است و مبتلا به بیماری هستند.

PPV = (TP) تعداد مثبت حقیقی / کل مثبت‌ها

ارزش اخباری منفی (NPV) نسبت افرادی است که تست آنها منفی شده و مبتلا به بیماری مورد نظر نیستند.

NPV = (TN) تعداد منفی حقیقی / کل منفی‌ها

توجه داشته باشید که بر خلاف محاسباتی که تاکنون انجام شد، محاسبات مربوط به این ارزشها با استفاده از ردیف‌های افقی جدول صورت می‌گیرد و در نتیجه وابسته به شیوع بیماری مورد نظر در نمونهٔ مطالعه است.

هر چه بیماری شایعتر باشد احتمال صحیح بودن نتیجهٔ مثبت تست و ناصحیح بودن نتیجه منفی آن بیشتر است. ارزش اخباری یکی از سودمندترین سنجه‌ها برای استفاده در طبابت بالینی و تفسیر نتیجهٔ تست یک بیمار خاص است. ارزش اخباری در حقیقت

نشان می‌دهد فردی که به پزشک مراجعه کرده است، بسته به مثبت یا منفی بودن نتیجه تستش با چه احتمالی بیمار یا سالم است. در حالیکه ویژگی و حساسیت مستقیماً اطلاعی راجع به بیمار بدست نمی‌دهند. ارزش اخباری در صورتیکه احتمال پیش از تست برابر شیوع بیماری در نمونه مطالعه فرض شود برابر احتمال پس از تست برای بیماری مورد نظراست.

(با توجه به محاسبات گفته شده می‌توانید این مطلب را ثابت کنید!)

در مجموع قضاوت راجع به اندازهٔ اثر یا اهمیت بالینی مطالعه، بر مبنای سنجه‌های دقت تشخیصی تست امکانپذیر است. بخاطر معامله[1] موجود بین حساسیت و ویژگی قضاوت درباره اهمیت ارزش آنها به سادگی ممکن نیست. بطور کلی می‌توان گفت تستهایی که مجموع ویژگی و حساسیت آنها کمتر از ۱ است از نظر بالینی اطلاعات چندانی را در اختیار ما قرار نمی‌دهند. بهترین حالت این است که حساسیت و ویژگی برابر هم و در بیشینه مقدار خود باشند، در غیر اینصورت تستهایی که ارزش یکی از دو سنجه حساسیت یا ویژگی در آنها نزدیک به صددرصد است نیز با توجه به شرایط طبابت از اهمیت بالینی برخوردار هستند. (تستی که حساسیت بسیار بالایی دارد نتیجه آن برای یک بیمار منفی شده است، بخوبی می‌تواند رد کننده بیماری مورد نظر باشد و برعکس تستی که ویژگی بسیار بالایی دارد نتیجه آن برای یک بیمار مثبت شده است بخوبی می‌تواند مطرح کننده بیماری مورد نظر باشد.) در مورد ارزش‌های اخباری نیز توصیه مشابه ویژگی و حساسیت است.

به عنوان یک راهنمای کلی برای استفاده از نسبت درست نمایی در ارزیابی نقادانه شواهد مربوط به دقت تستهای تشخیصی در زمینه اهمیت بالینی می‌توان گفت نسبت درست نمایی مثبت (LR+ve) بیش از ۱۰ و نسبت درست نمایی منفی (LR+ve) کمتر از ۰/۱ می‌توانند شواهد متقاعد کننده‌ای برای رد یا قبول تشخیص فراهم کند. در حالیکه نسبت درست نمایی بالای ۵ و زیر ۰/۲ نیز به نوبه خود شواهد تشخیصی قدرتمندی در اختیار ما می‌گذارد. با این وجود اهمیت بالینی این سنجه به عوامل دیگری چون احتمال پیش از تست و شرایط خاص بیمار نیز بستگی دارد.

1- Trade off

سطح مطالعه

در بعضی مطالعات مانند این مطالعه ممکن است طرح مطالعه مستقیماً ذکر نشده باشد ولی با توجه به نحوه تقسیم‌بندی گروهها، طرح مطالعه با طرح مورد ـ شاهدی سازگارتر است. (شمارهٔ ۱)

METHODS

Once a diagnosis was established, patients were classified according to whether they had temporal arteritis, polymyalgia rheumatica, or negative histologic findings and a diagnosis other than temporal arteritis or polymyalgia rheumatica. In addition, control patients matched for age and sex to the patients with temporal arteritis were included.

شمارهٔ ۱

کیفیت مطالعه

- آیا تست تشخیصی در طیف مناسبی از بیماران مورد ارزیابی قرار گرفته است؟ (شمارهٔ ۲)

با توجه به اینکه ابتلا یا عدم ابتلا به آرتریت تمپورال در ۴ گروه مطالعه از قبل مشخص بود، بدین صورت که گروه اول مبتلا به نوع فعال آرتریت تمپورال بودند و سایر گروهها از جمله گروه کنترل با تشخیص‌های دیگری بجز آرتریت تمپورال وارد مطالعه شده بودند، مقایسه تست آزمایشی در این گروهها می‌تواند با تخمین بیش از مقدار واقعی برای دقت تست همراه باشد.

ولی با توجه به محتوای گروهها از نظر گستره تشخیصی به نظر می‌رسد طیف مناسبی از بیماران مورد ارزیابی قرار گرفته است.

METHODS

Patients

All consecutive patients seen in the departments of rheumatology and ophthalmology between January 1994 and October 1996 with clinically suspected active temporal arteritis and polymyalgia rheumatica as well as all patients who had temporal-artery biopsies to rule out the condition were included in the study.

شمارهٔ ۲

■ آیا تست مرجع استفاده شده از لحاظ بالینی مناسب است؟ (شمارهٔ ۳)

The American College of Rheumatology requires three of the following five criteria to be met to establish the diagnosis: age ≥50 years, new onset of localized headache, temporal-artery tenderness or decreased pulse, erythrocyte sedimentation rate ≥50 mm per hour, and histologic findings.

Biopsy

In all cases biopsies were bilateral. Temporal arteritis was diagnosed if there was vasculitis with a predominance of mononuclear-cell infiltration or granulomatous inflammation with or without giant cells.[1]

شمارهٔ ۳

ارزیابی تست مرجع معمولاً علاوه بر اطلاعات موجود در مقاله تا حدودی نیازمند دانش قبلی فرد ارزیابی کننده است. در این مطالعه نوع تست مرجع به روشنی بیان نشده است ولی نتایج تست آزمایشی با دو گروه از نتایج تحت عنوان تست مرجع مقایسه شده است و در پایان نیز سنجه‌های دقت تشخیصی بدست آمده در مورد هر کدام جداگانه محاسبه شده است. تست مرجع اول معیارهای کالج روماتولوژی آمریکا[1] است، که در مورد همه شرکت کنندگان سنجیده شده است و تست مرجع دوم بیوپسی شریان تمپورال است. بیوپسی شریان تمپورال به این علت که ممکن است نمونه از قسمتهای بدون ضایعه گرفته شود از نظر حساسیت احتمالاً به میزان کافی دقیق نیست.

1- American College of Rheumatology

■ آیا مقایسه نتایج تست آزمایشی و تست مرجع به طور کامل و مستقل انجام شده است؟ (شمارهٔ ۴)

RESULTS

TABLE 1. COMPARISON OF RESULTS OBTAINED BY ULTRASONOGRAPHY AND HISTOLOGIC ANALYSIS IN 30 PATIENTS WITH TEMPORAL ARTERITIS.*

No. of Patients	Ultrasonography		Histologic Analysis
	Halo	Stenosis or occlusion†	
14	+	+	+
3	+	+	ND
2	+	−	+
1	+	−	ND
1	+	+	−
1	+	−	−
4	−	+	+
1	−	+	−
1	−	+	ND
1	−	−	+
1	−	−	−

*ND denotes not done or that there was insufficient material for analysis. A plus sign indicates positive findings, and a minus sign negative findings.

†Twelve patients had both stenosis and occlusion, six patients had occlusion alone, and six patients had stenosis alone.

RESULTS

TABLE 2. SENSITIVITY AND SPECIFICITY OF DUPLEX ULTRASONOGRAPHY OF THE TEMPORAL ARTERIES FOR THE DIAGNOSIS OF TEMPORAL ARTERITIS AND TO CONFIRM HISTOLOGIC FINDINGS.

Finding	Diagnosis*		Confirmation of Histologic Findings†	
	Sensitivity	Specificity	Sensitivity	Specificity
	positive tests/total (%)	negative tests/total (%)	positive tests/total (%)	negative tests/total (%)
Halo	22/30 (73)	82/82 (100)	16/21 (76)	24/26 (92)
Stenosis or occlusion	24/30 (80)	76/82 (93)	18/21 (86)	23/26 (88)
Halo, stenosis, or occlusion	28/30 (93)	76/82 (93)	20/21 (95)	22/26 (85)

*Thirty patients had temporal arteritis, and 82 patients had been given other diagnoses.

†Twenty-one patients had positive histologic findings, and 26 patients had negative histologic findings (4 in the temporal-arteritis group, 7 in the group with polymyalgia rheumatica, and 15 with other diagnoses).

شمارهٔ ۴

همه شرکت کنندگان با معیارهای کالج روماتولوژی آمریکا (ACR) مورد بررسی قرار گرفته بودند که ۳۰ بیمار به آرتریت تمپورال و ۸۲ بیمار غیرمبتلا به آن تشخیص داده شدند. همه بیماران تحت تست آزمایشی (اولتراسونوگرافی داپلکس) نیز قرار گرفتند. ولی از میان شرکت کنندگان فقط ۵۷ نفر با بیوپسی شریان تمپورال تست شدند. بنابراین محاسبه سنجه‌های دقت تست آزمایشی در مقایسه با بیوپسی شریان تمپورال به علت احتمال تورش تأیید ناکامل از روایی بالایی برخوردار نخواهد بود.

ضمن اینکه قسمتی از تفاوت بین سنجه‌های محاسبه شده در جدول ۵ می‌تواند به تورش تأیید افتراقی مربوط باشد.

METHODS

Two trained physician ultrasonographers performed the examination, which takes an experienced sonographer 20 to 30 minutes. One investigator, who was unaware of the patients' diagnoses, looked only for the existence of a halo, stenosis, or occlusion. If their results disagreed, the investigators examined the results together and reached a consensus on the findings.

شمارهٔ ۵

- آیا مقایسه و تفسیر نتایج تست آزمایشی و تست مرجع بدون آگاهی از نتایج هر کدام بوده است؟ (شمارهٔ ۵)

در مورد کور بودن مقایسه تست آزمایشی و بیوپسی شریان تمپورال مطلبی در گزارش مقاله ذکر نشده است ولی کور بودن مقایسه نتایج تست آزمایشی با معیارهای ACR در حد ذکر شده قابل قبول به نظر می‌رسد.

- آیا انجام تست مرجع و تست آزمایشی پیش از شروع هر درمانی برای بیماری مورد نظر بوده است؟

در هر گروهی از بیماران تعدادی از آنها اخیراً تحت درمان با کورتیکواستروئید بوده‌اند ولی موردی وجود نداشته که بین انجام تست مرجع و آزمایشی تحت درمان قرار بگیرد. بنابراین وجود تورش تناقص درمان بعید به نظر می رسد.

- آیا نتایج تستهای همه شرکت‌کنندگان در تحلیل مطالعه در نظر گرفته شده است؟

با توجه به نتایج مطالعه (Table1, Table 2) نتایج تست‌های تمام شرکت کنندگان در تحلیل مطالعه آورده شده است.

اهمیت بالینی

- آیا مطالعه نشان‌دهنده توانایی کافی تست تشخیصی برای تعیین دقیق افراد مبتلا و غیرمبتلا به بیماری مورد نظر است؟

حساسیت و ویژگی یافته‌های اولتراسونوگرافیک در مقایسه با معیارهای ACR برای تشخیص آرتریت تمپورال برابر ۹۳٪ و در مقایسه با نتایج بیوپسی شریان تمپورال به ترتیب ۹۵٪ و ۸۵٪ بوده است (Table1, Table 2) که هر دو از لحاظ بالینی دارای اهمیت است. در این مطالعه سایر سنجه‌های دقت تشخیصی محاسبه نشده است.

در مجموع با وجود اینکه نتایج مطالعه از لحاظ بالینی از اهمیت کافی برخوردار است ولی به لحاظ روایی و کیفیت در سطح خیلی بالایی قرار نمی‌گیرد. استفاده از این مطالعه باید با توجه به شرایط بیمار و احتیاط کافی در تفسیر نتایج صورت گیرد.

References

1- Egger M, Smith GD, Altman DG, Systematic Reviews in Health Care: Meta-analysis in context. 2nd ed. BMJ Publishing Group 2001.

2- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine, How to Practice and Teach EBM, 2nd ed. Churchill Livingstone 2000.

3- Guyatt G, Rennie D, Hayward R. User's Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice.American Medical Association; 2002.

4- Riegelman RK. Studying a Study and Testing a Test. 4th ed. Lippincott Williams and Wilkins; 2000.

5- How to Review the Evidence: Systematic identification and review of the scientific literature. NHMRC; 1999.

فصل چهارم

شواهد مربوط به غربالگری

فصل چهارم
شواهد مربوط به غربالگری

در مطب خود نشسته‌اید، یک زوج ۵۲ ساله به شما مراجعه می‌کنند. یکی از دوستان آنها که تاکنون مشکلی نداشته است به علت علائم گوارشی مختصری به پزشک مراجعه کرده و بعد از انجام اقدامات تشخیصی برای وی سرطان کولون تشخیص داده شده است. این امر آنها را شدیداً مضطرب کرده است. دوست آنها که هم اکنون به علت عمل جراحی کلکتومی در بیمارستان بستری است، انجام تست خون مخفی در مدفوع[1] (FOBT) را به آنها توصیه کرده و تأکید نموده که پیشگیری بهتر از درمان است. اکنون این دو برای انجام این تست به شما مراجعه کرده‌اند. هیچ یک به تازگی تغییری را در اجابت مزاج ذکر نمی‌کنند و همچنین هیچ سابقه مثبتی در این زمینه ندارند. از قبل می‌دانستید که تست FOB قدرت غربالگری[2] دارد ولی در عین حال میزان مثبت‌های کاذب آن زیاد است (که در این صورت برای اطمینان از وضعیت بیمار باید کولونوسکپی انجام داد.)

خب، حال چه می‌کنید!؟ این زوج چندان مسن نیستند، علامتی ندارند و تقریباً جزء افراد با ریسک متوسط برای سرطان کولون محسوب می‌شوند. آیا توصیه به انجام تست FOB در این زوج کار صحیحی است؟

بهتر است برای آسانتر شدن جستجو ابتدا مشکل را به صورت یک سوال بالینی قابل پاسخگویی درآورید: در یک زوج ۵۲ سالهٔ بدون علامت، اثربخشی انجام تست FOB در کاهش مرگ و میر ناشی از سرطان کولون چقدر است؟

پس از جستجو در پایگاههای داده‌ای به مقاله‌ای از نشریهٔ پزشکی نیوانگلند[3] برمی‌خورید که به نظر بیشترین ارتباط را با سوال شما دارد. عنوان مقاله "اثر غربالگری با تست خون مخفی در مدفوع بر بروز سرطان کولورکتال"[4] می‌باشد که متن کامل آن در ضمیمهٔ

1- Fecal Occult Blood Test
2- Screening
3- New England Journal of Medicine
4- "The Effect of Fecal Occult-Blood Screening on the Incidence of Colorectal Cancers" (N Engl J Med 2000; 343: 1603-7)

شمارهٔ ۲ آمده است. حال نوبت ارزیابی مقاله با چک لیست زیر است که برای ارزیابی نقادانهٔ مقالات مربوط به غربالگری طراحی شده است.

آیا مطالعه انجام شده برای تعیین اثربخشی تست غربالگری از کیفیت مناسب برخوردار است؟

- تا چه حد تورشهای زمان تشخیص[1] و ازمان بیماری[2] در نظر گرفته شده است؟
- آیا گروهها درآغاز مطالعه از نظر عوامل موثر بر پیش‌آگهی با هم مشابه بودند؟
- آیا در ارزیابی داده‌ها به مقولهٔ کور[3] بودن توجه شده است؟
- آیا در مطالعه یک معیار مقایسه کننده تحت عنوان استاندارد طلایی[4] وجود دارد؟
- آیا فاصله‌ی زمانی لحاظ شده بین تست‌های غربالگری منطقی است؟
- آیا افرادی که برای انجام تست غربالگری انتخاب شده‌اند، شامل طیف مناسبی از جمعیت هستند؟

آیا مطالعه از دقت آماری کافی برخوردار است؟

- چقدر تخمین مطالعه برای اثر بخشی تشخیص زود هنگام از نظر آماری دقیق است؟

آیا نتایج این مطالعه از نظر بالینی اهمیت دارد؟

- آیا اندازهٔ اثری که برای تست غربالگری که در مطالعه نشان داده شده است اهمیت بالینی دارد؟
- آیا چنانچه بیماری مورد نظر تشخیص داده شود، درمان در دسترس است؟
- آیا آنچه که قرار است با غربالگری تشخیص داده شود، از لحاظ بالینی مهم است؟

1- Lead time bias
2- Length bias
3- Blinding
4- Gold Standard

منظور از غربالگری تشخیص بیماری در مراحل اولیه می‌باشد و جمعیت هدف آن افراد بدون علامت هستند. (برخلاف تستهای تشخیصی که در آن معمولاً افراد بیمار و علامتدار جمعیت هدف را تشکیل می‌دهند.)

در بسیاری موارد چنانچه نتیجهٔ تست غربالگری فردی مثبت شود، وی مبتلا به آن بیماری تلقی شده و به دنبال آن انجام اقدامات تشخیصی و درمانی برای فرد غیرقابل اجتناب خواهد بود ولی آیا همیشه غربالگری سلامت بیشتر را برای جامعه به همراه دارد؟

شواهد بسیاری مؤید این مطلب است که تست‌های غربالگری علاوه بر هزینه‌ای که به نظامهای سلامت تحمیل می‌کنند، به علت اینکه به یک فرد بدون علامت و ظاهراً سالم برچسب بیماری می‌زنند همیشه از نظر پایین آوردن کیفیت زندگی[1] افراد دردسرساز تلقی می‌شوند. به بیان دیگر تشخیص زودهنگام بیماری الزاماً منجر به بیشتر زنده ماندن افراد نمی‌شود ولی الزاماً منجر به افزایش سالهای زندگی توأم با بیماری خواهد شد. از طرف دیگر ممکن است بیمار به دلایلی قادر به انجام درمان نباشد یا بیماری تشخیص داده شده فاقد درمان شناخته شده و اثربخش باشد.

تمام این مشکلات در صورتی پیش می‌آید که تست غربالگری از دقت قابل قبولی برخوردار باشد. حال فرض کنید تست، نتایج مثبت کاذب یا منفی کاذب زیادی هم داشته باشد.

چنین نتایجی علاوه بر تأثیرات بالقوه سوء و طولانی مدت بر کیفیت زندگی افراد می‌تواند انجام رویکردهای بعضاً تهاجمی و پرهزینه را به فرد تحمیل کند و یا منجر به بدبینی وی نسبت به ارائه دهندگان خدمات سلامت شود، بنابراین با توجه به حساسیت‌های ذکر شده تصمیم‌گیری در مورد انجام یا عدم انجام یک تست غربالگری همیشه برای پزشکان مقوله‌ای چالش برانگیز بوده است.

1- Quality of Life

از این رو استفاده از بهترین شواهد در مورد غربالگری اهمیت زیادی دارد.اکنون که اهمیت استفاده از شواهد مربوط به غربالگری روشن شد در ادامه نکات مهم در ارزیابی این دسته از شواهد به طور کاملتر بحث می‌شود.

سطح شواهد

از آنجاییکه غربالگری در حقیقت مداخله‌ای است که برای کاهش عوارض و مرگ و میر بیماریها انجام می‌شود، بهترین نوع مطالعات از نظر سطح شواهد، در این زمینه کارآزمایی‌های تصادفی شدهٔ شاهددار[1] (RCT) هستند. چرا که این نوع مطالعات تورش‌های معمول در سنجش اثربخشی مداخلات را به حداقل می‌رسانند.

اگر در جستجویی که انجام داده‌اید کارآزمایی تصادفی شده شاهددار یافت نشد می‌توانید از طرحهای دیگر مانند مطالعات هم‌گروهی[2] استفاده کنید که طبیعتاً به لحاظ سطح شواهد در جایگاه دوم قرار خواهد گرفت. در مطالعهٔ نیوانگلند طرح مطالعه از نوع کارآزمایی تصادفی شدهٔ شاهددار ذکر شده است.(شماره ۱)

بعد از ارزیابی سطح شواهد باید کیفیت، دقت آماری و اهمیت بالینی نتایج مطالعه را بررسی کنید.

- **آیا مطالعهٔ انجام شده برای تعیین اثربخشی تست غربالگری از کیفیت مناسب برخوردار است؟**
- تا چه اندازه اثر تورشهای زمان تشخیص و ازمان بیماری در نظر گرفته شده است؟

METHODS

Participants

The Minnesota study was a randomized, controlled clinical trial conducted to evaluate the effectiveness of fecal occult-blood testing in reducing the rate of death from colorectal cancer.

شماره ۱

1- Randomized Controlled Trial
2- Cohort Study

از تورش‌هایی که به طور معمول در شواهد مربوط به غربالگری مطرح می‌شوند می‌توان به تورش زمان تشخیص و ازمان بیماری اشاره کرد.

ـ تورش زمان تشخیص

در بعضی شواهد مربوط به غربالگری (عمدتاً در مطالعات همگروهی) میزان بقا[1] (تعداد سالهای زنده بودن بعد از تشخیص) را در افرادی که غربالگری شده‌اند با افرادی که بدون غربالگری بیماریشان تشخیص داده شده است مقایسه می‌کنند. در این موارد ممکن است بدنبال غربالگری بقا افزایش یابد ولی نه به علت اینکه بیماران بیشتر زندگی می‌کنند بلکه بدین علت که غربالگری مدت زمانی را که آنها می‌دانند بیمار هستند، افزایش می‌دهد، به این اثر در اصطلاح تورش زمان تشخیص گفته می‌شود.

برای مثال فرض کنید ۲ فرد ۵۰ ساله را همزمان وارد مطالعهٔ همگروهی غربالگری کرده‌اید برای یکی تست غربالگری انجام می‌دهید و در سال اول به تشخیص می‌رسید. برای دیگری این کار را انجام نمی‌دهید و وی در سن ۷۵ سالگی علامت‌دار شده و بیماری‌اش تشخیص داده می‌شود.

چنانچه هردوی آنها در سن ۸۵ سالگی بمیرند بقای فرد اول پس از تشخیص ۳۵ سال و بقای فرد دوم ۱۰ سال است ولی در واقع تعداد سالهایی که هر دو زندگی کرده‌اند یکی بوده است. به همین علت در ارزیابی یک مطالعهٔ غربالگری، مدت زمان بین تشخیص تا مرگ اطلاعات صحیحی در زمینهٔ اثربخشی تست غربالگری بدست نمی‌دهد و بهتر است سراغ سنجه‌های دیگری از جمله مرگ و میر کلّی در هر سن خاص[2] برویم.

در این مطالعه کاهش بروز[3] سرطان کولون به عنوان سنجهٔ اثربخشی تست غربالگری مورد استفاده قرار گرفته است که احتمال تورش

1- Survival
2- Total age- Specific mortality
3- Incidence

زمان تشخیص را تا حد زیادی کاهش می‌دهد.(شماره ۲)

Statistical Analysis

The ratios of the 18-year cumulative incidence of colorectal cancer in the two screening groups to the incidence in the control group and the 95 percent confidence intervals for these ratios were computed as a measure of the extent to which screening affected incidence.

شمارهٔ ۲

METHODS

Briefly, 46,551 people, the majority of whom were 50 to 80 years old when recruited between 1975 and 1978, were randomly assigned to annual screening, biennial screening, or usual care (control group).

شمارهٔ ۳

ـ تورش ازمان بیماری

ماهیت بیماریها ناهمگون است. یک بیماری در برخی افراد با سیر آهسته پیشرفت می‌کند و در برخی سریع. مسلماً احتمال اینکه به هنگام غربالگری افرادی که بیماری آنها سیر آهسته‌تری دارد بیشتر شناسایی می‌شوند، بیشتر است، که به این اثر در اصطلاح تورش ازمان بیماری گفته می‌شود. با وجود اینکه پیش‌بینی سیر بیماری در افراد مورد مطالعه در ابتدای امر ممکن نیست،

ولی با تخصیص تصادفی افراد به دو گروه کنترل و مطالعه می‌توان فرض کرد که هر دو گروه نسبت برابری را از افراد با سیر بیماری آهسته و سریع در خود جای داده‌اند. در این مطالعه همانطور که ذکر شد تخصیص افراد به گروهها به صورت تصادفی بوده است. (شماره ۳)

- آیا گروهها در آغاز مطالعه از نظر عوامل موثر بر پیش‌آگهی با هم مشابه بودند؟

درصورتی می‌توان از کفایت تصادفی بودن مطالعه مطمئن شد که گروههای مطالعه تا حد امکان از نظر عوامل موثر بر پیش‌آگهی

مشابه هم باشند. برای این کار عمدتاً سن، جنس و سایر عوامل خطر شناخته شده برای بیماری در نظر گرفته می‌شود. تصادفی انتخاب کردن کمک می‌کند تا عوامل خطر شناخته نشده نیز به طور مشابه بین گروهها تقسیم شود.

در این مطالعه تعداد افراد و نسبت زن و مرد در هر گروه ذکر شده است ولی اطلاعات بیشتری مثلاً راجع به میانگین سنی افراد و سایر ویژگیها داده نشده است. بنابراین قضاوت در مورد مشابه بودن دو گروه از نظر عوامل موٴثر بر پیش‌آگهی ممکن نیست. (شماره ۴)

RESULTS

TABLE 1. RESULTS OF FECAL OCCULT-BLOOD TESTING ACCORDING TO STUDY GROUP AFTER 18 YEARS OF FOLLOW-UP.*

VARIABLE	ANNUAL SCREENING	BIENNIAL SCREENING	CONTROL
Subjects — no.			
Total	15,532	15,550	15,363
Men	7,474	7,430	7,419
Women	8,058	8,120	7,944
Person-years of observation — no.	235,584	235,513	232,612
Cases of colorectal cancer — no.	417	435	507
Cumulative incidence/1000	32 (29–35)	33 (30–36)	39 (36–43)
Cumulative incidence ratio†	0.80 (0.70–0.90)	0.83 (0.73–0.94)	1.0
P value‡	<0.001	0.002	—

*Values in parentheses are 95 percent confidence intervals.

†Values are the ratios of cumulative incidence in each screening group to that in the control group.

‡P values are for the comparison with the control group.

شمارهٔ ۴

■ آیا در ارزیابی داده‌ها به مقولهٔ کور بودن توجه شده است؟

بر خلاف مطالعات تصادفی در زمینهٔ درمان، کور بودن در غربالگری کامل (بطوریکه هم بیمار و هم پزشک نسبت به اطلاعات ناآگاه باشند) نیست. در غربالگری پزشک نمی‌تواند نسبت به نتایج مطالعه کور باشد، چرا که باید در صورت مثبت شدن تست، در مورد ارزیابی بیشتر و یا درمان با توجه به مرحلهٔ بیماری تصمیم‌گیری کند. ضمن اینکه در این مطالعه مطلبی راجع به کور بودن بیماران ذکر نشده است.

■ آیا در مطالعه، یک معیار مقایسه کننده تحت عنوان استاندارد طلایی وجود دارد؟

در شواهد مربوط به غربالگری پس از اینکه تست بر روی افراد گروه غربالگری انجام شد، اصولاً باید افرادیکه نتایج تست آنها مثبت شده است با یک تست مرجع به عنوان استاندارد طلایی نیز مورد آزمایش قرار بگیرند. همانطور که قبلاً هم گفته شد، استاندارد طلایی باید از دقت بسیار بالایی برخوردار باشد. تا بتوان براساس آن افراد سالم و بیمار را با اطمینان بالایی از هم افتراق داد.

در این مطالعه علاوه بر بررسیهای مختلف کلینیکی و پاراکلینیکی، کولونوسکوپی به عنوان یک استاندارد طلایی مناسب در تشخیص پولیپ‌ها و سرطانهای کولون مورد استفاده قرار گرفته است. (شماره ۵)

Study Design

Participants with at least one positive slide were invited to undergo a diagnostic evaluation, which initially included history taking and a physical examination, single-column barium enema, rigid proctosigmoidoscopy, urinalysis, a complete blood count, routine tests of blood chemistry, an upper gastrointestinal x-ray series, chest radiography, electrocardiography, and colonoscopy. Single-column barium enema was discontinued in 1978. Double-contrast barium enemas were administered to participants who had at least one positive screening test and whose colonoscopic examinations were incomplete or suboptimal (about 5 percent of the subjects). Rigid proctosigmoidoscopy and the upper gastrointestinal series were discontinued in 1982. Throughout the trial, colonoscopy was the dominant diagnostic procedure used to identify colorectal polyps and cancers.

شمارهٔ ۵

■ آیا فاصله‌های زمانی لحاظ شده بین تست‌های غربالگری منطقی است؟

فاصله‌های زمانی لحاظ شده بین تست‌های غربالگری بر اساس خصوصیات خود بیماری تعیین می‌شود. هر چه مرحله قبل از علامت‌دار شدن بیماری کوتاهتر باشد، فواصل کمتری برای غربالگری لازم است در بیماریهایی با دورهٔ نهفته طولانی، احتمال اینکه در غربالگری اولیه افراد بیشتری یافت شوند بیشتر است علاوه بر آن به نظر می‌رسد هر چه فاصلهٔ زمانی بین دو تست غربالگری کمتر باشد، به همان نسبت هم موارد بیماری بیشتری یافت می‌شود و برنامهٔ غربالگری ظاهراً در یافتن موارد جدید بیماری اثربخش‌تر است. در این مطالعه در یکی از گروهها، غربالگری به طور سالانه و در گروه دیگر غربالگری هر دو سال یکبار صورت می‌گرفت که با توجه به نوع بیماری قابل قبول است.

■ آیا افرادیکه برای انجام تست غربالگری انتخاب شده‌اند شامل طیف مناسبی از جمعیت هستند؟

به طور کلی تست‌های غربالگری براساس جمعیت هدف به دو دسته تقسیم می‌شوند:

الف) غربالگری جمعیت[1]: این دسته از تست‌ها همهٔ افراد را غربالگری می‌کنند، در نتیجه باید ارزان، در دسترس و ایمن باشند و بیماری هدف آنها نیز باید شایع یا خیلی مهم باشد. غربالگری سرطان پستان، سرطان دهانهٔ رحم، سرطان کولون و سونوگرافی دوران حاملگی از جمله مثالهای این دسته است.

ب) غربالگری گروههای پرخطر[2]: این دسته فقط گروههای پرخطر را ارزیابی می‌کنند و انجام آنها عمدتاً به خود بیمار بستگی دارد. داشتن عوامل خطر از جمله سابقهٔ بیماریهای ژنتیکی، سابقه مواجهه با بیماریهایی عفونی، افراد با شرایط شناخته شدهٔ پیش سرطانی، افراد مسن، ابتلاء به بیماری زمینه‌ای خاص و برخوردهای شغلی (برخورد با سرب، رنگ‌های شیمیایی و ...) افراد را مستعد انجام این تستها می‌کند.

1- Population Screening
2- High risk group screening

پس برای ما مهم است بدانیم بیماری مورد نظر در گروه اول غربالگری قرار می‌گیرد یا گروه دوم یا مثلاً چنانچه غربالگری سرطان کولون را که در گروه اول قرار می‌گیرد، فقط برای افراد پرخطر انجام دهیم از روایی و قدرت تعمیم‌پذیری مطالعه کاسته می‌شود.

در این مطالعه توضیحی راجع به نحوهٔ نمونه‌گیری از جمعیت مطالعه وجود ندارد.

آیا مطالعه از دقت آماری کافی برخوردار است؟

چقدر تخمین مطالعه از اثربخشی تشخیص زود هنگام از نظر آماری دقیق است؟

از آنجاییکه ممکن است تخمین اثربخشی بدست آمده در مطالعه ناشی از تصادف یا شانس باشد ارزیابی دقت آماری نیز ضرورت خاصی دارد؛ برای این کار، از مفهوم بازهٔ اطمینان[1] و P value استفاده کنید.

با توجه به P value کاملاً معنادار در حضور طول کوتاه بازهٔ اطمینان در این مطالعه، دقت آماری آن قابل قبول است.(شماره ۶)

RESULTS

The ratios of the cumulative incidence rates in the screening groups to that in the control group were 0.80 (95 percent confidence interval, 0.70 to 0.90; P<0.001) for the annual-screening group and 0.83 (95 percent confidence interval, 0.73 to 0.94; P=0.002) in the biennial-screening group (Table 1).

شمارهٔ ۶

آیا نتایج این مطالعه از نظر بالینی اهمیت دارد؟

- آیا اندازهٔ اثری که برای تست غربالگری در مطالعه نشان داده شده است اهمیت بالینی دارد؟

میزان اثربخشی تست غربالگری معمولاً با محاسبهٔ سنجه‌هایی چون کاهش ریسک مطلق[2] (ARR)، کاهش ریسک نسبی[3] (RRR)

1- Confidence Interval
2- Absolute Risk Reduction
3- Relative Risk Reduction

و تعداد لازم برای غربالگری[1] (NNS) برآورد می‌شود. نحوهٔ محاسبهٔ این شاخص‌ها در فصل بعد بطور کامل بیان شده است. (NNS تعداد افرادی است که لازم است غربالگری شوند تا از وقوع یک پیامد نامطلوب مانند ابتلا به سرطان کولون جلوگیری شود.)

در بعضی مطالعات به غربالگری از دیدگاه یک تست تشخیصی نگاه شده است و مانند مطالعات مربوط به دقت تستهای تشخیصی میزان اثربخشی در قالب سنجه‌های تعیین دقت تشخیصی مثل ارزش اخباری مثبت[2] (PPV) بیان می‌شود. این مطلب در مورد مطالعهٔ مورد بررسی ما نیز صدق می‌کند. (شماره ۷)

For the annual-screening group, the positive predictive value ranged from 0.87 percent for one positive slide to 4.53 percent for six positive slides (Spearman's rho=0.94, P=0.02). For the biennial-screening group, the positive predictive value ranged from 1.12 percent for one positive slide to 6.13 percent for six positive slides (Spearman's rho=0.94, P=0.02).

شمارهٔ ۷

- چنانچه بیماری مورد نظر تشخیص داده شود، آیا درمان در دسترس است؟

اصولاً وقتی می‌توان درباره کارایی یک تست غربالگری نظر داد که بدنبال تشخیص، درمان نیز در دسترس باشد و بتواند پیامدهای بیمار را تغییر بدهد. البته گاهی اوقات این خود بیمار است که در فرآیند درمان شرکت نمی‌کند.

در این مطالعه ضمن انجام غربالگری پولیپ‌های پیدا شده

DISCUSSION

Our findings demonstrate a significant reduction in the incidence of colorectal cancer after fecal occult-blood testing. This reduction occurred after both annual and biennial screening. The most plausible explanation is the identification and removal of the precursor lesions for colorectal cancer — that is, adenomatous polyps.

شمارهٔ ۸

1- Number Needed to Screen
2- Positive Predictive Value

که ضایعات پیش سرطانی هستند به طور روتین با کولونوسکوپی برداشته می‌شدند که این کار در درازمدت به کاهش بروز سرطان کولون انجامیده است.(شماره ۸)

- آیا آنچه که قرار است با غربالگری تشخیص داده شود، از لحاظ بالینی مهم است؟

کاستن از مرگ و میر بیماریها با تشخیص زود هنگام اهمیت زیادی دارد. بنابراین انجام تست غربالگری برای بیماریهایی که موربیدیته و مرگ و میر زیادی ندارند چندان مفید نیست. بعلاوه گاهی اوقات بعد از تشخیص بیماری، انجام درمان فایده چندانی ندارد؛ به عنوان مثال در یک پیرمرد ۸۰ ساله تشخیص زود هنگام سرطان تیروئید از لحاظ بالینی اهمیت چندانی ندارد؛ چرا که فرد یا با سرطان تیروئید می‌میرد یا از سرطان تیروئید.

در این مطالعه با توجه به موربیدیته و مرگ و میر بالای سرطان کولون و شیوع نسبتاً زیاد آن غربالگری توجیه‌پذیر است.

در مجموع می‌توان گفت مطالعهٔ مورد بررسی از لحاظ طرح مطالعه یک RCT است که در بالاترین سطح شواهد مربوط به غربالگری قرار می‌گیرد. این مطالعه براساس چک لیست ارزیابی نقادانه از نظر کیفیت و دقت آماری نسبتاً قابل قبول است.

بعلاوه اهمیت بالینی نتایج مطالعه تا حد ذکر شده راضی کننده است، بنابراین می‌توانید براساس آن برای بیمارتان تصمیم‌گیری کنید.

References

1- Guyatt G, Rennie D, Hayward R. User's Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice.American Medical Association; 2002.

2- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine, How to Practice and Teach EBM, 2nd ed. Churchill Livingstone 2000.

3- Riegelman RK. Studying a Study and Testing a Test. 4th ed. Lippincott Williams and Wilkins; 2000.

4- Welch HG, Blach WC. Evaluating Randomised Trials of Screening, 2002.

فصل پنجم

شواهد مربوط به درمان

فصل پنجم
شواهد مربوط به درمان

یک پزشک بازنشسته که ۶۹ سال دارد با شما در مورد تشدید سوء هاضمهٔ خود که به مدت ۲۰ سال به صورت خفیف وجود داشته، مشاوره می‌کند. او اضافه وزن دارد ولی در کل در وضعیت سلامت خوبی است. وی هم‌اکنون روزی ۳ کپسول امپرازول ۲۰ میلی گرمی می‌خورد، در هفته ۱ واحد الکل می‌نوشد و برای کنترل درد مفصلی در هفته بین ۲ تا ۳ بار از ناپروکسن استفاده می‌کند.

اندوسکوپی انجام شده از مری، معده و دوازدهه یک هرنی هیاتال کوچک ۵ سانتی‌متری و گاستریت خفیف آنترال را نشان می‌دهد. اما بجز این دو مورد بقیه قسمتها نرمال است. در طی اندوسکوپی یک بیوپسی از ناحیه آنترال تهیه می‌کنید و تست CLO[1] را برای هلیکوباکتر انجام می‌دهید که کاملاً مثبت است. بررسی بافت‌شناسی گاستریت و عفونت با هلیکوباکتر را تأیید می‌کند.

می‌خواهید بدانید آیا شواهد جدیدی در مورد اثربخشی از بین بردن هلیکوباکتر در بیماران مبتلا به سوء هاضمه بدون زخم وجود دارد یا خیر.

برای آغاز جستجو ابتدا سوال بالینی خود را به شکل زیر درمی‌آورید.

آیا در بیماران مبتلا به سوء هاضمهٔ بدون زخم و عفونت با هلیکوباکترپیلوری، درمان ریشه کنی هلیکوباکتر به کاهش علائم سوء هاضمه می‌انجامد.

با جستجو در پایگاه داده‌ای پاب مد[2] با واژه‌های "سوء هاضمهٔ بدون زخم"[3] و "هلیکوباکترپیلوری"[4] و "RCT" به مقاله‌ای با عنوان "بهبود نشانه‌ها

1- Urease Slide Test
2- PubMed
3- Nonulcer dyspepsia
4- Helicobacter pylori

در نتیجهٔ ریشه‌کنی عفونت هلیکوباکترپیلوری در بیماران مبتلا به سوءهاضمهٔ غیر زخمی"[1] می‌رسید که به نظر می‌رسد دقیقاً به سوال شما مربوط است. متن کامل این مقاله در ضمیمهٔ شماره ۳ آورده شده است. برای ارزیابی این مطالعه از چک لیست زیر استفاده کنید.

- **آیا مطالعه از سطح مناسبی برای پاسخ دادن به سوالات درمانی برخوردار است؟**
- **آیا مطالعه انجام شده برای تعیین اثربخشی مداخلهٔ درمانی از کیفیت مناسبی برخوردار است؟**
  - آیا افراد دو گروه کنترل و مداخله بجز درمان از نظر سایر خصوصیات یکسان بوده‌اند؟
  - آیا لیست افراد دو گروه مداخله و کنترل مخفی نگاه داشته شده است؟
  - جدا از مداخله‌ای که تحت بررسی بوده آیا بیماران به شیوه یکسانی مراقبت می‌شدند؟
  - آیا بیماران هر دو گروه برای ظاهر شدن اثرات مداخله به اندازه کافی پیگیری شده‌اند؟
  - آیا همه بیماران در مدت زمان معین شده تحت پیگیری بوده‌اند؟
  - ارزیابی پیامدهای بیماران ذهنی بوده است یا عینی؟
  - آیا کسانی که اثرات مداخله را ارزیابی می‌کردند از نوع گروه مطالعه (مداخله یا کنترل) مطلع بوده‌اند؟
  - آیا افراد گروه کنترل و مداخله به یک شکل مورد ارزیابی قرار گرفته‌اند؟
  - آیا از سنجه مناسبی برای ارزیابی اثرات مداخله استفاده شده است؟
  - آیا هر یک از بیماران (گروه مداخله یا کنترل) در گروهی که به طور تصادفی در آن قرار گرفته‌اند، تحلیل شده‌اند؟

1- Symptomatic benefit from eradicating Helicobacter pylori in patients with nonulcer dyspepsia N EngIJ Med 1998; 339: 1869-74)

- آیا افرادیکه (چه از گروه مداخله و چه از گروه کنترل) به هر علتی نتوانسته‌اند تا انتها در مطالعه شرکت کنند در ارزیابی لحاظ شده‌اند؟ (آیا تحلیل مطالعه بر مبنای تحلیل قصد درمان[1] انجام شده است؟)

آیا دقت آماری در نظر گرفته شده برای محاسبهٔ اثر درمانی قابل قبول است؟

P value

بازهٔ اطمینان

آیا این نتایج روا از نظر بالینی دارای اهمیت است؟

دستهٔ بزرگی از سوالاتی که ممکن است در طبابت برای پزشک مطرح شود در مورد اثربخشی مداخلات می‌باشد.

این مداخلات در طبابت طیف وسیعی را شامل می‌شود؛ از جمله مداخلات درمانی، مداخلات توانبخشی، تشخیص زودهنگام بیماریها (غربالگری) و آموزش بیماران. از این بین، مداخلات درمانی طبق شواهد بیشترین سوالات بالینی را به خود معطوف کرده است. درمانهای دارویی، جراحی، تغییر رژیم غذایی، تغییر شیوه زندگی، روان درمانی و ... مثالهایی از مداخلات درمانی هستند. در این فصل نحوهٔ ارزیابی نقادانهٔ شواهد مربوط به اثر بخشی مداخلات با تأکید بر مداخلات درمانی بیان خواهد شد.

آیا مطالعه از سطح مناسبی برای پاسخ دادن به سوالات درمانی برخوردار است؟

بهترین مطالعات برای ارزیابی اثربخشی یک مداخله مقالات مروری نظام‌مند هستند، که در آنها از نتایج چندین کارآزمایی تصادفی شدهٔ شاهددار (RCT) استفاده شده است. بنابراین بهتر است برای یافتن پاسخ سوال بالینی خود ابتدا این مطالعات را جستجو کنید. گزینهٔ بعدی RCTها هستند که بعد از مقالات مروری نظام‌مند، بهترین منبع شواهد برای ارزیابی اثربخشی مداخلات

1- Intention to treat analysis

محسوب می‌شوند، چرا که فرایند تخصیص تصادفی افراد به گروههای مطالعه بالقوه خطای تصادفی و تورش را به حداقل می‌رساند. مقالات مروری نظام‌مندی که از مطالعات همگروهی استفاده کرده‌اند نیز از نظر سطح شواهد در جایگاه بعدی قرار می‌گیرند. سایر طرحهای پژوهشی مانند مطالعات هم گروهی و مورد شاهدی از این نظر ضعیف‌تر هستند و بهتر است استفاده از آنها به شرایط خاص محدود شود.

برای مطالعه مورد بررسی همانطور که در قسمت روشها توضیح داده شده است طرح مطالعه از نوع RCT است که از نظر سطح شواهد در رتبه بالایی قرار دارد.(شماره ۱)

Methods We performed a randomized, placebo-controlled trial comparing the efficacy of treatment for two weeks with 20 mg of omeprazole orally twice daily, 500 mg of amoxicillin three times daily (with 500 mg of tetracycline three times daily substituted for amoxicillin in patients allergic to penicillin), and 400 mg of metronidazole three times daily (160 patients) with that of omeprazole alone (158 patients) for resolving symptoms of dyspepsia in patients with *H. pylori* infection but no evidence of ulcer disease on upper gastrointestinal endoscopy.

شمارهٔ ۱

آیا مطالعهٔ انجام شده برای تعیین اثربخشی مداخلهٔ درمانی از کیفیت مناسبی برخوردار است؟

- آیا افراد در گروههای کنترل و مداخله بجز در دریافت درمان از نظر سایر خصوصیات یکسان بوده‌اند؟

تصادفی بودن تخصیص افراد به گروههای کنترل و مداخله الزاماً تضمین کننده معادل بودن این گروهها نیست بلکه به خاطر پدیدهٔ شانس گاهی اوقات ممکن است نسبت افرادیکه ویژگی خاصی (مانند جنس مذکر) دارند در یکی از گروهها بیشتر باشد.

بنابراین در اکثر مطالعات RCT ویژگیهایی که به نظر می‌رسد در پیش‌آگهی بیماران موثر هستند (مانند سن، جنس، نژاد، شرایط جغرافیایی، عوامل خطر، شدت بیماری، بیماریهای زمینه‌ای و ...) در یک جدول با هم مقایسه می‌شوند. در صورتیکه تفاوت بین این عوامل موثر در پیش‌آگهی که اصطلاحاً به آنها عوامل مداخله‌گر گفته می‌شود معنادار باشد باید اثر این متغیرها به نوعی در تحلیل پژوهش در نظر گرفته شود.

در این مطالعه گروه کنترل و گروه مداخله از نظر کلیهٔ عوامل شناخته شده موثر بر پیش‌آگهی مشابه هم بوده‌اند و این مطلب در گزارش مقاله به خوبی ذکر شده است. (شماره ۲)

شمارهٔ ۲

The two groups were well matched with respect to age, sex, initial results on the Glasgow Dyspepsia Severity Score, prevalence of previously prescribed antisecretory-drug therapy, prevalence of smoking, predominant symptom at presentation, duration of symptoms, and quality of life (Table 1).

TABLE 1. BASE-LINE CHARACTERISTICS OF 318 PATIENTS WITH NONULCER DYSPEPSIA.*

CHARACTERISTIC	OMEPRAZOLE AND ANTIBIOTICS (N=160)	OMEPRAZOLE ALONE (N=158)
Male sex (%)	51	47
Age (yr)		
Mean ±SD	42±12	42.2±13
Range	18–70	17–70
Glasgow Dyspepsia Severity Score†		
Mean ±SD	11.4±2.2	11.5±2.5
Range	6–17	4–18
Qualify-of-life score‡		
Mean ±SD	520±176	507±164
Range	25–786	93–764
Duration of symptoms (%)§		
<2 yr	39	38
2–5 yr	25	21
>5–10 yr	11	16
>10 yr	24	25
Predominant symptom (%)		
Epigastric pain	57	54
Retrosternal pain	21	16
Reflux	12	13
Other	10	17
History of smoking in past 6 mo (%)	34	33
Medication prescribed in past 6 mo (%)	84	80
Histamine H_2-receptor antagonists	59	54
Proton-pump inhibitors	13	11
Both	8	11
Other	4	4
Medication prescribed for >3 of the past 6 mo (%)	44	45

*There were no significant differences between the two groups.

†Scores can range from 0 (no symptoms or limitations) to 20 (severe symptoms most days, requiring regular medication).

‡Quality of life was measured with the SF-36; scores can range from 0 (worst) to 800 (best).

§Because of rounding, percentages may not total 100.

■ آیا انتخاب بیماران برای قرار گرفتن در گروه مداخله یا کنترل تصادفی بوده است؟

برای اینکه تخصیص بیماران برای قرار گرفتن در گروه مداخله و کنترل تصادفی باشد روش‌های متنوعی وجود دارد ولی مشخصهٔ مشترک همهٔ این روشها این است که احتمال انتخاب یک نفر برای دریافت (گروه مداخله) یا عدم دریافت درمان (گروه کنترل) یکسان و برابر می‌باشد. ذکر روش تخصیص تصادفی که در مطالعه استفاده شده است از نظر ارزیابی گزارش مقاله اهمیت زیادی دارد.

METHODS

Recruitment and Evaluation of Patients

The randomization was carried out independently by the pharmacy department, which used tables of random numbers to assign half the patients to receive active treatment with omeprazole, amoxicillin, and metronidazole and half to receive omeprazole and placebo.

شمارهٔ ۳

در مطالعهٔ مورد بررسی ذکر شده است که تخصیص بیماران به دو گروه مداخله و کنترل به صورت کاملاً تصادفی بوده و برای این کار از جدول اعداد تصادفی استفاده شده است.(شماره ۳)

در طراحی یک کارآزمایی تصادفی شده استفاده از روشی برای پیشگیری از مشخص شدن اسامی افراد گروه مداخله و کنترل پیش از ارائه مداخلهٔ آزمایشی بسیار با اهمیت است. این کار به منظور جلوگیری از اعمال نظر پژوهشگران و شرکت‌کنندگان مطالعه در انتخاب تصادفی انجام می‌شود.

اگر لیست افرادی که در هر گروه قرار می‌گیرند مخفی نباشد امکان دارد که به هر دلیلی بیمار یا پژوهشگر تمایل خاص خود را در مورد قرار گرفتن در یک گروه خاص اعمال کند. مثلاً پژوهشگر برای اینکه روش درمانی خود را موثرتر از آنچه در واقعیت هست جلوه دهد بیماران را بر اساس عوامل موثر بر پیش‌آگهی به گروههای مداخله و کنترل تقسیم می‌کند بدین ترتیب که در گروه مداخله

افرادی را قرار می‌دهد که به نظر خودش به آن درمان بهتر پاسخ می‌دهند. از طرف دیگر بیماری که از احتمال اثربخش بودن مداخلهٔ آزمایشی آگاه است ممکن است سعی کند به هر طریقی در گروه درمان قرار بگیرد. بدیهی است چنین اتفاقاتی از روایی نتایج مطالعه می‌کاهد. آگاه نبودن بیمار و پژوهشگر از نوع گروه مطالعه (مداخله یا کنترل) تا قبل از ارائه مداخله درمانی اغلب امکانپذیر است در حالیکه بعد از آغاز درمان در بعضی موارد مانند مداخلات جراحی این امر امکانپذیر نیست. براساس شواهد تجربی، مخفی نگه داشتن لیست افراد[1] پیش از ارائه مداخلهٔ آزمایشی بیش از سایر اجزاء انتخاب تصادفی، تورش انتخاب[2] را کاهش می‌دهد.[3]

در کارآزماییهای بالینی باید اطلاعات مربوط به مخفی نگه داشتن لیست افراد پیش از ارائه مداخله، به طور کامل ذکر شده باشد. در اینصورت می‌توان در مورد روش مخفی نگه داشتن لیست در آنها قضاوت کرد. ذیلاً به چند نمونه از رویکردهای مطمئنی که برای مخفی ماندن لیست انتخاب می‌توان مورد استفاده قرار داد، اشاره شده است:

۱ استفاده از پاکت‌های نامه که کاملاً مات هستند (منظور این است که در نور نمی‌توان از محتویات آن آگاه شد)، درون این پاکت‌ها، مشخصات بیمار نوشته شده و سپس به طور کامل بسته می‌شود و روی آنها به ترتیب شماره می‌خورد، انتخاب بیماران بر اساس شمارهٔ هر پاکت و بدون آگاهی از محتویات آن خواهد بود.

۲ استفاده از نرم‌افزارهای رایانه‌ای که لیست افراد در آنها به صورت یک فایل قفل شده و غیرقابل خواندن ذخیره می‌شود و تنها بعد از وارد کردن مشخصات فرد شرکت کننده اطلاعات قابل دستیابی خواهد بود.

۳ استفاده از محفظه‌های شماره دار مشابه[4] که به طور سریالی بین شرکت کنندگان توزیع می‌شود.

۴ تخصیص مرکزی[5] بیماران، مثلاً تخصیص بیماران به گروه کنترل و مداخله توسط یک اداره مرکزی که از مشخصات افراد آگاهی ندارد و یا انتخاب تصادفی بیماران در محل تولید یا پخش دارو.[6]

1- Allocation concealment
2- Selection bias
3- Chalmers 1983, Schulz 1995, Moher 1998
4- Prenumbered or coded identical containers
5- Central allocation
6- Pharmacy

تعدادی از رویکردهای مورد استفاده برای مخفی نگه داشتن تخصیص بیماران ناکافی در نظر گرفته می‌شود. مثلاً تخصیص بیماران براساس روز تولد، روز مراجعه یا هر روش دیگری که در آن انتخاب فرد برای گروهی خاص از قبل مشخص باشد.

به عنوان یک توصیه، چنانچه در گزارش مطالعه توضیحی در مورد نحوهٔ مخفی نگه داشتن لیست افراد نیافتید، این مورد را ناکافی در نظر بگیرید. در این مطالعه هم بیماران و هم ارائه دهندگان درمان از نحوهٔ تخصیص هر فرد به گروه درمان یا کنترل بی‌اطلاع بوده‌اند. نحوه مخفی نگه داشتن انتخاب تصادفی در مطالعه اینگونه ذکر شده است:(شماره ۳)

METHODS

Recruitment and Evaluation of Patients

The randomization was carried out independently by the pharmacy department, which used tables of random numbers to assign half the patients to receive active treatment with omeprazole, amoxicillin, and metronidazole and half to receive omeprazole and placebo.

شمارهٔ ۳

- جدا از مداخله‌ای که تحت بررسی بوده آیا بیماران به شیوهٔ یکسانی مراقبت می‌شدند؟

یکی دیگر از معیارهای بررسی روایی این است که مطمئن شویم به جز مداخله‌ای که تحت بررسی است اختلافی در مراقبت ارائه شده به دو گروه کنترل و درمان وجود نداشته است. برای جلوگیری از این اختلاف ناخواسته بین گروه مداخله و کنترل ارائه دهندگان و دریافت کنندگان این مراقبت‌ها بهتر است از گروهی که بیمار به آن تعلق دارد آگاه نباشند.[1]

این مورد در مطالعهٔ مورد بررسی به خوبی رعایت شده است.(شماره ۴)

The active and placebo tablets for each antibiotic were identical in appearance, and the trial was conducted in a double-blind fashion.

شمارهٔ ۴

با این حال در مطالعه ذکر شده است که بیماران

1- Patient and care provider blinding

می‌توانستند از هر دارویی که می‌خواهند پس از پایان درمان آزمایشی استفاده کنند. این کار می‌تواند بالقوه در نتایج مطالعه تورش ایجاد کند.(شماره ۵)

They were also told that they could take any medication necessary, including a proton-pump inhibitor, if they had recurrent or persistent symptoms after the completion of treatment. However, their doctors were requested by letter not to prescribe any treatment to eradicate *H. pylori*.

شمارهٔ ۵

- آیا بیماران هر دو گروه برای ظاهر شدن اثرات مداخله به اندازهٔ کافی پیگیری شده‌اند؟

پاسخ این سوال در هر مداخله‌ای براساس پاتوفیزیولوژی اثر آن مداخله متفاوت خواهد بود، منظور این است که در مطالعه از لحاظ زمانی فرصت کافی برای ظاهر شدن اثرات مداخله وجود داشته باشد. جواب این سوال منوط به داشتن اطلاعات کافی در مورد مداخله است مثلاً در مورد ارزیابی اثر یک دارو دانستن مدت زمان شروع اثر دارو رسیدن آن به حداکثر اثربخشی لازم است.

Symptoms were assessed with the Glasgow Dyspepsia Severity Score, with resolution of symptoms defined as a score of 0 or 1 in the preceding six months (maximal score, 20). One year later the patients were assessed to determine the frequency of the resolution of symptoms.

شمارهٔ ۶

در این مطالعه زمان پیگیری بیماران یک سال بوده است، که با توجه به مداخلات انجام شده کافی به نظر می‌رسد.(شماره ۶)

- آیا همهٔ بیماران در مدت زمان تعیین شده تحت پیگیری بوده‌اند؟

همهٔ افراد شرکت کننده در مطالعه چه در گروه مداخله وچه در گروه کنترل باید تا انتهای دورهٔ تعیین شده برای پیگیری تحت

نظر باشند و مراقبت‌ها و ارزیابی‌های مد نظر روی آنها اعمال شود. اگر فردی از این مراقبت‌ها و ارزیابی‌ها بازمانده باشد باید علت آن دقیقاً بررسی شود، زیرا ممکن است علت آن برای مثال عارضهٔ جانبی مداخلهٔ درمانی باشد. از آنجاییکه معمولاً در گزارش مطالعه اطلاعات کافی در مورد علت از دست دادن افراد داده نمی‌شود باید در این مورد محتاط بود.

در این مطالعه، با توجه به اینکه ریزش بیماران در طول پیگیری فقط ۳ درصد و تقریباً در هر دو گروه یکسان بوده است بنابراین می‌توان فرض کرد این تعداد حذف از پیگیری قابل قبول بوده است. (شماره ۷)

Ten patients were lost to follow-up: six patients in the group assigned to omeprazole and antibiotics and four patients in the group assigned to omeprazole alone.

شمارهٔ ۷

- آیا نحوهٔ ارزیابی بیماران ذهنی[1] بوده است یا عینی[2]؟

در مواردی که ارزیابی پیامدها از نوع عینی است می‌توان تا حدودی از تداخل نظرات بیمار و فردی که اثرات مداخله را در بیماران ارزیابی می‌کند در نتایج مطالعه اطمینان حاصل کرد.

در این مطالعه از مقیاس شدت سوءهاضمه «گلاسگو»[3] استفاده شده است که بعضی معیارهای آن از نوع عینی (مانند دفعات ویزیت داخلی توسط پزشک) و بعضی از نوع ذهنی (مثل اثر سوءهاضمه روی فعالیت طبیعی زندگی فرد) است.

تست تنفسی اوره نیز یک ماه بعد از پایان درمان و یک سال بعد از درمان برروی بیماران انجام شده است که یافته‌های آن از نوع عینی می‌باشد.

- آیا کسانی که اثرات مداخله را ارزیابی می‌کردند از نوع گروه مطالعه (مداخله یا کنترل) مطلع بوده‌اند؟[4]

1- Subjective
2- Objective
3- Glasgow dyspepsia severity score
4- Outcome assessor blinding

مطالعاتی که در آنها ارزیابی کننده اثرات مداخله از نوع گروه شرکت‌کنندگان مطلع نیست نسبت به سایر مطالعات از روایی بالاتری برخوردارند. این ناآگاهی بخصوص در مطالعاتی که اثرات مداخله به طور ذهنی ارزیابی می‌شود (مانند درد و کیفیت زندگی) اهمیت بیشتری خواهد داشت.

در این مورد در مطالعهٔ مورد بررسی توضیحی داده نشده است.

■ آیا افراد گروه کنترل و مداخله به یک شکل مورد ارزیابی قرار گرفته‌اند؟

در این مطالعه برای ارزیابی پیامدها در هر دو گروه از تست تنفسی اوره، مقیاس گلاسگو و مقیاس کیفیت زندگی استفاده شده است. (شماره۸)

■ آیا از سنجهٔ مناسبی برای ارزیابی اثرات مداخله استفاده شده است؟

مناسب بودن سنجه براساس اینکه چه اثری از مداخله مدنظر است تعیین می‌شود. معمولاً روایی سنجه‌های اثربخشی اثر مداخله باید در مطالعات قبلی تایید شده

Of the 154 patients in the group assigned to receive omeprazole and antibiotics, 150 had a urea breath test one month after the completion of treatment. The test was negative in 132 patients (88 percent), of whom 125 (95 percent) also had a negative test at one year. Of the 154 patients in the group assigned to receive omeprazole alone, 152 had a urea breath test one month after the end of treatment. The test was positive in 145 patients (95 percent), of whom 134 (92 percent) also had a positive test at one year.

The distribution of Glasgow Dyspepsia Severity Scores in the two groups at one year is shown in Figure 1.

شمارهٔ ۸

باشد، این امر غالباً در قسمت بحث انتهایی مطالعه مطرح می‌شود، که در این مطالعه در قسمت تحلیل آماری آورده شده است.(شماره۹)

Statistical Analysis

The main end point was the resolution of symptoms, defined as a score of 0 or 1 on the Glasgow Dyspepsia Severity Score, because validation studies indicated that this cutoff point was the most appropriate indication of the resolution of dyspepsia.

شمارهٔ ۹

- آیا هر یک از بیماران (گروه مداخله یا کنترل) در گروهی که به طور تصادفی در آن قرار گرفته‌اند، تحلیل شده‌اند؟
- آیا افرادیکه (چه از گروه مداخله و چه از گروه کنترل) به هر علتی نتوانسته‌اند تا انتها در مطالعه شرکت کنند در ارزیابی لحاظ شده‌اند؟

گاهی اوقات به علل مختلف پژوهشگر مجبور می‌شود بعضی بیماران را از گروهی که در بدو مطالعه بطور تصادفی در آن قرار گرفته‌اند به گروه دیگر یا خارج از مطالعه انتقال دهد. این کار می‌تواند در نتایج مطالعه تورش ایجاد کند. برای جلوگیری از این تورش پژوهشگران در تحلیل پژوهش از روشهایی استفاده می‌کنند که پیامدهای مربوط به این افراد را در گروه اولیه آنها تحلیل می‌کند. به این روشها در اصطلاح تحلیل قصد درمان گفته می‌شود. به عبارت دیگر تحلیل به قصد درمان تحلیلی است که بر پایه آن تمام شرکت کنندگان در یک کارآزمایی بنابر گروهی که ابتدا به آن تخصیص داده شده‌اند مورد تحلیل قرار می‌گیرند؛ حال چه آن مداخله را دریافت کرده باشند، چه دریافت نکرده باشند، تحلیل قصد درمان در ارزیابی اثربخشی مداخلات اهمیت ویژه‌ای دارد چرا که از این طریق اثر عدم تبعیت بیمار[1] از رژیم درمانی و تغییرات درمانی که ممکن است در حین انجام آن مداخله در طبابت بوجود آید، حذف می‌شود و اینگونه احتمال تورش پیگیری هنگامیکه شرکت کنندگان از تحلیل مطالعه خارج می‌شوند، کاهش می‌یابد.

1- Non compliance

در این مطالعه افراد دو گروه تا انتها در گروههای خود پیگیری شده و در نتیجه‌گیری نهایی لحاظ شده‌اند.(شماره ۱۰)

شمارهٔ ۱۰

analysis was based on the intention-to-treat principle. The

آیا دقت آماری در نظر گرفته شده برای محاسبهٔ اثر درمانی قابل قبول است؟

در این قسمت میزان نزدیک بودن اثر درمانی محاسبه شده به واقعیت مورد ارزیابی قرار می‌گیرد. برای این کار باید از P value و بازهٔ اطمینان که در تخمین اثر واقعی درمان بسیار مهم هستند استفاده کنید.

شمارهٔ ۱۱

On an intention-to-treat basis, the symptoms resolved (defined as a score of 0 or 1) in 33 of 154 patients (21 percent) in the group assigned to receive omeprazole and antibiotics, as compared with 11 of 154 patients (7 percent) in the group assigned to receive omeprazole alone (95 percent confidence interval for the difference between groups, 7 to 22 percent; P<0.001).

TABLE 2. OUTCOMES OF THE PATIENTS AT ONE YEAR.*

CHARACTERISTIC	OMEPRAZOLE AND ANTIBIOTICS (N=154)	OMEPRAZOLE ALONE (N=154)	P VALUE
Resolution of symptoms (%)	21	7	<0.001
Glasgow Dyspepsia Severity Score†			
Mean ±SD	5.4±4.0	6.2±3.6	0.07
Range	0–14	0–16	
Quality-of-life score‡			
Mean ±SD	566±179	556±175	0.61
Range	50–786	113–776	
H. pylori–negative (%)	85	12	<0.001
Medication prescribed in past 6 mo (%)	43	53	0.09
Histamine H_2-receptor antagonists	25	28	
Proton-pump inhibitors	16	22	
Both	1	3	
Medication prescribed for >3 of the past 6 mo (%)	20	30	0.05
Eradication therapy prescribed during follow-up (no.)	3	9	0.08

*Ten patients were lost to follow-up: six patients in the group assigned to omeprazole and antibiotics and four patients in the group assigned to omeprazole alone.

†Total scores can range from 0 (no symptoms or limitations) to 20 (severe symptoms most days, requiring regular medication).

‡Quality of life was measured with the SF-36; scores can range from 0 (worst) to 800 (best).

با وجود اینکه این دو مقدار بهم مرتبط هستند، در این موارد (دقت آماری)، P value نسبت به بازهٔ اطمینان اطلاعات بیشتری را در اختیار می‌گذارد. همانطور که در فصلهای گذشته گفته شد هر چه P value کوچک‌تر باشد اطمینان از واقعی بودن اثر درمان بیشتر است.

اختلاف دو گروه مطالعه بر مبنای برطرف شدن نشانه‌ها با توجه به P value و بازهٔ اطمینان ذکر شده کاملاً معنادار است.(شماره ۱۱)

آیا این نتایج روا از نظر بالینی دارای اهمیت است؟

بعد از اینکه در مورد واقعی بودن اثر مداخله مطمئن شدید، باید بررسی کنید این اثر به اندازهٔ کافی بزرگ هست که از نظر بالینی نیز دارای اهمیت باشد. در هر کارآزمایی بالینی، برای مشخص شدن اثر مداخله روی دادن یک رخداد را به عنوان معیار در نظر می‌گیرند و از بررسی تغییرات بروز آن در گروههای مطالعه به عنوان مقیاسی برای ارزیابی اثر درمان استفاده می‌کنند. این رخداد به طور مستقیم یا غیرمستقیم با اثر مداخله در ارتباط است و اندازه‌گیری آن می‌تواند کمّی یا کیفی باشد. به میزان بروز این رخداد در گروه کنترل[1] CER و در گروه مداخله[2] EER گفته می‌شود. توجه داشته باشید که سنجه‌های فوق همانطور که قبلاً هم ذکر شد جزء سنجه‌های پیامد دوتایی هستند (بروز یا عدم بروز رخداد) بنابراین در مطالعاتی که از این پیامدها استفاده می‌کنند (مثلاً برطرف شدن علائم سوءهاضمه یا برطرف نشدن آن در مطالعهٔ مورد بررسی) می‌توان اهمیت بالینی یا اندازهٔ اثر را بر مبنای فراوانی رخداد مورد نظر در گروه مداخله و گروه کنترل در پایان دورهٔ پیگیری محاسبه کرد.

سنجه‌هایی که به طور شایع در این زمینه استفاده می‌شوند شامل کاهش خطر نسبی[3] (RRR) و کاهش خطر مطلق[4] (ARR) است. (در صورتیکه رخداد مورد نظر یک پیامد نامطلوب محسوب شود).

1- Control Event Rate
2- Experiment Event Rate
3- Relative Risk Reduction
4- Absolute Risk Reduction

نحوه محاسبه این سنجه‌ها در مطالعه این گونه است؟

RRR = (CER- EER) / CER

ARR = CER- EER

چنانچه در مطالعه، رخداد مورد نظر یک پیامد مطلوب باشد (مانند برطرف شدن علائم سوءهاضمه در مطالعهٔ مورد بررسی) از سنجه‌های معادل کاهش خطر یعنی افزایش فایدهٔ نسبی[1] (RBI) و افزایش فایدهٔ مطلق[2] (ABI) استفاده می‌شود که نحوهٔ محاسبهٔ آن مشابه RRR و ARR است.

سنجه‌هایی مانند RBI یا RRR که اندازه اثر را به صورت نسبی مشخص می‌کنند، ممکن است نتوانند بین یک اثر درمانی بزرگ و یک اثر درمانی بسیار کوچک و بی‌اهمیت تمایز قائل شود. این مشکل از آنجا منشأ می‌گیرد که به علت وارد کردن CER در مخرج کسر به نوعی اثر آن که نشاندهندهٔ میزان پایهٔ احتمال پیامد دو گروه کنترل است، از محاسبه حذف می‌شود مثال زیر گویای این مشکل است.

Table 5.3 Clinically useful measures of the effects of treatment

	Event rate = progression of disability by 33 months		Relative risk reduction (RRR = |CER – EER|/CER)	Absolute risk reduction (ARR = |CER – EER|)	Number needed to treat (NNT = 1/ARR)
	Control event rate (on placebo) (CER)	Experimental event rate (on interferon) (EER)			
In the actual trial Lancet 1998; 352: 1491–7	50%	39%	(50% – 39%)/50% = 22%	50% – 39% = 11%	1/11% = 9
In the hypothetical trivial case	0.00050%	0.00039%	(0.00050% – 0.00039%)/ 0.00050% = 22%!	0.00050% – 0.00039% = 0.00011%	1/0.00011% = 909 090

1- Relative Benefit increase
2- Absolute Benefit increase

در این جدول مشاهده می‌کنید که خط اول نتایج از یک مثال واقعی دربارهٔ اثربخشی اینترفرون در کاهش پیشرفت معلولیت در طی ۳۳ ماه‌پیگیری بیماران گرفته شده است ولی در خط انتهایی جدول یک سری نتایج فرضی برای نشان دادن ضعف RRR در انعکاس تفاوت اثرهای بزرگ واثرهای کوچک و بی‌اهمیت آورده شده است. براساس این نتایج فرضی چنانچه نسبت پیشرفت معلولیت در گروه کنترل (0.00050%) و گروه مداخله (0.00039%) بسیار کوچک و بی‌اهمیت باشد باز هم کاهش خطر نسبی به اندازهٔ ۲۲ درصد باقی می‌ماند. برخلاف آن سنجه‌های مطلق (قدر مطلق تفاوت CER و EER) به سادگی بین این دو مثال تمایز قایل می‌شوند.

در مثال کارآزمایی اینترفرون ARR برابر 50%-39%=11% است. در حالیکه در مثال فرضی که خطر پایهٔ آن بسیار کوچک و بی‌اهمیت است، ARR نیز مقدار کوچک و بی‌اهمیتی خواهد داشت (0.00011%) بنابراین ARR سنجهٔ معنادارتری نسبت به RRR است.

در سالهای اخیر برای کاربردی‌تر شدن نتایج بدست آمده از مطالعات، مفهوم جدیدی با عنوان تعداد لازم برای درمان[1] (NNT) شکل گرفته است. NNT از تقسیم عدد یک بر میزان کاهش خطر مطلق بدست می‌آید (NNT=1/ARR). ارائه نتایج مطالعات به صورت NNT مزایای متعددی دارد از جمله اینکه عدد NNT به صورت یک عدد کامل بیان می‌شود و بنابراین به خاطر سپردن آن آسانتر است (در صورت اعشاری بودن 1/ARR عدد بدست آمده به صورت اولین عدد کامل بعدی به عنوان NNT گزارش می‌شود).

از طرف دیگر مفهوم NNT نیز به کاربردی تر شدن آن کمک کرده است، NNT بیان می‌کند چه تعداد بیمار لازم است که با درمان آزمایشی به مدت تعیین شده در کارآزمایی درمان شوند تا از پیشامدن یک پیامد نامطلوب جلوگیری شود (1/ARR) یا یک پیامد مطلوب بوجود آید (1/ABI). در مثال کارآزمایی اینترفرون NNT برابر با ۹=٪۱۱÷۱ است و بدین معناست که باید ۹ نفر در طی ۳۳

1- Number Needed to Treat

ماه با اینترفرون درمان شوند تا از یک مورد پیشرفت معلولیت در مبتلایان به MS جلوگیری شود. در مثال فرضی در انتهای جدول ... اهمیت بالینی NNT مشخص‌تر می‌شود. برای چنین اثر درمانی کوچکی نیاز است تا تقریباً یک میلیون (۹۰۹۰۹۰) نفر به مدت ۳۳ ماه با اینترفرون درمان شوند تا از یک مورد پیشرفت معلولیت جلوگیری شود! (NNT=909090)

امروزه با استفاده از مفهوم NNT به راحتی و با سرعت بالا می‌توانید به میزان اثربخشی مداخلات پی ببرید و از نتایج مطالعات با سهولت بیشتری در تصمیم گیری بر بالین بیماران استفاده کنید.

حال به نظر شما چه اندازه از NNT از نظر بالینی با اهمیت است؟

برای یافتن پاسخ این سوال می‌توانید NNT گزارش شده در مطالعه را با NNTهای بدست آمده از سایر مداخلات درمانی و طول درمان آنها مقایسه کنید و از تجربه و تخصص بالینی خود نیز در این زمینه کمک بگیرید. بدیهی است که هر چه مقدار NNT کمتر باشد اثر مداخله بیشتر بوده و مطلوب‌تر است. حال به مطالعهٔ اصلی خود یعنی "اثر ریشه‌کنی عفونت هلیکوباکترپیلوری در بهبود نشانه‌های بیماران مبتلا به سوءهاضمه بدون زخم" بازگردیم. در جدول ۲ مطالعه، مشاهده کردید که برطرف شدن نشانه‌ها در گروه مداخله (آنتی‌بیوتیک و امپرازول) بعد از یک سال در ۲۱ درصد افراد دیده شد (EER=21%) که این رقم در گروه کنترل (فقط امپرازول) برابر ۷ درصد بود (CER=7%). اکنون با محاسبات زیر می‌توانید میزان اهمیت بالینی مطالعه را مشخص کنید.

$$RBI= \frac{21\% \div 7\%}{7\%} =2 \qquad ABI= 21\% - 7\% = 14\% \qquad NNT=1 \div 14\%= 8$$

بیان مفهوم NNT=8 به عهدهٔ شماست!...

در مجموع با توجه به ارزیابی قسمتهای مختلف مطالعه و رعایت معیارهای کیفیت، دقت آماری بالا و حجم اثر معنادار، استفاده از توصیهٔ این مطالعه در کنار ارزشها و ترجیحات بیمار برای تصمیم‌گیری در مورد وی منطقی به نظر می‌رسد.

References

1- How to Use the Evidence: assessment and application of scientific evidence. National Health and Medical Research Council; 2000.

2- Egger M, Smith GD, Altman DG, Systematic Reviews in Health Care: Meta-analysis in context. 2nd ed. BMJ Publishing Group 2001.

3- Guyatt G, Rennie D, Hayward R. User's Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice.American Medical Association; 2002.

4- Cochrane Reviewer's Handbook; The Cochrane Collaboration; 2004.

فصل ششم
شواهد مربوط به اتیولوژی و آسیب

در یک درمانگاه عمومی در حال ویزیت بیماری هستید که بخاطر سابقهٔ فشار خون بالا به مدت ۵ سال است که روزانه ۳۰mg نیفدیپین مصرف می‌کند. او در طی دوران بازنشستگی خود نحوه استفاده از اینترنت را فرا گرفته است و عنوان مطالبی را از چند روزنامه که در جستجوی اینترنتی یافته است، چاپ کرده و برای شما آورده است.

لس‌آنجلس تایمز[1]: داروهای ضد فشار خون قربانی می‌گیرد.

واشنگتن پست[2]: داروهای ضد فشار خون منجر به حملات قلبی و سرطان می‌شوند. پژوهشگران هشدار دادند این خطر شش میلیون نفر را تهدید می‌کند و ...

او بسیار نگران است و می‌خواهد بداند آیا لازم است نیفدیپین را قطع کند. قبلاً در مورد خطرات قلبی عروقی آدالات شنیده‌اید ولی در مورد سرطان‌زایی آن مطلبی به خاطرتان نمی‌رسد. سوالی که در ذهنتان شکل گرفته است اینگونه بیان می‌کنید: آیا بیمارانی که از بلوک کننده‌های کانال کلسیم[3] برای درمان فشار خون بالا استفاده می‌کنند، بدنبال مصرف آنها دچار سرطان می‌شوند؟

شروع به جستجو می‌کنید. در پایگاه داده‌ای BRC[4] مقاله‌ای تحت عنوان "بلوک کانال کلسیم و بروز سرطان در جمعیت سالمند"[5]، توجه شما را جلب می‌کند. متن کامل مقاله را از اینترنت دریافت می‌کنید (ضمیمه ۴). حال نوبت ارزیابی نقادانهٔ مطالعه بوسیله چک لیست روبرو است:

1- LA Times
2- Washington Post
3- Calcium-Channel Blockers
4- Biomedical Reference Collection Database
5- "Calcium-Channel Blockade and Incidence of Cancer in Aged Population" (Lancet 1996: 348: 443- 7)

آیا مطالعه از کیفیت مناسبی برخوردار است؟

- آیا گروههای تعریف شده‌ای از بیماران وجود دارند که از هر جهت به جز از نظر مواجهه با مداخلهٔ درمانی یا عامل مورد نظر مشابه باشند؟
- آیا روش اندازه‌گیری میزان مواجهه با مداخلهٔ درمانی یا عامل مورد نظر در دو گروه یکسان است؟
- آیا ارزیابی پیامدها نسبت به مواجهه به صورت عینی و کور انجام شده است؟
- آیا مدت پیگیری بیماران برای وقوع پیامد کافی بوده است؟
- آیا نتایج مطالعه شواهدی به نفع علیت دارد؟

آیا همراهی مواجهه و پیامد از نظر دقت آماری قابل قبول است؟

آیا این نتایج روا از نظر بالینی اهمیت دارند؟

- آیا ارتباط بین مواجهه و پیامد (تخمینی که از احتمال بروز پیامد در مطالعه بدست آمده است) به اندازه‌ای هست که از نظر بالینی اهمیت داشته باشد؟

شما به عنوان پزشک خیلی اوقات ناچارید در مورد اینکه آیا یک مداخلهٔ درمانی یا یک عامل محیطی به بیمارتان ضرر خواهد زد یا خیر، قضاوت کنید. سوالات زیادی در این زمینه هم برای شما و هم برای بیمارتان ایجاد می‌شود. سوالاتی مانند "آیا داروهای ضد بارداری خوراکی باعث سرطان پستان می‌شوند؟"، "آیا مصرف ویتامین آ در بارداری خطر نقص‌های مادرزادی را بالا می‌برد؟"، "آیا چاقی خطر ابتلا به سرطان پستان را بالا می‌برد؟" و ...

برای پاسخ به چنین سوالاتی باید بتوانید پس از یافتن شواهد مربوط، آنها از نظر روایی، دقت و اهمیت ارزیابی کنید. از فصلهای گذشته به خاطر دارید که برای پاسخ به سوالات بالینی از جمله سوالات مربوط به اتیولوژی و آسیب مقالات مروری نظام‌مند بالاترین سطح شواهد را تشکیل می‌دهد. با این حال متأسفانه در حال حاضر تعداد مقالات مروری نظام‌مند که به سوالات مربوط به اتیولوژی و عوارض پاسخ داده باشند محدود است.

کارآزمایی‌های تصادفی شده شاهددار (RCT) از نظر سطح شواهد بعد از مقالات مروری نظام‌مند قرار می‌گیرد ولی استفاده از آنها نیز در پاسخ به سوالات مربوط به اتیولوژی و عوارض مشکلات خاص خود را دارد. کارآزماییهای تصادفی شده به علت محدودیت‌هایی که از نظر حجم نمونه و طول مدت پیگیری دارند غالباً برای بررسی اتیولوژی، عوامل خطر و آسیب خصوصاً در مورد شرایط و بیماریهای نادر، چندان مناسب نیستند. از طرف دیگر استفاده از این نوع مطالعات برای یافتن آسیب‌ها و عوارض در بسیاری موارد غیر اخلاقی است. با این حال کارآزمایی‌های تصادفی شده شاهدداری که به قصد ارزیابی اثربخشی مداخلات درمانی انجام می‌شوند، چنانچه از دورهٔ پیگیری مناسبی برخوردار باشند می‌توانند عوارض جانبی مداخلات را نیز مشخص کنند.

در هر صورت برای یافتن آسیب و عوارض نیاز به شواهد دیگری داریم که محدودیت‌های RCT را نداشته باشد. قدم بعدی استفاده از مطالعات همگروهی است. از آنجاییکه در مطالعات همگروهی مواجهه در گروههای مطالعه تصادفی شده نیست، نسبت به کارآزمایی تصادفی شده سطح پایین‌تری از شواهد را به خود اختصاص می‌دهد.

در مواردیکه بیماری یا عارضهٔ مورد نظر نادر باشد یا به مدت طولانی برای وقوع نیاز داشته باشد. مطالعات همگروهی بزرگ نیز غیرقابل اجراست بنابراین باید به مطالعات گذشته‌نگر مانند مطالعات مورد–شاهدی روی آورد.

در مرحلهٔ آخر، ممکن است فقط یک یا چند مورد گزارش از آسیب بدنبال دریافت یک درمان خاص بیابید. درصورتیکه پیامد

غیرعادی و قابل توجه باشد (مانند فوکوملیا در کودکانی که مادر آنها از تالیدومید استفاده کرده است) مطالعات غیرتجربی مانند گزارش موردی و بررسی موردها نیز می‌توانند برای پاسخ به سوال شما کافی باشند. با این حال چون این مطالعات فاقد گروه کنترل هستند معمولاً صرفاً برای فرضیه‌سازی کفایت می‌کنند. در کل می‌توان گفت متأسفانه انجام پذیری مطالعات مربوط به اتیولوژی و عوارض با جایگاه طرح آنها از نظر سطح شواهد نسبت عکس دارد. پس از انتخاب مطالعه‌ای که حتی‌الامکان از بالاترین سطح شواهد برخوردار است اکنون می‌توانید سایر ابعاد مطالعه را نیز ارزیابی کنید.

در مطالعهٔ مورد بررسی طرح مطالعه از نوع همگروهی و آینده‌نگر بوده است.(شمارهٔ ۱)

Methods Between 1988 and 1992 we carried out a prospective cohort study of 5052 people aged 71 years or more and who lived in three regions of Massachusetts, Iowa, and Connecticut USA. Those taking calcium-channel blockers (n=451) were compared with all other participants (n=4601).

شماره ۱

آیا مطالعه از کیفیت مناسبی برخوردار است؟

- آیا گروههای تعریف شده‌ای از بیماران وجود دارند که از هر جهت به جز از نظر مواجهه با مداخلهٔ درمانی یا عامل مورد نظر مشابه باشند؟

یکی از پیش‌نیازهای رعایت این معیار، این است که افراد گروههای کنترل و مطالعه به طور تصادفی به این گروهها راه یافته باشند ولی همانطور که اشاره شد این کار معمولاً عملی نیست. در سایر مطالعات از جمله مطالعات همگروهی ممکن است افرادیکه

با عامل مورد نظر مواجهه داشته‌اند (گروه مطالعه) و افرادی که مواجهه نداشته‌اند (گروه کنترل) در شاخصهایی متفاوت باشند که نقش تعیین کننده در پیامد مطالعه دارد (متغیرهای مداخله‌گر). مثلاً چنانچه احتمال سرطان در افراد دارای فشار خون بالا از افراد با فشار خون طبیعی بیشتر باشد، فشارخون بالا می‌تواند به عنوان یک عامل مداخله‌گر عمل کند. بنابراین پژوهشگران باید مشخصات گروههای مطالعه را از نظر عوامل موثر بر پیش‌آگهی ثبت کنند و قابل مقایسه بودن گروهها را نشان بدهند و یا اثر مداخله‌گرهایی را که می‌شناسند به نوعی حذف کنند. از آنجاییکه از این طریق در بهترین شرایط فقط می‌توان اثر مداخله‌گرهای شناخته شده را حذف کرد، باید در تفسیر مطالعات همگروهی بیشتر احتیاط کنید.

در مطالعات مورد-شاهدی مشکل مداخله‌گرها باز هم بیشتر است چرا که امکان سنجش مداخله‌گرهای گذرا یا مداخله‌گرهایی که اثرات آنها به مرگ می‌انجامد وجود ندارد. مثلاً اگر گروه مورد از بیماران بستری انتخاب شود، در صورتیکه احتمال بستری در بیماران مواجه شده با عامل مورد بررسی بیش از بیماران مواجه نشده باشد، رابطه بین مواجهه و پیامد ممکن است به علت عامل مداخله‌گر بستری شدن توأم با تورش باشد.

نکتهٔ مهم دیگر این است که در مطالعات مورد–شاهدی بیماران گروه کنترل باید شانس برابر با گروه مورد برای مواجهه با عامل مورد بررسی داشته باشند. مثلاً چنانچه شما یک مطالعهٔ مورد–شاهدی در مورد بلوک کننده‌های کانال کلسیم و سرطان پیدا کردید که گروه مورد آن بیماران سرطانی هستند ولی مبتلایان به بیماری عروق کرونر (که شانس مواجهه با این داروها در آنها بیشتر است) از گروه کنترل حذف شده‌اند، دیگر نمی‌توانید راجع به درستی ارتباط مشاهده شده مطمئن باشید در مطالعه نشریهٔ لنست دو گروه از نظر شاخص‌هایی مانند فراوانی افراد دیابتی و افراد مبتلا به بیماریهای قلبی عروقی اختلاف معنادار دارند که از نقاط ضعف مطالعه محسوب می‌شود. (شمارهٔ ۲)

Results

Compared with participants who were not taking calcium-channel blockers those taking these drugs were significantly more likely to have cardiovascular diseasesand diabetes, to use cardiovascular drugs, to be disabled, to have a lower diastolic blood pressure, and to be admitted to hospital (table 1). Age, sex, ethnic origin, cigarette smoking, alcohol intake, systolic blood pressure, body-mass index, and use of non-cardiovascular drugs did not differ significantly between the two groups.

شمارهٔ ۲

- آیا روش اندازه‌گیری میزان مواجهه با مداخلهٔ درمانی یا عامل مورد نظر در دو گروه یکسان است؟

مطالعاتی که در آنها مواجهه با عامل مورد نظر در هر دو گروه به طور یکسان و با یک روش اندازه‌گیری شده‌است از روایی بیشتری برخوردار در مطالعهٔ مورد بررسی برای سنجش میزان مواجهه به جز مصاحبه با بیماران از آنها خواسته شده است تا تمام داروهای مصرفی خود را نشان بدهند. (شمارهٔ ۳)

During the section of the baseline interview on the use of medications, participants were asked to show all containers for all prescription and non-prescription drugs taken over the past 2 weeks. The interviewer transcribed the drug product name from the container label. If the label was not seen, the interviewer asked the participant the name of the drug. These ascertainment methods are similar to those of other studies and have proven valid and reliable.[25,26] Participants taking calcium-channel blockers were compared with participants who were not.

شمارهٔ ۳

■ آیا ارزیابی پیامدها به صورت عینی و کور نسبت به مواجهه انجام شده است؟

یکسان بودن سنجش پیامدهای بالینی در هر دو گروه اهمیت ویژه‌ای دارد. مثلاً اگر پژوهشگران برای اثبات فرضیهٔ خود، در بیمارانی که مورد شناخته شدهٔ مصرف بلوک کننده‌های کانال کلسیم هستند با دقت بیشتری به دنبال شواهدی دال بر ابتلا به سرطان باشند، ممکن است در نتایج مطالعه تورش ایجاد شود، همچنین بیمارانی که دچار عارضه شده‌اند احتمالاً انگیزه و توانایی بیشتری برای به یادآوردن مواجههٔ احتمالی دارند. برای حذف اثر این تورشها بهتر است که بیماران و کسانی که پیامدهای بیماران را مورد ارزیابی قرار می‌دهند نسبت به نوع گروه مطالعه ناآگاه باشند. از طرف دیگر عینی بودن معیارهای ارزیابی پیامد در هر دو گروه تا حدی از تداخل نظر پژوهشگران در نتیجهٔ مطالعه پیشگیری می‌کند.

در این مطالعه بروز سرطان از طریق بررسی تشخیص بیماران در هنگام ترخیص و علل مرگ آنها تعیین شده است ولی به کور بودن ارزیابی پیامدها نسبت به مواجهه اشاره‌ای نشده است. (شمارهٔ ۴)

The incidence of cancer was assessed by survey of hospital discharge diagnoses and causes of death. These outcomes were validated by the cancer registry in the one region where it was available.

شمارهٔ ۴

■ آیا مدت پیگیری بیماران برای وقوع پیامد کافی بوده است؟

مدت و کفایت پیگیری بیماران در شواهد مربوط به اتیولوژی و عوارض نیز مانند بقیه شواهد اهمیت ویژه‌ای دارد. به ویژه در مورد بیماریهایی مانند سرطان طول دورهٔ پیگیری کوتاه، تخمین کمتری از بروز واقعی پیامد بدست می‌دهد. به همین ترتیب چنانچه تعداد

افرادی که به هر دلیل از فرآیند پیگیری حذف شده‌اند، بیشتر از حد قابل قبول باشد، دیگر نمی‌توان به نتیجهٔ مطالعه اعتماد کرد.

در این مطالعه مدت پیگیری به طور متوسط ۳/۷ سال بوده است، که برای وقوع پیامدهای مورد نظر کافی است. (شمارهٔ ۵)

During 18774 person-years of follow-up (mean follow-up time 3.7 years), 420 participants had a cancer event (140 in East Boston, 185 in Iowa, and 95 in New Haven; overall rate 22.4 per 1000 person-years; table 2).

شمارهٔ ۵

- آیا نتایج مطالعه شواهدی به نفع علیّت دارد؟

معیارهای متفاوتی برای رابطهٔ علّی در منابع مختلف ذکر شده است. در اینجا برای سادگی کار، معیارهای رابطه علّی به صورت سه معیار اصلی و سه معیار فرعی بیان می‌شود. برای وجود رابطهٔ علّی سه معیار اصلی حتماً باید وجود داشته باشد.

معیارهای اصلی

- همراهی[1]: آیا پژوهش توانسته است بین متغیرهای مورد نظر (در اینجا مواجهه و پیامد) همراهی معناداری از نظر آماری نشان دهد؟ این همراهی معنادار باید بتواند شواهد قانع‌کننده‌ای از همراهی علّت و معلول بیش از آنچه که صرفاً به علت شانس قابل انتظار است، ارائه کند.
- تقدم زمانی[2]: آیا در پژوهش مقدم بودن علت بر معلول از نظر زمانی اثبات شده است؟
- تغییر در علت باید به تغییر در معلول بیانجامد؛ آیا پژوهش می‌تواند نشان دهد که تغییر یا تعدیل شدت یا فرکانس علت (در اینجا مواجهه) قادر است در شدت یا فرکانس معلول (پیامد مورد بررسی) تغییر ایجاد کند؟

1- Association
2- Prior association

معیارهای فرعی

- قدرت همراهی[1]: این معیار بیانگر قدرت همراهی مشاهد شده بین مواجهه با یک عامل خاص و بروز پیامد مورد نظر (بیماری یا آسیب) است که با سنجه‌هایی مانند خطر نسبی نشان داده می‌شود.
- سازگاری در همراهی[2]: سازگاری زمانی وجود دارد که پژوهشهای مختلف که در محیط‌های متفاوت و برروی افراد متفاوت انجام شده است، نتایج مشابهی داشته باشد.
- امکانپذیری زیست‌شناختی[3]: برای این معیار منطقی بودن رابطه علّی باید با اصول زیست‌شناختی و پاتوفیزیولوژی تأیید شود.

در این مطالعه با وجود رعایت معیارهای اصلی علّیت، در قسمت بحث مطالعه[4] به علّت ضعف طرح پژوهش نشان دادن رابطهٔ علّی به کارآزمایی‌های بالینی طولانی مدت موکول شده است. (شمارهٔ ۶)

Discussion

We found that after adjustment for confounding factors, older people taking calcium-channel blockers for any indication had a significant 70% excess hazard ratio of developing cancer during 3.7 years (mean) of follow-up compared with those not taking these drugs. This association was consistent in all three regions in this study, and for most types of cancer. The dose-response gradient was significant. The excess risk of cancer associated with use of calcium-channel blockers was even more evident when the SEER cancer registry information was used to define the events.

We have identified six potential limitations of the study. First, although several confounding factors were examined, observational data such as these cannot conclusively establish causal links. Whether calcium-channel blockers truly promote cancer needs to be confirmed in long-term clinical trials. Physical disabilities and the consequent incapacity of a number of participants in this study would probably preclude their attendance for follow-up visits. Thus, these individuals would probably be excluded from clinical trials.

شمارهٔ ۶

1- Strength of association
2- Consistency of association
3- Biological plausibility
4- Discussion

آیا همراهی مواجهه و پیامد از نظر دقت آماری قابل قبول است؟

در مطالعه مورد بررسی نسبت مخاطره[1] بدست آمده به عنوان سنجهٔ تفاوت دو گروه از نظر مواجهه برای ابتلا به سرطان از بازهٔ اطمینان قابل قبول و P value کاملاً معنادار برخوردار است. (شمارهٔ ۷)

Findings The hazard ratio for cancer associated with calcium-channel blockers (1549 person-years, 47 events) compared with those not taking calcium-channel blockers (17225 person-years, 373 events) was 1.72 (95% CI 1.27-2.34, p=0.0005), after adjustment for confounding factors.

شمارهٔ ۷

آیا این نتایج روا از نظر بالینی نیز اهمیت دارد؟

پس از اینکه از روایی و دقت آماری نتایج مطالعه اطمینان حاصل کردید اکنون نوبت ارزیابی اهمیت بالینی این نتایج در تصمیم‌گیری برای بیمارتان است.

- آیا ارتباط بین مواجهه و پیامد (تخمینی که از احتمال بروز پیامد در مطالعه بدست آمده است) به اندازه‌ای هست که از نظر بالینی اهمیت داشته باشد؟

تخمین قدرت ارتباط بین مواجهه و پیامد بسته به طرح مطالعه با روشهای مختلف صورت می‌گیرد. برای کارآزمایی‌های تصادفی شده و مطالعات همگروهی شایعترین روش، محاسبهٔ بروز رخداد نامطلوب در افراد مواجه شده با مداخلهٔ درمانی یا عامل

1- Hazard Ratio

مورد نظر (a/a+b) و افراد شاهد (c/c+d) است. میزان بروز رخداد نامطلوب در اصطلاح خطر[1] نامیده می‌شود و به نسبت آن در گروه مواجهه به گروه کنترل خطر نسبی[2] (RR) می‌گویند [a/(a+b)]/[c/(c+d)]. (جدول ۱)

جدول ۱

		پیامد نامطلوب		مجموع
		عدم وجود	وجود	
مواجهه با مداخلهٔ درمانی یا عامل خطرزا	بله	b	a	a + b
	خیر	d	c	c + d

برای مثال چنانچه در مطالعه‌ای از ۱۰۰۰ بیمار گروه درمان ۲۰ نفر پیامد نامطلوب (عارضهٔ درمان) را نشان بدهند (a=20, a+b=1000) و از ۱۰۰۰ بیمار گروه کنترل که آن مداخله درمانی را دریافت نمی‌کنند ۲ نفر دچار عارضه شوند (c+d=1000, c=2) ، خطر نسبی (RR) برابر ۱۰ خواهد بود (RR=[a/(a+b)]/[c/(c+d)]=0.02/0.002=10) خطر نسبی برابر ۱۰ بدین معناست که بیمارانی که مداخلهٔ درمانی مورد نظر را دریافت می‌کنند نسبت به بیمارانی که این مداخله را دریافت نمی‌کنند ۱۰ برابر بیشتر در خطر نشان دادن پیامد نامطلوب (عارضه درمان) هستند.

در مطالعات مورد ـ شاهدی از آنجاییکه بیماران در گروه مورد و افرادیکه بیماری مورد نظر را ندارند در گروه شاهد قرار می‌گیرند، محاسبهٔ بروز پیامد نامطلوب در هر یک از گروهها امکان‌پذیر نیست. بنابراین باید از سنجهٔ دیگری به جز خطر نسبی

1- Risk
2- Relative Risk

استفاده کرد. در این نوع مطالعات قدرت ارتباط بین مواجهه با پیامد مورد نظر به طور غیرمستقیم و با استفاده از نسبت شانس[1] (OR) تخمین زده می‌شود.

(نحوه محاسبهٔ شانس و تفاوت آن با خطر که از جنس احتمال است در فصل ۳ بیان شده است)

شانس در گروه موردها برابر است با نسبت تعداد افرادیکه در سابقهٔ آنها مواجهه با عامل مورد نظر وجود دارد به تعداد افرادیکه با این عامل مواجهه نداشته‌اند (a/c) و در گروه شاهد نیز به همین ترتیب (b/d) بنابراین نسبت شانس (OR) اینگونه محاسبه می‌شود. (جدول ۲)

جدول ۲

		مواجهه با درمان یا عامل خطرزا	
		خیر	بله
پیامد نامطلوب	وجود (گروه مورد)	c	a
	عدم وجود (گروه شاهد)	d	b

$$OR = (a/c)/(b/a) = ad / bc$$

تفسیر نسبت شانس در مطالعات مورد ـ شاهدی مانند خطر نسبی در مطالعات همگروهی است. حال به نظر شما چه اندازه‌ای از نسبت شانس یا خطر نسبی از نظر بالینی اهمیت دارد. تفسیر اهمیت بالینی این سنجه‌ها وابسته به روایی محاسبهٔ آنها در مطالعه است. نسبت شانس و خطر نسبی بیشتر از ۱ نشاندهندهٔ همراهی مواجهه با افزایش خطر پیامد نامطلوب است.

به عنوان یک توصیهٔ کلی در مورد پیامدهای نامطلوب خفیف معمولاً OR کمتر از ۴ ارزش چندانی ندارد ولی اگر این پیامد

3- Odds ratio (relative odds)

نامطلوب خطیر یا کشنده باشند مقادیر کمتر از ۴ نیز ارزشمند تلقی می‌شود. از آنجاییکه احتمال تورش در مطالعات همگروهی کمتر است خطر نسبی بیش از ۳ برای مضر بودن مداخله یا مواجهه با عامل مورد نظر متقاعد کننده است.

با اینکه OR و RR برای نشان دادن قدرت ارتباط مواجهه و پیامد به کار می‌روند ولی مانند شواهد مربوط به درمان نیازمند سنجه‌ای هستیم که مفهوم ساده‌تر و کاربرد بیشتری در طبابت داشته باشد. بدین منظور از سنجهٔ تعداد لازم برای آسیب (NNH) استفاده می‌شود. NNH از نظر مفهومی بیان می‌کند چه تعداد بیمار باید در معرض یک عامل فرضی قرار بگیرند تا یک رخداد زیانبار ایجاد شود. محاسبهٔ NNH در کارآزماییهای تصادفی شده و مطالعات همگروهی مشابه محاسبهٔ NNT در شواهد مربوط به درمان است (NNH=1/[a/(a+b)]-[c/(c+d)]).

محاسبهٔ NNH برای مطالعات مورد شاهدی قدری پیچیده‌تر است، زیرا همانطور که گفته شد میزان بروز پیامد در این نوع مطالعه به طور مستقیم قابل محاسبه نیست. برای راحتی محاسبهٔ NNH در چنین شرایطی از جدولهایی مانند جدول ۳ استفاده می‌شود. (جدول ۳)

Deriving the NNH From the Odds Ratio*

Control Event Rate	Therapeutic Intervention (OR)								
	1.1	1.2	1.3	1.4	1.5	2	2.5	3	3.5
0.05	212	106	71	54	43	22	15	12	9
0.1	112	57	38	29	23	12	9	7	6
0.2	64	33	22	17	14	8	5	4	4
0.3	49	25	17	13	11	6	5	4	3
0.4	43	23	16	12	10	6	4	4	3
0.5	42	22	15	12	10	6	5	4	4
0.7	51	27	19	15	13	8	7	6	5
0.9	121	66	47	38	32	21	17	16	14

جدول ۳

(CERمیزان بروز رخداد (پیامد نامطلوب) در بیمارانی است که با عامل مورد نظر مواجهه نداشته‌اند)

در مطالعهٔ مورد بررسی که از نوع همگروهی است، اندازه اثر مشاهده شده را با توجه به یافته‌های مطالعه می‌توان اینگونه محاسبه کرد.

a/a+b= 47÷1549=3.03%

c/c+d= 373÷17225=2.17%

RR= 3.03% ÷2.17%=1.4 (p=0.032) (CI 95%:1.27-2.34)

NNH= 1÷ (3.03%-2.17%)=116

به طور کلی با توجه به چک لیست ارزیابی، نتایج مطالعه مورد بررسی از روایی و دقت آماری قابل قبولی برخوردار است ولی قدرت ارتباط (NNH=116, RR=1.4) نشان داده شده چندان قابل توجه نمی‌باشد. بنابراین نتایج این مطالعه شواهد کافی را برای توصیه به قطع داروهای بلوک کنندهٔ کانال کلسیم فراهم نکرده است.

References

1- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine, How to Practice and Teach EBM, 2nd ed. Churchill Livingstone 2000.

2- Guyatt G, Rennie D, Hayward R. User's Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice.American Medical Association; 2002.

3- Riegelman RK. Studying a Study and Testing a Test. 4th ed. Lippincott Williams and Wilkins; 2000.

فصل هفتم
شواهد مربوط به پیش‌آگهی

بیماری به مطب شما مراجعه می‌نماید. او آقای ۷۰ ساله‌ای است که سه ماه قبل با تشخیص سکتهٔ مغزی ایسکمیک[1] از بیمارستان مرخص شده است. در حال حاضر ریتم قلب وی به صورت سینوسی و نرمال است. فشار خون او ۱۲۰/۸۰mmHg است و در سمع ریه و سایر معاینات و بررسی‌ها هیچ نکتهٔ غیرطبیعی به استثناء ضعف واضح در نیمهٔ چپ بدن وی یافت نمی‌شود.

بیمار به تازگی بعد از مطالعه مقاله‌ای درباره شانس تشنج بعد از سکتهٔ مغزی در وب سایت BMJ دچار نگرانی و اضطراب در مورد وقوع چنین اتفاقی برای خود شده است و از شما در این رابطه سوال می‌نماید. وی در مورد سیر بیماری خود جویای اطلاعات است. اجزای این سوال بالینی را در ذهن خود مرتب می‌کنید:

در یک آقای ۷۰ ساله که در حال حاضر مشکلی ندارد، آیا سابقهٔ سکته مغزی خطر تشنج را افزایش می‌دهد؟

شروع به جستجو می‌کنید؛ در پایگاه داده‌ای پاب‌مد مقاله‌ای تحت عنوان "حملات صرع بعد از اولین سکتهٔ مغزی"[2] می‌یابید که به نظر می‌رسد دقیقاً در راستای پاسخ به سوال شماست (متن کامل مقاله در ضمیمهٔ شمارهٔ ۷ آمده است).

اکنون می‌توانید مطالعه را با چک لیست زیر مورد ارزیابی قرار دهید.

1- Ischemic Stroke

2- Epileptic seizures after a first stroke: The Oxfordshire Community Stroke Project BMJ 1997; 315: 1582-1587

- **آیا مطالعه از کیفیت مناسبی برخوردار است؟**
  - آیا شرکت‌کنندگان در مطالعه، نمونهٔ مناسب و تعریف شده‌ای از بیماران هستند؟
  - آیا بیماران در زمان مناسب و مشابهی از دورهٔ بیماری وارد مطالعه شده‌اند؟
  - آیا پیگیری بیماران کامل و به مدّت کافی بوده است؟
  - آیا در مطالعه از پیامدهای عینی و مناسب استفاده شده است؟
  - آیا سنجش پیامدها در مطالعه به صورت کور بوده است؟
  - اگر در مطالعه پیش‌آگهی برای بعضی زیرگروهها گزارش شده، از چه روشی برای تصحیح سایر عوامل موثر در پیش‌آگهی بین این زیرگروهها استفاده شده است؟
- **آیا تخمین مطالعه از پیش‌آگهی از دقت آماری کافی برخوردار است؟**
- **آیا نتایج مطالعه از نظر بالینی دارای اهمیت است؟**
  - پیامدهای مورد نظر با چه احتمالی در طول زمان ظاهر می‌شود؟

مطمئناً شما نیز به عنوان یک پزشک با سوالات بی‌پایان بیماران در مورد سیر بیماریشان مواجه شده‌اید. امروزه بسیاری از بیماران خصوصاً مبتلایان به بیماریهای مزمن زمان زیادی را صرف جمع‌آوری اطلاعات در مورد پیش‌آگهی بیماری خود می‌کنند. این اطلاعات معمولاً از منابع مختلفی مانند تجربهٔ دوستان، رسانه‌های گروهی، اینترنت و ... بدست می‌آید که در بسیاری مواقع غیرقابل اعتماد است. چنین بیمارانی در نهایت برای رد و یا تأیید اطلاعات خود به

شما مراجعه می‌کنند. سوالات مربوط به پیش‌آگهی ممکن است به صورت مستقیم توسط بیمار مطرح شود (همانند سناریوی ذکر شده در ابتدای فصل) و یا اینکه ذهن شما را به طور غیرمستقیم برای ادامه یا انتخاب بهترین روش درمانی به خود مشغول سازد؛ در هر صورت دانستن پیش‌آگهی بیماری نه تنها بیمار را برای پذیرفتن واقعیت‌ها آماده ساخته و در افزایش اعتماد آنها نسبت به شما نقش بسزایی دارد، بلکه در بسیاری موارد به عنوان ابزاری مفید برای انتخاب راهبرد تشخیصی و درمانی مناسب مورد استفاده قرار می‌گیرد.

سوالات مربوط به پیش‌آگهی از سه جزء تشکیل می‌شود:

- وقوع چه پیامدهایی مورد انتظار است؟ این سوال به جنبهٔ کیفی[1] پیش‌آگهی اشاره دارد. مثلاً حملهٔ صرع یکی از پیامدهای مورد انتظار بعد از سکته‌های مغزی است.
- این پیامدها با چه احتمالی به وقوع می‌پیوندد؟ این سوال بیانگر جنبهٔ کمّی[2] پیش‌آگهی است. مثلاً احتمال وقوع حملهٔ صرع بعد از سکتهٔ مغزی چقدر است؟
- در چه بازهٔ زمانی باید منتظر وقوع این پیامدها بود؟ این سوال نشان‌دهندهٔ جنبهٔ زمانی[3] پیش‌آگهی است. مثلاً احتمال وقوع حملهٔ صرع در سال بعد از سکتهٔ مغزی چقدر است؟

برای پاسخ به چنین سوالاتی پس از یافتن شواهد مربوط باید از مهارت کافی برای ارزیابی روایی، دقت و اهمیت بالینی آنها بر خوردار باشید.

سطح شواهد

در سوالات مربوط به پیش‌آگهی نیز دستیابی به بالاترین سطح شواهد از طریق مقالاتی که به صورت مرور نظام‌مند کلیهٔ شواهد مربوط و روا طراحی شده‌اند، امکانپذیر است ولی احتمال یافتن چنین شواهدی در پاسخ به سوالات مربوط به پیش‌آگهی در حال

1- Qualitative aspect
2- Quantitative aspect
3- Temporal aspect

حاضر نسبتاً کم است. از دیگر طرحهای مناسب برای بررسی پیش‌آگهی می‌توان به مطالعات همگروهی و مورد ـ شاهدی اشاره نمود. در این بین نوع خاصی از مطالعات همگروهی تحت عنوان مطالعهٔ همگروهی آغازین[1] که در آن بیماران از مراحل اوّلیهٔ پس از تشخیص بیماری، مورد مطالعه قرار می‌گیرند از دیگر طرحها مناسب‌تر است. چرا که در این نوع مطالعات عدم توجه به دورهٔ خاصی از سیر بیماری باعث ایجاد تورش در نتایج نخواهد شد. در مقالهٔ BMJ طرح مطالعه از نوع همگروهی بوده است که دورهٔ پیگیری آن بلافاصله بعد از اوّلین حملهٔ سکتهٔ مغزی آغاز شده است.(شماره ۱)

Design: Cohort study following up stroke survivors for 2 to 6.5 years; comparison with age specific incidence rates of epileptic seizures in the general population.

شمارهٔ ۱

پس از ارزیابی سطح مطالعه، اکنون نوبت ارزیابی سایر ابعاد آن است.

آیا مطالعه از کیفیت مناسبی برخوردار است؟

برای ارزیابی کیفیت شواهد مربوط به پیش‌آگهی معیارهایی که به طور گسترده مورد توافق باشد وجود ندارد. به علاوه شواهد تجربی که نشان دهد ویژگیهای خاصی از مطالعه، روایی و پایایی یافته‌های آن را تحت تاثیر قرار می‌دهد، بسیار محدود است با این حال بر پایهٔ ملاحظات نظری، بسیاری از جنبه‌های روش شناختی در ارزیابی کیفیت شواهد با اهمیت به نظر می‌رسد. همانطور که پیش از این نیز اشاره شد، در ارزیابی شواهد شبه تجربی مانند اکثر مطالعات مربوط به پیش‌آگهی بررسی معیارهای کیفیت

1- Inception Cohort

بیشتر بر پایهٔ قضاوت فرد ارزیابی کننده است.

- آیا شرکت‌کنندگان در مطالعه نمونه مناسب و تعریف شده‌ای از بیماران هستند؟ آیا این بیماران در زمان مناسب و مشابهی از دورهٔ بیماری وارد مطالعه شده‌اند؟

مطالعات مربوط به پیش‌آگهی به طور ایده‌آل باید تمام افرادی را که تاکنون به بیماری مورد نظر مبتلا شده‌اند از لحظهٔ شروع بیماری مورد بررسی قرار داده باشند. ولی از آنجاییکه رسیدن به این ایده‌آل به جز در موارد خاص (مانند بیماریهای نادر و نوظهور) غیرممکن است، می‌توان میزان دستیابی به آن را در رویکرد مطالعه ارزیابی نمود. در این راستا علاوه بر وجود تعریف مشخصی از بیماری، نمونهٔ مطالعه باید شامل طیفی از بیماری باشد که ممکن است در زمان طبابت با آن برخورد کنیم. به عنوان مثال مطالعه‌ای را در نظر بگیرید که احتمال عود حملات تشنج در کودکانی را که یک بار بدنبال تب دچار تشنج شده‌اند بررسی می‌کند. این احتمال در بین کودکانی که اختلالات نرولوژیک همراه دارند بین ۹/۷۶-۶/۲ درصد و در کودکان فاقد هر گونه اختلال همراه ۶/۴-۵/۱ درصد گزارش شده است، بنابراین انتخاب نمونهٔ نامناسب می‌تواند نتایج را شدیداً تغییر دهد.

از این مطلب می‌توان نتیجه گرفت که ذکر نحوهٔ نمونه‌گیری و توصیف کافی از معیارهای تشخیصی برای قضاوت دربارهٔ این معیار اهمیت زیادی دارد.

در مورد زمان شروع پیگیری (در دورهٔ بیماری)، بهتر است تمام افرادی که وارد مطالعه شده‌اند در مراحل ابتدایی بیماری و تقریباً مشابه هم باشند. به عبارت دیگر شروع پیگیری به محض قابل قضاوت شدن تظاهرات بیماری همانگونه که در مطالعات همگروهی آغازین دیده می‌شود. بهترین رویکرد است. اگر پژوهشگر مدتی بعد از شروع بیماری اقدام به پیگیری افراد کند ممکن است پیامدهای مربوط به افرادی که بیماری آنها به سرعت بهبود یافته و یا به سرعت باعث مرگ آنها شده است، در نتایج مطالعه

لحاظ نشود و نتایج مطالعه همراه با تورش باشد. استثنائاً در مواردی که بدنبال تعیین پیش‌آگهی بیماران در مراحل انتهایی بیماری هستیم، نمونهٔ مطالعه باید از بیمارانی که به طور مشابه در مرحلهٔ پیشرفتهٔ بیماری هستند تشکیل شده باشد. در این مطالعه، نمونه شامل بیمارانی است که بعد از اوّلین سکتهٔ مغزی وارد شده‌اند. جزئیات بیشتری راجه به نحوهٔ نمونه‌گیری و فراخوانی بیماران ذکر نشده است. (شمارهٔ ۲)

Method

The Oxfordshire community stroke project registered 675 patients with a first stroke, 357 (53%) women and 318 (47%) men, from a study population of about 105 000 over 4 years. Patients of all ages were included: 163 (24%) were under 65 years, 195 (29%) 65-74 years, 228 (34%) 75-84 years, and 89 (13%) over 84 years (mean age 72.2 years).

شمارهٔ ۲

- آیا پیگیری بیماران کامل و برای مدت کافی بوده است؟

اگر کلیهٔ بیمارانی که وارد مطالعهٔ همگروهی آغازین شده‌اند تا زمان بهبودی کامل و یا بروز یک پیامد خاص که اغلب مرگ بیمار است پیگیری شوند می‌توان از کافی بودن زمان پیگیری کاملاً مطمئن بود. در غیراینصورت قضاوت در مورد کافی بودن طول دورهٔ پیگیری بسته به نوع بیماری و سیر طبیعی آن بر پایهٔ دانش قبلی فرد ارزیابی کننده و مطالعات دیگر خواهد بود.

مثلاً اگر در مطالعه‌ای به خاطر کوتاه بودن طول دورهٔ پیگیری تعداد اندکی از بیماران پیامدهای نامطلوب را نشان دهند نمی‌توان از خوب بودن پیش‌آگهی بیماری اطمینان حاصل کرد ولی با نتایج مشابه اگر طول دورهٔ پیگیری کافی باشد با اطمینان بیشتری

می‌توانید راجع به خوب بودن پیش‌آگهی قضاوت کنید.

با این حال در بسیاری موارد پیگیری تا زمان بهبود کامل یا مرگ بیمار غیرممکن است چرا که در چنین شرایطی به قیمت افزایش طول مدّت پیگیری احتمال از دست دادن افراد مورد مطالعه به علل مختلف از جمله عدم مراجعه برای پیگیری بیشتر خواهد شد. از آنجاییکه در اکثر مواقع علّت ریزش افراد شرکت‌کننده مشخص نمی‌شود و یا حداقل در گزارش مقاله ذکر نمی‌گردد، نتایج چنین مطالعاتی همراه با تورش خواهد بود. به طور توافقی معمولاً خروج بیش از ۲۰ درصد از شرکت‌کنندگان در مطالعه از روند پیگیری غیرقابل قبول و خروج کمتر از ۵ درصد قابل قبول در نظر گرفته می‌شود. یکی از روش‌های ساده و کاربردی برای تعیین میزان ریزش قابل قبول در نظر گرفتن شرایط در بدبینانه‌ترین حالت ممکن است. در صورتیکه این حالت تفاوت زیادی در نتایج مطالعه ایجاد کند، نمی‌توان از چنین ریزشی در شرکت‌کنندگان صرف نظر کرد. به عنوان مثال مطالعه‌ای را در نظر بگیرید که در طی دورهٔ پیگیری از ۱۰۰ نفر افراد مورد مطالعه ۴ نفر مرده‌اند و ۱۶ نفر نیز به دلایل، نامعلوم از لیست افراد تحت پیگری خارج شده‌اند.در این مطالعه خروج این ۱۶ نفر نادیده گرفته شده و میزان مرگ و میر برابر ۴/۸ درصد (۴/۸۴) گزارش می‌شود. ولی از آنجاییکه از وضعیت این ۱۶نفر اطلاعی در دست نیست می‌توانید با دو رویکرد خوش بینانه و بدبینانه نیز نتایج را بررسی کنید، بدین صورت که یک بار تمام ۱۶ نفر را در گروه افراد زنده قرار می‌دهید (خوش بینانه) و دفعهٔ بعد با فرض مرگ همهٔ آنها (بدبینانه) میزان مرگ و میر را محاسبه می‌کنید. این میزان در رویکرد خوش‌بینانه و بدبینانه به ترتیب ۴ درصد (۴/۱۰۰) و ۲۰ درصد (۲۰/۱۰۰) خواهد بود. بنابراین به علت تفاوت زیاد بین میزان مرگ و میر در حالت بدبینانه و نتیجهٔ گزارش شده در مطالعه این میزان ریزش غیرقابل صرف نظر کردن است. در مطالعهٔ مورد بررسی طول دورهٔ

پیگیری بین ۲ تا ۶/۵ سال ذکر شده است که برای بروز پیامدها کافی به نظر می‌رسد و پیگیری نیز کامل بوده است. (شماره ۳)

Follow up

The Oxfordshire community stroke project followed up every patient for 2 years or until death. Patients recruited earlier in the study were followed for longer, and no patient was lost to follow up. Seventeen patients died on day 1, 129 by 30 days, 208 by 1 year, and 254 by 2 years. A total of 421 patients were followed up at 2 years, 274 were seen at 3 years, 182 at 4 years, and 92 at 5 years; 324 patients survived to a final examination by a study neurologist between August 1987 and November 1988.

شمارهٔ ۳

- آیا در مطالعه از پیامدهای عینی و مناسب استفاده شده است؟

بیماریهای مختلف پیامدهایی متعددی برای بیماران به همراه دارد. سنجش بعضی از این پیامدها آسان و بعضی دیگر دشوار و پیچیده است. به طور کلی سنجش پیامدهایی که در دو سوی طیف بیماری قرار می‌گیرند یعنی مرگ و بهبود کامل نسبتاً ساده است و با روایی بالاتری انجام می‌شود، (به شرطی که علّت مرگ بدرستی تعیین شده باشد) ولی در این طیف پیامدهای متعدد و متنوعی قرار دارد که سنجش آنها با مشکلات بیشتری همراه است. بسیاری اوقات سنجش این پیامدها (مانند آمادگی برای بازگشت به سر کار، شدّت درد باقیمانده و ...) بر مبنای قضاوت پژوهشگران انجام می‌گیرد. برای کاهش تورش در سنجش پیامدها، پژوهشگران

باید ضمن انتخاب پیامد مناسب (رجوع کنید به فصل دوم)، معیارهای مشخصی برای تعریف هر یک از پیامدهای مورد نظر ارائه دهند و از آنها در طول دورهٔ پیگیری استفاده کنند.

حتماً به خاطر دارید که معیارهای عینی برای سنجش پیامدها نسبت به معیارهای ذهنی ارجحیت دارد و در کاهش تورش سنجش موثر است. در این مطالعه از بیماران در طی دورهٔ پیگیری دربارهٔ تشنج سؤال شد. اگر پاسخ بیماران مثبت بود، این افراد به نرولوژیست ارجاع می‌شدند. (شماره ۴)

At each follow up the nurse asked the patient and caregivers specifically about whether any possible seizure had occurred. As a check the nurse also reviewed the records kept by the primary healthcare team. Patients with suspected seizures were reassessed clinically by the study neurologist. Long term survivors were visited at the close of the study by one of the study neurologists (JPSB), and all available information about possible seizures was reviewed again with the patient. The hospital and primary care records of patients who died before this final visit were reviewed by one of the study neurologists (JPSB), and data relating to possible seizures were discussed at regular consensus meetings of the study team. Generalised seizures were diagnosed with reference to statements of witnesses. Focal seizures were distinguished clinically from clonus, spasms, and recurrent strokes.[8 9] Electroencephalograms were not routinely performed. Each visit included an assessment of functional ability using the Oxford handicap scale[10]; in this report patients in grades 0-2 were described as independent and patients in grades 3-5 as dependent.

شمارهٔ ۴

■ آیا سنجش پیامدها در مطالعه به صورت کور بوده است؟

گاهی اوقات با وجود استفاده از معیارهای عینی، احتمال وجود تورش در سنجش پیامدها وجود دارد. این حالت در صورتی اتفاق می‌افتد که پژوهشگرانی که مسئول سنجش پیامدها هستند، از شرایط بالینی بیمار به نحوی مطلع باشند. بنابراین بهتر است مسئول سنجش پیامدها از شرایط بالینی و عوامل تعیین کنندهٔ پیش‌آگهی بیماری بی‌خبر باشد. در این مطالعه توضیحی راجع به کور بودن نرولوژیست داده نشده است.

اگر در مطالعه پیش‌آگهی خاصی برای زیرگروهها تعیین شده، از چه روشی برای تصحیح سایر عوامل موثر بر پیش‌آگهی بین این گروهها استفاده شده است؟

در بسیاری از شرایط بالینی به دنبال تفاوتهای موجود در پیش‌آگهی گروههای خاص مبتلا به یک بیماری هستیم. مثلاً می‌خواهیم بدانیم در بین بیماران مبتلا به فیبریلاسیون دهلیزی غیر دریچه‌ای[1] آیا در افرادی که دهلیز چپ بزرگ شده دارند، نسبت به بیمارانی که دهلیز چپ در آنها نرمال است خطر ابتلا به سکتهٔ مغزی بیشتر است؟

مطالعات بسیاری وجود دارند که در آنها ادعا شده است پیش‌آگهی یک گروه از بیماران با گروههای دیگر متفاوت است.

نکته‌ای که باید در ارزیابی چنین مطالعاتی مورد توجه قرار گیرد، تصحیح سایر عوامل موثر در پیش‌آگهی است. بدین معنا که پژوهشگران باید اطمینان حاصل کنند، پیش‌آگهی پیش‌بینی شده برای زیرگروهها به علّت وجود همزمان یک عامل مهم دیگر خدشه‌دار نشده است. (مثلاً اگر بیمارانی که دهلیز چپ بزرگ شده دارند نسبت به بیمارانی که دهلیز چپ نرمال دارند سابقهٔ سکته‌های مغزی آمبولیک بیشتری نیز داشته باشند، نسبت دادن این پیش‌آگهی به دهلیز چپ بزرگ شده توام با تورش خواهد بود). بسیاری از چک لیست‌ها مساله درمانهای ارائه شده را در نظر نمی‌گیرند، اگر

1- Non-valvular atrial fibrilation

درمانهای ارائه شده برپایهٔ عوامل موثر در پیش‌آگهی متفاوت باشد، باز هم مطالعه نمی‌تواند ارزیابی قابل قبولی از توانایی تعیین پیش‌آگهی توسط یک عامل خاص داشته باشد.

شواهد نشان می‌دهد وجود چنین تفاوتهایی در درمان بسیار شایع است. بنابراین بهتر است کلیهٔ افراد شرکت‌کننده در مطالعه به طریق مشابه درمان شوند.

عوامل موثر در پیش‌آگهی به چند دستهٔ کلی تقسیم می‌شود.

- عوامل جمعیت شناختی[1] مانند سن، جنس، شرایط اجتماعی، اقتصادی، نژاد و ...
- عوامل مربوط به بیماری[2] مانند میزان پیشرفت بیماری ، شدّت بیماری، میزان غیرنرمال بودن آزمایش‌ها و ...
- بیماریهای همراه[3] وجود یا عدم وجود بیماریها و سایر شرایط بالینی همراه.

روشهای متعددی برای از بین بردن اثر این عوامل موثر بر پیش‌آگهی وجود دارد که طیفی از روشهای ساده مانند تحلیل طبقه‌ای[4] (پیش‌آگهی بیماران دارای دهلیز چپ بزرگ شده را جداگانه برای بیمارانی که سابقهٔ سکتهٔ مغزی امبولیک دارند و بیمارانی که این سابقه را ندارند تعیین می‌کند.) تا روشهای پیچیده‌تر مانند تحلیل رگرسیون چند گانه[5](که همزمان می‌تواند اثر چندین عامل را لحاظ کند) را شامل می‌شود، استفاده از این رویکردها معمولاً در گزارش مقاله در قسمت روشها ذکر می‌شود.

در صورتیکه تفاوت در پیش‌آگهی یک زیرگروه در مطالعه توسط مطالعات دیگر نیز تایید شده باشد با احتمال بیشتری می‌توانید به نتایج آن اطمینان کنید. در مطالعهٔ مورد بررسی برای زیرگروههای مبتلا به انواع مختلف سکتهٔ مغزی

1- Demographic
2- Disease-specific
3- Comorbidity
4- Stratified Analysis
5- Multiple regression Analysis

پیش‌آگهی‌های متفاوتی از نظر ابتلا به تشنج ذکر شده است. (شمارهٔ ۵)

The risk of seizures after ischaemic stroke was substantial only in patients presenting with severe strokes due to total anterior circulation infarction.

شمارهٔ ۵

آیا تخمین مطالعه از پیش‌آگهی از دقت آماری کافی برخوردار است؟

از آنجاییکه مطالعات مربوط به پیش‌آگهی نیز مانند سایر مطالعات معمولاً بر روی نمونه‌ای از جمعیت انجام می‌شود ممکن است مطالعه صرفاً به علّت پدیده شانس نتایج غیرواقعی بدست دهد. قضاوت در مورد دقت آماری مطالعه از طریق بررسی P value و بازهٔ اطمینان ۹۵ درصد انجام می‌شود و همانطور که در فصلهای گذشته گفته شد هر چه P value محاسبه شده کوچکتر باشد (کمتر از ۰/۰۵) و طول بازهٔ اطمینان کوتاهتر باشد، دقت آماری بیشترخواهد بود.(شمارهٔ ۶)

شمارهٔ ۶

Results:

The 5 year actuarial risk of a post stroke seizure in survivors (excluding 19 patients with a history of epilepsy and 3 patients in whom the seizure occurred shortly before death from another cause) was 11.5% (95% confidence interval 4.8% to 18.2%).

آیا نتایج مطالعه از نظر بالینی دارای اهمیت است ؟

اگر با در نظر گرفتن معیارهای فوق مطالعه از کیفیت و دقت آماری قابل قبول برخوردار است. اکنون می‌توانید اهمیت بالینی نتایج آن را ارزیابی کنید.

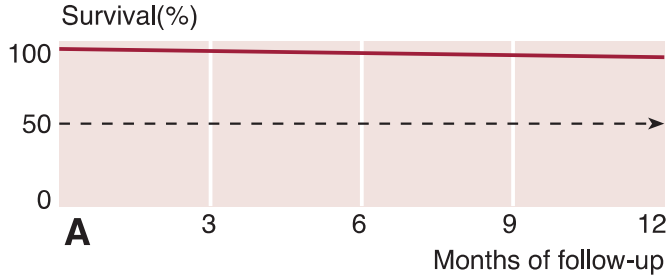
پیامدهای مورد نظر با چه احتمالی در طول زمان ظاهر می‌شود؟

نتایج حاصل از مطالعه معمولاً به یکی از سه شکل زیر در مطالعات ارائه می‌شوند با این حال از آنجا که هریک از این روشها اطلاعات مهمی را در اختیار خوانندگان قرار می‌دهد، درک کامل پیش‌آگهی بیماری معمولاً با در اختیار داشتن اطلاعات بدست آمده از هر سه روش امکانپذیر است.

۱ میزان بقای بیماران در یک زمان مشخص از دورهٔ پیگیری ابتلا میزان بقای بیماران در پایان یکسال، در ۵ سال اول و ...)

۲ میانهٔ بقا (مدت زمانی از دورهٔ پیگیری که تا آن زمان ۵۰ درصد از بیماران مرده‌اند)

۳ منحنی بقا[1] که در هر نقطه از زمان پیگیری نسبتی از نمونهٔ مطالعه که هنوز پیامد مورد نظر را نشان نداده‌اند به صورت درصد مشخص می‌کند. در ادامه۴ نمونه از منحنی بقا آورده شده است که هر یک پیش‌آگهی متفاوتی را نشان می‌دهد. شکل A

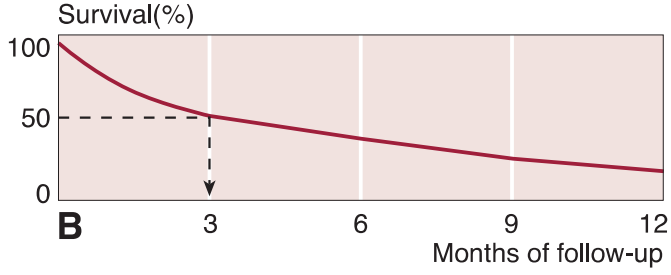


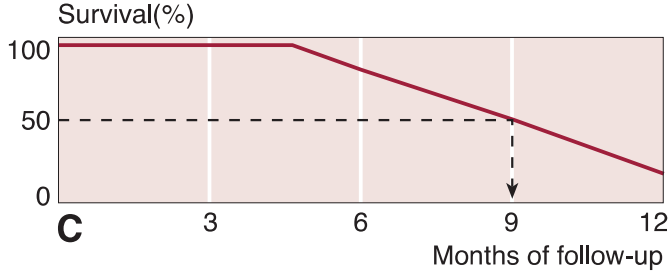
1- Survival curve

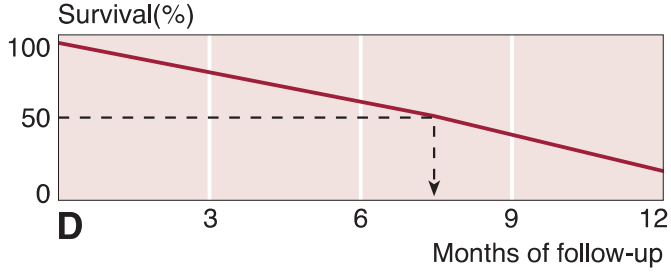
در این منحنی تا انتهای سال اول هیچگونه تغییر قابل توجهی در میزان بقای افراد حاصل نشده است که می‌تواند موید دو حالت باشد، در این موارد یا پیش‌آگهی بیماری بسیار عالی است و یا اینکه ختم زودهنگام مطالعه مانع از بروز پیامد مورد نظر شده است (دراینصورت نتایج مطالعه چندان کاربردی ندارد) شکل B

این منحنی مانند منحنی‌های C و D با توجه به میزان بقای ۲۰ درصد در انتهای ماه دوازدهم نشاندهندهٔ یک بیماری جدی و کشنده است. با این حال شکل منحنی در هر کدام کاملاً متفاوت است. در منحنیB ملاحظه می‌کنید که پس از ۳ ماه فقط نیمی از بیماران زنده هستند. شکل C

در این منحنی در ابتدای سیر بیماری (تا ماه پنجم) پیش‌آگهی خوب است ولی بعد از آن به شدت افت می‌کند بطوریکه میانهٔ میزان بقا برابر ۹ ماه است.شکل D







در این منحنی سیر بیماری یکنواخت و پیوسته در حال نزول است. با میانهٔ بقای ۷/۵ ماه و میزان بقای ۲۰ درصد در پایان ۱۲ ماه.

در این مطالعه خطر تشنج بعد از هر نوع سکتهٔ مغزی ۵/۷ درصد در سال اول تخمین زده شده است. (شمارهٔ ۷)

The probability of having a post stroke seizure was 5.7% (3.5% to 7.9%) within the first year and 11.5% (4.8% to 18.2%) within 5 years. The incremental risk averaged about 1.5% a year after the first 12 months.

شمارهٔ ۷

به طور کلی در مطالعهٔ مورد بررسی از BMJ به اختصار می‌توان به نکات مثبت و منفی زیر اشاره کرد:

در این مطالعه هیچگونه اشاره‌ای به کور بودن ارزیابی پیامدها نشده است، همچنین در مورد تصحیح عوامل موثر بر پیش‌آگهی در زیرگروههای تعیین شده نیز اطلاعات دقیقی در اختیار خواننده قرار نگرفته است. عدم وجود معیارهای مشخص برای وارد کردن افراد به گروه مطالعه نیز از نکات منفی این مقاله است.

با این حال تعریف معیارهای مشخص و از قبل تعیین شده برای سنجش پیامدهای بیماران نکته مثبت و مهمی است که نباید از آن صرف‌نظر کرد. با توجه به نکات فوق و دقت آماری مطالعه و اهمیت بالینی نتایج آن در کل به سختی می‌توان از اشکالات این مطالعه صرف‌نظر کرد.

References

1- Egger M, Smith GD, Altman DG, Systematic Reviews in Health Care: Meta-analysis in context. 2nd ed. BMJ Publishing Group 2001

2- Guyatt G, Rennie D, Hayward R. User's Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice.American Medical Association; 2002.

3- Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine, How to Practice and Teach EBM, 2nd ed. Churchill Livingstone 2000.

فصل هشتم
شواهد مربوط به شیوع و بروز

در مطب خود نشسته‌اید یکی از دوستان پدر ۶۰ سالهٔ خود را پیش شما می‌آورد و می‌گوید: پدرم مدتی است که دچار فراموشی پیشرونده شده و دیگر من را نمی‌شناسد. اولین تشخیصی که به ذهنتان می‌رسد بیماریهای نرودژنراتیو مانند آلزایمر است. به یاد می‌آورید از سه ماه پیش که برای طبابت وارد این شهر شده‌اید، این چندمین بیماری است که با شکایت‌های مشابه مانند فراموشی، از دست دادن تدریجی حافظه و تغییر خلق و خو به شما مراجعه می‌کند و شما برای وی آلزایمر تشخیص می‌دهید. این سوال برایتان پیش می‌آید که شیوع آلزایمر در این منطقه چقدر است. بعد از ویزیت آخرین بیمار به کتابخانهٔ نزدیک مطب رفته و شروع به جستجو می‌کنید. در حین جستجو به مقاله‌ای از نشریهٔ نرولوژی[1] برمی‌خورید که به نظر می‌رسد مستقیماً به پاسخ شما مربوط است. عنوان مقاله « شیوع و علل دمانس در افراد زیر ۶۵ سال»[2] است. متن کامل مقاله را از کتابخانه دریافت کرده و آمادهٔ ارزیابی آن با چک لیست زیر می‌شوید: (متن مقاله در ضمیمهٔ شماره ۶ آمده است.)

آیا مطالعه از کیفیت مناسبی برخوردار است؟

- آیا روش نمونه‌گیری متناسب با سوال طرح شده است؟
- آیا نمونهٔ انتخاب شده، نماینده جمعیت است؟

1- Journal of Neurology, Neurosurgery, and Psychiatry (JNNP)
2- "The Prevalence and causes of dementia in people under the age of 65 years" J. Neurol, Neurosurg. Psychiatry 2003; 74: 1206-1209

- آیا اندازهٔ نمونه کافی است؟
- آیا از معیارهای عینی و استاندارد برای سنجش پیامدهای بالینی استفاده شده است؟
- آیا میزان پیگیری کافی است؟
- آیا در صورت وجود، علت خروج افراد از مطالعه ذکر شده است؟
- **دقت آماری نتایج مطالعه چه میزان است؟**
- آیا تخمین بروز و شیوع، با توجه به فاصله اطمینان ذکر شده مناسب است؟

شواهد مربوط به شیوع و بروز بیماریها از جنبه‌های مختلفی حائز اهمیت هستند. از آنجاییکه شیوع و بروز بسیاری از بیماریها در مناطق مختلف متفاوت است، در اختیار داشتن شواهد کیفی در این زمینه می‌تواند در تشخیص بیماریها به پزشکان کمک کند.

همچنین همانطور که در فصل ۳ نیز گفته شد پیش از تقاضای هر تست تشخیصی پزشک باید تخمین مناسبی از احتمال پیش آزمون برای بیمار خود داشته باشد. شیوع و بروز بیماریها در هر منطقه در کنار شرایط خاص بیمار یکی از عوامل اصلی تعیین کنندهٔ احتمال پیش از تست است.

از طرف دیگر مشخص کردن شیوع و بروز بیماریها می‌تواند اطلاعات ارزشمندی جهت تأمین نیازهای پزشکی و سلامت جامعه در اختیار سیاست‌گزاران و برنامه‌ریزان قرار دهد و کمک ارزنده‌ای برای برنامه‌ریزی خدمات بهداشتی درمانی مورد نیاز در هر منطقه باشد.

پیش از توضیح چک لیست و ارزیابی نقادانهٔ مطالعه، جهت یادآوری توضیح مختصری در مورد شیوع و بروز داده می‌شود.

همانطور که می‌دانید شیوع عبارتست از نسبتی از افراد جامعه که در یک زمان خاص مبتلا به بیماری مورد نظر هستند. به عبارت دیگر شیوع در واقع فراوانی موردهای موجود بیماری است.

بروز نسبتی از افراد جامعه است که در یک بازهٔ زمانی مشخص، به بیماری مورد نظر مبتلا می‌شوند، یا به عبارت دیگر میزان موارد جدید یک مشکل بالینی.

بروز و شیوع از طریق مدت زمان بیماری با یکدیگر ارتباط دارند. (طول دورهٔ بیماری X بروز = شیوع) با توجه به این رابطه بالطبع با طولانی‌تر شدن مدت زمان بیماری شیوع آن نیز افزایش می‌یابد. بیماریهای مزمن که از بروز کمی برخوردارند به این علت شیوع بالایی دارند.

شواهد مربوط به شیوع و بروز عمدتاً به صورت مطالعات همه گیر شناختی و به ویژه مطالعات پیمایشی[1] است. این نوع مطالعات طرح ایده‌آلی برای تعیین شیوع یک مشکل بالینی هستند. مطالعاتی که بروز یک بیماری را بررسی می‌کنند باید طرح طولی[2] یا آینده‌نگر[3] داشته باشند.

طرح مطالعهٔ مورد بررسی از نوع پیمایشی است که برای تعیین شیوع کاملاً مناسب است. (شمارهٔ ۱)

Design: Epidemiological catchment area prevalence survey.

شمارهٔ ۱

آیا مطالعه از کیفیت مناسبی برخوردار است؟

- آیا روش نمونه‌گیری متناسب با سوال طرح شده است؟
- آیا نمونهٔ انتخاب شده نمایندهٔ[4] جمعیت است؟

اگر نمونهٔ مطالعه تمام افراد جمعیت تعریف شده را شامل شود، مطالعه در حقیقت به یک سرشماری[5] تبدیل می‌شود و دیگر

1- Survey Studies
2- Longitudinal
3- Prospective
4- representative
5- Census

مسائل مربوط به نمونه‌گیری در آن مطرح نخواهد شد. در غیر اینصورت بهترین روش، نمونه‌گیری تصادفی از افراد موجود در زیرمجموعه‌های تعریف شده از جمعیت است (با فرایند این کار قبلاً در فصل روش‌شناسی تحقیق آشنا شده‌اید). نمونه‌گیری تصادفی امکان تعمیم‌دادن[1] نتایج بدست آمده از مطالعه را به افراد دیگر و در زمان و مکان متفاوت با سهولت بیشتری فراهم می‌کند، مطالعهٔ مورد بررسی به صورت پیمایشی در چهار بخش از انگلستان انجام شده است و در تمام این چهار بخش اطلاعات به صورت سرشماری جمع‌آوری شده است. (شمارهٔ ۲)

Participants: All residents of the catchment area with dementia, where the illness began before the age of 65 years. Participants were notified to the study by medical and care professionals. The diagnosis and age of onset was established from all available health and social care records. In total, 227 people were identified, of whom 185 fulfilled the inclusion criteria of having a dementia which started before their 65th birthday.

شمارهٔ ۲

■ آیا اندازهٔ نمونه کافی است؟

هر چقدر حجم نمونه بزرگتر باشد، نتیجهٔ بدست آمده از نمونه به مقدار واقعی نزدیک‌تر است. حجم نمونهٔ بزرگتر از طریق کاهش خطای معیار، طول بازهٔ اطمینان را کاهش داده و بدین ترتیب باعث کاهش خطای تصادفی[2] و افزایش دقت مطالعه می‌شود.

در مواقعی که می‌خواهیم شیوع یا بروز یک بیماری نادر را بررسی کنیم، حجم نمونهٔ کافی اهمیت بیشتری پیدا می‌کند. حجم نمونه‌ای که برای بررسی شیوع یک بیماری لازم است با درجهٔ مشخصی از اطمینان قابل محاسبه است.

در این مطالعه کل جمعیت ناحیه ۵۶۷/۵۰۰ نفر بود که از بین آنها ۲۲۷ نفر مبتلا شناسایی شدند. (شمارهٔ ۳)

1- Generalization
2- Random error

Setting: The London boroughs of Kensington and Chelsea, Westminster, and Hillingdon with a total population of 567 500 people.

شمارهٔ ۳

Sampling

Stage 2

A hierarchical diagnostic algorithm was then applied. At the top level, subjects achieved a definite diagnosis if the dementia was associated with a known disease-specific genetic mutation, if there were neuropathological results available from a cerebral biopsy, or from autopsy if the patient had died after the "census" date.

At the second level of the algorithm, well recognised and validated clinical diagnostic criteria were applied to make the diagnoses of Alzheimer's disease (NINCDS/ADRDA criteria)[11], vascular dementia (NINDS/AIREN criteria)[12], dementia with Lewy bodies,[13] frontotemporal dementia (Lund and Manchester Criteria)[14], alcohol related dementia (DSM-IV criteria for alcohol induced persisting dementia)[15], or other dementias for which DSM-IV criteria are available. These criteria were not applied hierarchically, but rather on the basis of best clinical opinion.

At the third level of the algorithm, patients fulfilling the DSM-IV criteria for dementia, but not fulfilling criteria for one of the above diagnostic categories were assigned to a Dementia Not Otherwise Specified category.

The age at onset of disease was defined as the age of the patient at which the earliest conclusive dementia symptom was noted or documented in the medical notes or other correspondence.

شمارهٔ ۴

- آیا از معیارهای عینی و استاندارد برای سنجش پیامدهای بالینی استفاده شده است؟

به طور معمول در پژوهشهای انجام شده در سطح جامعه به علت بالا بودن هزینه تستهای تشخیصی، اغلب برای ارزیابی افراد از سنجه‌های پیامد خام[1] استفاده می‌شود. در حالیکه این سنجه‌ها عموماً برای تعیین شیوع و بروز مناسب نیستند.

نحوهٔ سنجش و ارزیابی پیامدها باید به تشریح در گزارش مقاله ذکر شود. مثلاً چنانچه یک سنجهٔ استاندارد و مورد قبول برای تشخیص یک بیماری وجود دارد و مطالعه از آن استفاده کرده است باید جزئیات آن سنجه و شواهد موید روایی آن در مطالعه ذکر شود همچنین باید چگونگی و میزان رابطهٔ سنجهٔ استفاده شده در مطالعه با سنجه‌های استاندارد بیان شود.

1- Crude Outcome measures

بطور کلی سنجه‌های به کار رفته باید از روایی کافی و حساسیت و ویژگی بالا برخوردار باشند.

در این مطالعه ذکر شده است که در مرحلهٔ دوم نمونه‌گیری، از یک الگوریتم تشخیصی برای ارزیابی بیماری استفاده شده است. این الگوریتم دارای معیارهای استاندارد جهانی برای تشخیص انواع مختلف دمانس است. (شمارهٔ ۴)

- آیا میزان پیگیری کافی بوده است؟ آیا علت خروج افراد از مطالعه ذکر شده است؟

هر چه تعداد بیشتری از افراد انتخاب شده در نمونه از مطالعه[1] خارج شوند و یا هنگام سنجش در دسترس نباشند، از روایی نتایج مطالعه بیشتر کاسته می‌شود. در مطالعاتی که تعداد افراد خارج شده از مطالعه بسیار محدود و ناچیز باشد، روایی خارجی و در نتیجه قدرت تعمیم‌پذیری مطالعه نیز بیشتر خواهد بود. بسته به نوع بیماری مورد بررسی خارج شدن کمتر از 20% افراد حین مطالعه قابل‌قبول است مثلاً در مورد بیماری دمانس، بسیاری از بیماران به علت ماهیت بیماری از مطالعه خارج می‌شوند و این امر باعث نادیده گرفتن میزان واقعی شیوع آن می‌شود. انجام یک سری اقدامات می‌تواند تعداد افرادی را که به هر علت از مطالعه خارج می‌شود، کاهش دهد. برای مثال اگر روش ارزیابی و تشخیص بیماری در دسترس باشد، زمان مناسبی برای انجام سنجش‌ها انتخاب شود، طول مدت هر ارزیابی قابل قبول باشد و نحوهٔ آن از نظر بیمار پذیرفتنی باشد، تعداد افراد خارج شده به حداقل می‌رسد. (به عنوان مثال، ویزیت‌های خانگی برای افراد سالمند پذیرفته‌تر است).

در مطالعهٔ نمونه، ۱۹ نفر از ۲۲۷ نفر، به علت اینکه معیارهای DSM=IV را پر نمی‌کردند از مطالعه خارج شده‌اند (کمتر از ۲۰ درصد) و علل خروج آنها نیز به خوبی ذکر شده است. (شمارهٔ ۵)

Of the 227 referrals, 19 were excluded as not fulfilling DSM-IV criteria for dementia (six were "worried well", seven had traumatic brain injuries, two had chronic schizophrenia, and four cases identified from hospital information systems were clearly miscoded when the notes were reviewed).

شمارهٔ ۵

1- Refusal / drop outs

برای تعیین بروز بیماری‌ها، لازم است حتی‌الامکان تمام افراد درون مطالعه برای یک دورهٔ معقول بسته به نوع بیماری و سن جمعیت پیگیری شوند. اگر دورهٔ پیگیری بیماریهایی مانند دمانس خیلی طولانی باشد، بسیاری از بیماران به علت سن بالا فوت کرده و بدین گونه از مطالعه خارج می‌شوند.

دقت آماری نتایج مطالعه چه میزان است؟

- آیا تخمین بروز و شیوع با توجه به فاصلهٔ اطمینان ذکر شده مناسب است؟

نتایج کمّی در مطالعات بروز و شیوع، به صورت نسبت‌هایی در طی یک دورهٔ زمانی ثابت بیان می‌شوند. میزان شیوع واقعی یک مشکل بالینی در جامعه معمولاً با احتمال ۹۵ درصد درون بازهٔ تعیین شده در مطالعه قرار می‌گیرد. با اطمینان ثابت هر چه طول بازه کمتر باشد تخمین ما شیوع یا بروز واقعی مشکل بالینی دقیق‌تر است. در این مطالعه، شیوع بدست آمده برای دمانس با توجه به حجم بزرگ نمونه از بازهٔ اطمینان مناسبی برخوردار است. (شمارهٔ ۶)

Table 1 Age and gender specific prevalence rates in the study population

Population			All causes of dementia									
			All			Male			Female			
Age range	Male (n)	Female (n)	n	Rate*	95% CI	n	Rate	95% CI	n	Rate	95% CI	Significance†
30–34	23898	23375	6	12.7	(4.7 to 26.7)‡	3	12.6	(2.6 to 36.7)	3	12.8	(2.7 to 37.5)	NS
35–39	18526	19106	3	8.0	(1.6 to 23.3)	1	5.4	(0.1 to 30.1)	2	10.5	(1.3 to 37.8)	NS
40–44	18982	19643	6	15.5	(5.7 to 33.8)	1	5.3	(0.1 to 29.4)	5	25.5	(8.3 to 59.4)	NS
45–49	16549	16799	11	33.0	(16.5 to 59.0)	6	36.3	(13.3 to 78.9)	5	29.8	(9.7 to 69.5)	NS
50–54	15185	15237	19	62.5	(37.6 to 97.5)	10	65.9	(31.6 to 121)	9	59.1	(27 to 112)	NS
55–59	13983	13626	42	152.1	(110 to 206)	28	200.2	(133 to 289)	14	102.7	(56.2 to 172)	NS
60–64	12716	13141	43	166.3	(120 to 224)	26	204.5	(134 to 300)	17	129.4	(75.4 to 207)	NS
30–64	119839	120927	130	54.0	(45.1 to 64.1)	75	62.6	(49.2 to 78.4)	55	45.5	(34.3 to 59.2)	NS
45–64	58433	58803	115	98.1	(81.1 to 118)	70	119.8	(93.4 to 151.0)	45	76.5	(55.8 to 102)	NS
Over 65§			55			33			22			

*Rate per 100 000 people at risk.
†Significance of difference between genders by inference from 95% CI.
‡95% confidence interval for the prevalence rate.
§Subjects who had a dementia starting before the age of 65 years, but were over 65 but still living on the study census day.

شمارهٔ ۶

به طور خلاصه طرح مطالعه از نوع پیمایشی است و نمونه‌گیری آن به صورت سرشماری در ۴ منطقهٔ وسیع از انگلستان انجام شده است، که جمعیت بالایی را تحت پوشش قرار می‌دهد. میزان شیوع دمانس در این ۴ ناحیه با توجه به سن و جنس در هر زیرگروه از انواع دمانس ذکر شده است. در این مطالعه برای تشخیص دمانس از معیارهای معتبری استفاده شده است و میزان حذف از پیگیری کمتر از ۲۰ درصد بوده است و علت خروج افراد کاملاً ذکر شده است. دقت آماری مطالعه نیز با توجه به بازهٔ اطمینان ۹۵ درصد مناسب است بنابراین این مطالعه اکثر معیارهای ذکر شده را رعایت کرده است.

References

1- How to Review the Evidence: Systematic identification and review of the scientific literature. NHMRC; 1999.

2- Greenhalgh T. How to Read a Paper. 2nd ed. BMJ publishing Group 2001.

3- Riegelman RK. Studying a Study and Testing a Test. 4th ed. Lippincott Williams and Wilkins; 2000.

4- Trochim WM. Research Methods Knowledge Base. New York. Cornell University; 2003.

5-Public Health Resource unit of NHS.

Table of Contents

APPENDICES

1 absolute benefit increase ..... افزایش فایده مطلق
2 absolute risk ..... خطر مطلق
3 accidental sampling ..... نمونه‌گیری اتفاقی
4 adjustment ..... تصحیح
5 aim ..... هدف
6 allocation concealment / randomisation blinding ..... مخفی بودن تخصیص
7 alternative hypothesis ..... فرضیه جایگزین
8 assignment / allocation ..... تخصیص
9 association ..... همراهی
10 attrition bias ..... تورش پیگیری
11 bell-shaped ..... زنگوله‌ای شکل
12 benefits ..... فواید
13 binary Scale ..... مقیاس دوتایی
14 biological plausibility ..... امکانپذیری زیست شناختی
15 case series ..... بررسی موردها
16 case-control study ..... مطالعه مورد شاهدی
17 case-report ..... گزارش موردی

18 causal ..... علّی
19 census ..... سرشماری
20 central allocation ..... تخصیص مرکزی
21 central tendency ..... تمایل به مرکز
22 checklist ..... چک لیست
23 clinical relevant outcome ..... پیامد مربوط به بالین
24 cluster random sampling ..... نمونه‌گیری تصادفی خوشه‌ایی
25 cohort ..... هم‌گروهی
26 cointervention ..... هم درمانی
27 comorbidity ..... بیماری همراه
28 confidence interval ..... بازهٔ اطمینان
29 consecutive sampling ..... نمونه‌گیری پی‌درپی
30 consistency of association ..... سازگاری در همراهی
31 continuous scale ..... مقیاس پیوسته
32 control event rate ..... میزان بروز رخداد در گروه کنترل
33 critical Appraisal ..... ارزیابی نقادانه شواهد
34 cross-section ..... مقطع زمانی
35 cross-sectional design ..... طرح مقطعی

36 crude outcome measure ..... سنجه پیامد خام
37 cut off level ..... سطح معناداری
38 data collection ..... جمع‌آوری داده‌ها
39 database ..... پایگاه داده‌ای
40 demographic characteristics ..... ویژگیهای جمعیت شناختی
41 descriptive ..... توصیفی
42 design ..... طرح
43 detection bias / measurement bias ..... تورش سنجش
44 diagnostic test accuracy ..... دقت تست تشخیصی
45 diagnostic test performace ..... عملکرد تست تشخیصی
46 differential verification bias ..... تورش تأیید افتراقی
47 dimensions ofevidence ..... ابعاد شواهد
48 dispersal ..... پراکنش
49 etiology ..... اتیولوژی
50 experiment event rate ..... میزان بروز رخداد در گروه مداخله
51 experimental ..... تجربی
52 external validity ..... روایی خارجی
53 factorial design ..... طرح عاملی

54 false positve rate ..... نسبت مثبت کاذب
55 harm ..... آسیب
56 hazard ratio ..... نسبت مخاطره
57 historical cohort ..... همگروهی تاریخی
58 historical control ..... کنترل پیشینه‌ای
59 hypothesis ..... فرضیه
60 hypothesis testing ..... آزمون فرضیه
61 inception cohort ..... همگروهی آغازین
62 incidence ..... بروز
63 inclusion & exclusion criteria ..... معیارهای ورود و خروج
64 incorporation bias ..... تورش مشارکت
65 inferential ..... استنباطی
66 intention to treat analysis ..... تحلیل قصد درمان
67 interaction ..... اثر متقابل
68 internal validity ..... روایی داخلی
69 interrupted time series ..... سری‌های زمانی منقطع
70 interview ..... مصاحبه
71 investigator bias ..... تورش پژوهشگر

72 lead time bias ..... تورش زمان تشخیص
73 length bias ..... تورش ازمان بیماری
74 likelihood ratio ..... نسبت درست‌نمایی
75 longitudinal ..... طولی
76 measurment of outcome ..... سنجش پیامد
77 median survival ..... میانهٔ بقاء
78 morbidity ..... موربیدیته
79 morning report ..... گزارش صبحگاهی
80 multiple regression analysis ..... تحلیل رگرسیون چندگانه
81 multistage sampling ..... نمونه‌گیری چند مرحله‌ای
82 N of one randomised trial ..... کارآزمایی تصادفی تکنفره
83 NNH ..... تعداد.لازم.برای آسیب.
84 NNS ..... تعداد لازم برای غربالگری
85 NNT ..... تعداد لازم برای درمان
86 noncompliance ..... عدم تبعیت
87 non-experimental ..... غیر تجربی
88 non-probability sampling ..... نمونه‌گیری غیراحتمالی
89 null hypotheses ..... فرضیه صفر

90 objective ..... عینی
91 observational methods ..... روشهای مشاهده‌ای
92 odds ..... شانس
93 Odds ratio ..... نسبت شانس
94 parameter ..... پارامتر
95 partial verification bias ..... تورش تایید ناکامل
96 patient relevant outcome ..... پیامد مربوط به بیمار
97 performance bias ..... تورش عملکرد
98 postest probability ..... احتمال پس آزمون
99 post-test ..... سنجش بعد از مداخله
100 predictive value ..... ارزش اخباری
101 prenumbered or coded identical container ..... محفظه‌های شماره‌دار مشابه
102 pretest probability ..... احتمال پیش‌آزمون
103 pretest / post test ..... سنجش قبل و بعد از مداخله
104 prevalence ..... شیوع
105 prior association ..... تقدم در همراهی
106 probabilistic measures ..... سنجه‌های احتمالی
107 probability sampling ..... نمونه‌گیری احتمالی

108 protocol ..... پروتکل
109 pseudo RCT ..... کارآزمایی شبه تصادفی شده شاهددار
110 purposive ..... هدفمند
111 qualitative ..... کیفی
112 quality of evidence ..... کیفیت شواهد
113 quality of life ..... کیفیت زندگی
114 quantitative ..... کمّی
115 quasi random selection ..... شبه تصادفی
116 quasi-experimental ..... شبه تجربی
117 random error ..... خطای تصادفی
118 Randomized crossover trial ..... کارآزمایی تصادفی شده متقاطع
119 range ..... دامنه
120 raw data ..... داده خام
121 reasearch strategies ..... راهبردهای پژوهش
122 recall bias ..... تورش یادآوری
123 relational ..... ارتباطی
124 relative benefit increase ..... افزایش فایده نسبی
125 relative risk ..... خطر نسبی

126 relative risk reduction ..... کاهش خطر نسبی
127 relevance of evidence ..... مرتبط بودن شواهد
128 relevancy ..... مرتبط بودن
129 relevant ..... مربوط
130 reliability ..... پایایی
131 reporting bias ..... تورش گزارش‌دهی
132 research methodology ..... روش‌شناسی پژوهش
133 research planning ..... طرح‌ریزی پژوهش
134 risk difference ..... تفاوت خطر
135 risk factor ..... عامل خطر
136 risk ratio ..... نسبت خطر
137 sample ..... نمونه
138 sample size ..... حجم نمونه
139 sampling ..... نمونه‌گیری
140 sampling distribution ..... توزیع نمونه‌گیری
141 sampling error ..... خطای نمونه‌گیری
142 sampling frame ..... چارچوب نمونه‌گیری
143 sampling methods ..... روشهای نمونه‌گیری

144 scale ..... مقیاس
145 scale of measurment ..... مقیاس سنجش
146 scientific method ..... روش علمی
147 screening ..... غربالگری
148 selection bias ..... تورش انتخاب
149 sensitivity ..... حساسیت
150 side effects ..... اثرات جانبی
151 signs & symptoms ..... علائم و نشانه‌ها
152 simple random ..... تصادفی ساده
153 size of effect ..... اندازه اثر
154 specificity ..... ویژگی
155 spectrum bias ..... تورش طیف
156 standard error ..... خطای معیار
157 statistic ..... آماره
158 statistical precision ..... دقت آماری
159 statistical significance testing ..... آزمون معناداری آماری
160 strata ..... طبقه
161 stratified analysis ..... تحلیل طبقه‌ای

162 stratified random ..... تصادفی طبقه‌ای
163 stratistical power ..... توان آماری
164 strength of evidence ..... قدرت شواهد
165 stroke ..... سکته مغزی
166 study population ..... جمعیت مطالعه
167 subjective ..... ذهنی
168 subjective outcome measure ..... سنجه پیامد ذهنی
169 surogate outcome ..... پیامد جایگزین
170 survey designs ..... طرحهای پیمایشی
171 survival ..... بقاء
172 survival curve ..... منحنی بقاء
173 survival rate ..... میزان بقاء
174 systematic review ..... مقاله مروری نظام‌مند
175 test failure ..... نارسایی تست
176 theoretical population ..... جمعیت تئوری
177 total age-specific mortality ..... مرگ و میر کلی هر سن خاص
178 trade off ..... معامله
179 treatment paradox ..... تناقض درمان

180 true random selection ..... انتخاب تصادفی واقعی

181 type I error ..... خطای نوع ۱

182 type II error ..... خطای نوع ۲

183 validity ..... روایی

184 variability ..... تغییرپذیری

185 .verification bias ..... تورش تأیید

References

1- How to Review the Evidence: Systematic identification and review of the scientific literature. NHMRC; 1999.

2- Greenhalgh T. How to Read a Paper. 2nd ed. BMJ publishing Group 2001.

3- Riegelman RK. Studying a Study and Testing a Test. 4th ed. Lippincott Williams and Wilkins; 2000.

4- Trochim WM. Research Methods Knowledge Base. New York. Cornell University; 2003.

5-Public Health Resource unit of NHS.

APPENDICES and Glossary

APPENDIX 1

The New England Journal of Medicine

COLOR DUPLEX ULTRASONOGRAPHY IN THE DIAGNOSIS OF TEMPORAL ARTERITIS

WOLFGANG A. SCHMIDT, M.D., HELGA E. KRAFT, M.D., KLAUS VORPAHL, M.D., LUTZ VÖLKER, M.D., AND ERIKA J. GROMNICA-IHLE, M.D.

ABSTRACT

Background The diagnosis of temporal arteritis usually requires a biopsy of the temporal artery. We examined the usefulness of color duplex ultrasonography in patients suspected of having temporal arteritis.

Methods In this prospective study, all patients seen in the departments of rheumatology and ophthalmology from January 1994 to October 1996 who had clinically suspected active temporal arteritis or polymyalgia rheumatica were examined by duplex ultrasonography. The final diagnoses, made according to standard criteria, were temporal arteritis in 30 patients, 21 with biopsy-confirmed disease; polymyalgia rheumatica in 37; and negative histologic findings and a diagnosis other than temporal arteritis or polymyalgia rheumatica in 15. We also studied 30 control patients matched for age and sex to the patients with arteritis. Two ultrasound studies were performed and read before the biopsies; one ultrasonographer was unaware of the clinical information.

Results In 22 (73 percent) of the 30 patients with temporal arteritis, ultrasonography showed a dark halo around the lumen of the temporal arteries. The halos disappeared after a mean of 16 days (range, 7 to 56) of treatment with corticosteroids. Twenty-four patients (80 percent) had stenoses or occlusions of temporal-artery segments, and 28 patients (93 percent) had stenoses, occlusions, or a halo. No halos were identified in the 82 patients without temporal arteritis; 6 (7 percent) had stenoses or occlusions. For each of the three types of abnormalities identified by ultrasonography, the interrater agreement was ≥95 percent.

Conclusions There are characteristic signs of temporal arteritis that can be visualized by color duplex ultrasonography. The most specific sign is a dark halo, which may be due to edema of the artery wall. In patients with typical clinical signs and a halo on ultrasonography, it may be possible to make a diagnosis of temporal arteritis and begin treatment without performing a temporal-artery biopsy. (N Engl J Med 1997;337:1336-42.)

TEMPORAL arteritis is sometimes diagnosed clinically, but a temporal-artery biopsy is generally recommended to confirm the diagnosis. The American College of Rheumatology requires three of the following five criteria to be met to establish the diagnosis: age ≥50 years, new onset of localized headache, temporal-artery tenderness or decreased pulse, erythrocyte sedimentation rate ≥50 mm per hour, and histologic findings.[1] Although temporal-artery biopsy is a minor operation, not all patients agree to this procedure. There are occasional complications, including damage of the facial nerve,[2] skin necrosis,[3] drooping of the eyebrow,[4] and stroke due to an interruption of a collateral circulation.[5] It takes time to obtain the histologic results, and they are sometimes normal because the biopsy specimen was obtained from an area without lesions.

Color duplex ultrasonography is an excellent method of examining blood vessels. It combines the imaging capabilities of B-mode ultrasonography with the flow-velocity determinations of Doppler ultrasonography and permits accurate assessment of both the arterial anatomy and the flow characteristics of the vessel at specific sites. The new high-frequency ultrasound scanners with a high resolution make it possible to examine even small vessels such as the superficial temporal artery, which is located 3 to 4 mm below the surface of the skin. We have observed a hypoechoic halo around the perfused lumen of inflamed temporal arteries, which disappeared with treatment.[6] We undertook this study to assess the use of this method in the diagnosis of temporal arteritis.

METHODS

Patients

All consecutive patients seen in the departments of rheumatology and ophthalmology between January 1994 and October 1996 with clinically suspected active temporal arteritis and polymyalgia rheumatica as well as all patients who had temporal-artery biopsies to rule out the condition were included in the study. Once a diagnosis was established, patients were classified according to whether they had temporal arteritis, polymyalgia rheumatica, or negative histologic findings and a diagnosis other than temporal arteritis or polymyalgia rheumatica. In addition, control patients matched for age and sex to the patients with temporal arteritis were included. Patients with a history of temporal arteritis but no evidence of active disease and patients who had had a previous biopsy of the temporal arteries were not included.

Patients with Temporal Arteritis

Thirty patients (21 of them women) with a mean age of 73 years (range, 52 to 86) had active temporal arteritis. Five had not

From the Clinic of Rheumatology, Berlin-Buch (W.A.S., E.J.G.-I.), and the Departments of Ophthalmology (H.E.K.), Pathology (K.V.), and Angiology (L.V.), Klinikum Buch — both in Berlin, Germany. Address reprint requests to Dr. Schmidt at the Clinic of Rheumatology, Zepernicker Strasse 1, 13125 Berlin, Germany.

received corticosteroids before undergoing ultrasonography, 10 had received corticosteroids for less than 24 hours beforehand, 11 had received corticosteroids for 1 to 10 days beforehand, and 4 had been treated with doses of prednisolone that were too low to suppress disease activity. They all met at least three criteria for the diagnosis of temporal arteritis[1]: 7 patients met three of the criteria, 9 met four, and 14 met all five criteria. The clinical diagnosis was confirmed by two rheumatologists and one ophthalmologist who examined every patient.

Twenty-seven patients underwent biopsy, and three declined to undergo biopsy. The findings were positive in 21, and negative in 4. In two patients the biopsy yielded insufficient material for analysis.

Clinically, 16 patients had symptoms of temporal arteritis and polymyalgia rheumatica, 12 of temporal arteritis alone, and 2 of polymyalgia rheumatica alone. The last two were classified as having temporal arteritis on the basis of histologic findings. Thirteen patients (43 percent) had visual disturbances: seven had definite anterior ischemic optic neuropathy (repeated episodes in three patients and affecting both eyes in two), three had amaurosis fugax (repeated episodes in one), and one had a single definite episode of occlusion of the central artery of one eye (which did not resolve), one had a single transitory episode of cotton-wool exudates affecting both eyes, and one had a single transient episode of bilateral abducens paralysis.

Patients with Polymyalgia Rheumatica

Thirty-seven patients (28 women and 9 men; mean age, 70 years; range, 51 to 86) were given a diagnosis of polymyalgia rheumatica because they met three or more of the seven criteria of the Bird classification.[7] Four of the patients met three criteria, 12 met four, 9 met five, 10 met six, and 2 met all seven criteria. All patients met fewer than three of the criteria for temporal arteritis.[1] The diagnosis was confirmed by two rheumatologists who examined every patient. Seven patients had biopsies of the temporal arteries, all of which were negative. No patient had been treated with corticosteroids for more than 10 days. Sixteen patients had received corticosteroids for 1 to 10 days before ultrasonography, 7 had received corticosteroids for less than 24 hours beforehand, and 14 had not received corticosteroids beforehand.

Control Patients

Thirty patients with an established diagnosis of rheumatoid arthritis and no signs of temporal arteritis were chosen from the clinic of rheumatology and matched for age and sex to the patients with temporal arteritis. Fifteen patients received corticosteroids before undergoing ultrasonography.

Patients with Negative Histologic Findings and Diagnoses Other Than Temporal Arteritis or Polymyalgia Rheumatica

Fifteen patients (six women and nine men; mean age, 70 years; range, 49 to 90) were given the following diagnoses after completing the diagnostic procedure: noninflammatory anterior ischemic optic neuropathy in six, bronchial carcinoma in two, and noninflammatory central-artery occlusion, stroke, Hodgkin's lymphoma, glaucoma plus osteomyelitis, plasmacytoma, and idiopathic abducens paralysis in one patient each. The diagnosis remained undefined in one patient. Four patients received corticosteroids for 1 to 10 days before undergoing ultrasonography.

Ultrasonographic Evaluation

Simultaneous color Doppler and duplex ultrasonography was performed with a high-resolution linear scanner (L, 10 to 5; 10 to 5 MHz; length of scanner, 38 mm; ATL Ultramark 9 HDI, Advanced Technology Laboratories, Bothell, Wash.). We examined both common superficial temporal arteries and the frontal and parietal rami as completely as possible in a longitudinal and transverse plane to see whether they were perfused, whether there was a halo around the lumen, and (using simultaneous pulsed-wave Doppler ultrasonography) whether there was a stenosis. Stenosis was considered to be present if blood-flow velocity was more than twice the rate recorded in the area before the stenosis, perhaps with wave forms demonstrating turbulence and reduced velocity behind the area of stenosis. In addition, the diameters of the systolic lumen and wall as well as peak systolic blood-flow velocity were measured at four defined points on each side of the head (Fig. 1).

Two trained physician ultrasonographers performed the examination, which takes an experienced sonographer 20 to 30 minutes. One investigator, who was unaware of the patients' diagnoses, looked only for the existence of a halo, stenosis, or occlusion. If their results disagreed, the investigators examined the results together and reached a consensus on the findings.

The arterial wall was defined to include the intima, media, adventitia, and temporal fascia. The wall could not be seen in the proximal common superficial temporal artery, which is located outside the temporal fascia. The anterior and posterior diameters of the wall were measured throughout the longitudinal plane, and the mean value was calculated as the diameter of each point. When there was a stenosis at one of these points the measurement was made 3 mm proximal to it. Ultrasonography was always performed before biopsy. If there was a halo, both investigators repeated the ultrasound examination every three to four days until the halo disappeared.

For the measurements, the settings on the scanner were always uniform. The settings for color Doppler sonography with our ATL scanner were as follows: dynamic range, 50 dB; signal-processing characteristic, G6; dynamic-contrast gain, K4; color gain, 78 percent; type of color gain, V; color scale, 4; color sensitivity, 10; color wall filter, 100 Hz; color persistence, 4; pulse-repetition frequency, 2500 per second; dynamic-movement differentiation, D3; and focus-point position, 7 mm.

Biopsy

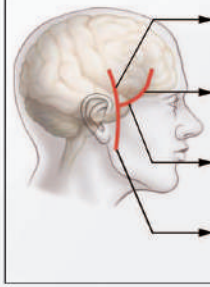
In all cases biopsies were bilateral. Temporal arteritis was diagnosed if there was vasculitis with a predominance of mononuclear-cell infiltration or granulomatous inflammation with or without giant cells.[1]

Statistical Analysis

The SPSS statistical package was used for statistical analysis. The Mann–Whitney U test or the chi-square test was used to compare the results between groups.

RESULTS

Figure 1 shows the diameters of the lumen and wall as well as peak systolic blood-flow velocity at four defined points along the superficial temporal artery in the four groups of subjects. The diameter of the artery wall was significantly larger in patients with temporal arteritis than in those in the other three groups. The diameters of the temporal arteries, including the lumen and wall, measured by ultrasonography can be compared with those determined in anatomical studies. Stock et al.[8] recorded the following diameters in cadavers: common superficial temporal artery, 2.03 mm; frontal branch, 1.74 mm; and parietal branch, 1.83 mm. The mean (±SD) diameters in all subjects except the patients with temporal arteritis were as follows: common superficial temporal artery, 1.7±0.43 mm (our value was smaller because we could not see the wall); frontal ramus (1 cm distal to the bifurcation), 2.2±0.38 mm; and



Site	Patients with Temporal Arteritis (N=30)	Patients with Polymyalgia Rheumatica (N=37)	Control Subjects (N=30)	Patients with Negative Histologic Findings and Other Diagnoses (N=15)
Parietal ramus (15 mm distal to bifurcation)				
Systolic lumen (mm)	0.79±0.29	0.76±0.20	0.89±0.24	0.81±0.30
Wall (mm)	0.94±0.28*	0.70±0.08	0.72±0.13	0.79±0.11
Maximal velocity (cm/sec)	52±18	59±14	54±14	57±18
Frontal ramus (25 mm distal to bifurcation)				
Systolic lumen (mm)	0.67±0.20	0.66±0.22	0.74±0.24	0.68±0.23
Wall (mm)	0.95±0.20*	0.66±0.07	0.65±0.13	0.72±0.09
Maximal velocity (cm/sec)	48±13	53±16	47±15	55±19
Frontal ramus (10 mm distal to bifurcation)				
Systolic lumen (mm)	0.74±0.24	0.71±0.17	0.86±0.26	0.78±0.30
Wall (mm)	0.95±0.22*	0.69±0.09	0.71±0.13	0.76±0.10
Maximal velocity (cm/sec)	50±14	56±15	48±13	59±20
Common superficial temporal artery (8 mm below skin surface)				
Systolic lumen (mm)	1.51±0.44	1.54±0.41	1.70±0.35	1.85±0.54
Maximal velocity (cm/sec)	62±22	61±16	55±13	64±16

Figure 1. Measurement of the Superficial Temporal Arteries.

Plus–minus values are means ±SD of the right and left sides. The arterial wall was defined to include the intima, media, adventitia, and temporal fascia. Asterisks indicate a significant difference ($P<0.01$ by the Mann–Whitney U test) between the patients with temporal arteritis and the other three groups.

parietal ramus (1.5 cm distal to the bifurcation), 2.3±0.38 mm (our value was higher because we also measured the temporal fascia).

Unexpectedly, there was no significant difference between the group with temporal arteritis and the other three groups with respect to the diameter of the systolic lumen and maximal velocity of systolic blood flow along the nonoccluded segments. Men tended to have bigger lumina and higher velocities than women. In some regions of measurement the difference between men and women was significant. The forms of the Doppler curve did not differ significantly between the groups.

Of the 30 patients with temporal arteritis, 22 (73 percent) had a hypoechoic halo around the perfused lumen of the temporal arteries (Fig. 2). The sagittal diameter of the halo was between 0.3 and 1.2 mm. On the anterior and posterior sides of the halo a hyperechoic region remained. We assume that this structure is the temporal fascia, which is not involved in the inflammatory process. The halo was bilateral in 17 patients. The common superficial temporal artery was involved in 15 patients, the frontal ramus in 16, and the parietal ramus in 14. Five patients had the halo in all eight segments analyzed (the common superficial temporal artery, proximal frontal ramus, distal frontal ramus, and parietal ramus on both sides of the head), one had it in seven segments, four had it in six segments, one had it in five segments, three had it in four segments, one had it in three segments, three had it in two segments, and four had it in one segment. On repeated ultrasonography, the halos were always seen in the same segments.

The halo disappeared a mean of 16 days (range, 7 to 56) after the start of therapy with corticosteroids, with no significant difference between the investigators in the mean values. After the biopsy the lumen was smaller and velocity was slower in the ramus proximal to the site of the biopsy. None of the 82 patients in the other three groups had a halo around the perfused lumen of the temporal arteries; thus, this finding seems to be specific for temporal arteritis.

Table 1 compares the results of ultrasonography with those obtained by histologic analysis. Table 2 shows the sensitivity and specificity of the test.

In two patients a halo was seen in the proximal frontal ramus, but the histologic findings were negative in the biopsy specimens, which were taken from another location to avoid the branches of the facial nerve at the proximal frontal ramus. Both patients (age, >50 years) had a clinically established diagnosis of temporal arteritis with new localized headache, tender temporal arteries with decreased pulse, and elevated erythrocyte sedimentation rates. The diagnosis has remained unchanged after more than one year. Both patients had received prednisolone in the past, and the daily doses had been reduced to 5 mg and 2 mg. The discrepancy between the ultrasonographic and histologic findings may be due to the low dose

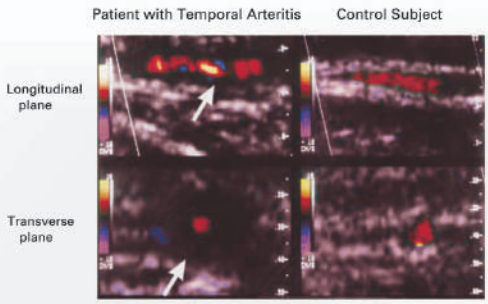


Figure 2. Parietal Ramus of the Superficial Temporal Artery in a Patient with Temporal Arteritis and a Control Subject in a Longitudinal and a Transverse Plane.
The hypoechoic (black) area is indicated by the arrows.

of corticosteroids and, perhaps, to the presence of segmental temporal arteritis.

In 12 of the 16 patients with both positive biopsy findings and a halo on ultrasonography, the biopsy specimen included the region with the halo. All 12 biopsy specimens had mononuclear-cell infiltrates, and 9 of the 12 contained giant cells. There was no association between the presence of a halo and the intensity of mononuclear-cell infiltration, involvement of the intima, or the presence of granulomas or giant cells. Three patients underwent biopsy of occluded segments, which showed giant cells and obliteration due to inflammation.

It was possible to detect nonperfused vessels. The superficial temporal artery is located between the superficial and profound laminae of the temporal fascia. Both laminae are usually found close together but separate at the temporal artery. It is possible to locate the artery at the point of separation by normal B-mode ultrasonography even without color Doppler ultrasonography. Color Doppler ultrasonography is necessary, however, to show whether the lumen is perfused and whether a halo is present.

Patients with polymyalgia rheumatica, control subjects, and patients with negative histologic findings and a diagnosis other than temporal arteritis or polymyalgia rheumatica had significantly fewer stenoses and occlusions than the patients with temporal arteritis ($P<0.01$ by the chi-square test). Two patients with polymyalgia rheumatica had occlusions; one had negative biopsy findings, and the other had not undergone biopsy. Two control subjects had stenoses, and two had occlusions. All four of these control subjects had arteriosclerotic disease. None of the 4 and none of the 10 other subjects with arteriosclerotic disease but not temporal arteritis had a halo.

The rates of agreement between the two ultrasonographers were 100 percent for the initial diagnosis of the halo (112 of 112 tests), 98 percent for stenoses (110 of 112), 96 percent for occlusions (108 of 112), and 95 percent for all three features (106 of 112). Each investigator missed one apparent stenosis and misidentified two areas as occlusions.

Among the patients with temporal arteritis, 11 with a halo and 2 without a halo had eye involvement; 11 with a halo and 6 without a halo had no eye involvement (P not significant by the chi-square test). Among the patients with temporal arteritis, there was no relation between the presence of a halo and the clinical appearance. Of the 22 patients with a halo, 9 had symptoms of isolated temporal arteritis, 12 of temporal arteritis and polymyalgia rheumatica, and 1 of isolated polymyalgia rheumatica. Of the eight patients without a halo, three had symptoms of isolated temporal arteritis, one of polymyalgia rheumatica, and four of both syndromes. Thirteen patients with a halo had hard and prominent temporal arteries, whereas this clinical sign was not found in nine with a halo. Two patients with hard and prominent arteries had no halo; six had neither a halo nor hard and prominent arteries.

There was no association between the presence of a halo and the erythrocyte sedimentation rate. One patient with no visual disturbance had a unilateral headache, with the halo found only on that side. The other patients reported bilateral headache. A halo was found to be topographically related to a diffuse swelling of the right frontal region in another patient.

TABLE 1. COMPARISON OF RESULTS OBTAINED BY ULTRASONOGRAPHY AND HISTOLOGIC ANALYSIS IN 30 PATIENTS WITH TEMPORAL ARTERITIS.*

No. of Patients	Ultrasonography		Histologic Analysis
	Halo	Stenosis or occlusion†	
14	+	+	+
3	+	+	ND
2	+	−	+
1	+	−	ND
1	+	+	−
1	+	−	−
4	−	+	+
1	−	+	−
1	−	+	ND
1	−	−	+
1	−	−	−

*ND denotes not done or that there was insufficient material for analysis. A plus sign indicates positive findings, and a minus sign negative findings.

†Twelve patients had both stenosis and occlusion, six patients had occlusion alone, and six patients had stenosis alone.

TABLE 2. SENSITIVITY AND SPECIFICITY OF DUPLEX ULTRASONOGRAPHY OF THE TEMPORAL ARTERIES FOR THE DIAGNOSIS OF TEMPORAL ARTERITIS AND TO CONFIRM HISTOLOGIC FINDINGS.

Finding	Diagnosis*		Confirmation of Histologic Findings†	
	Sensitivity	Specificity	Sensitivity	Specificity
	positive tests/total (%)	negative tests/total (%)	positive tests/total (%)	negative tests/total (%)
Halo	22/30 (73)	82/82 (100)	16/21 (76)	24/26 (92)
Stenosis or occlusion	24/30 (80)	76/82 (93)	18/21 (86)	23/26 (88)
Halo, stenosis, or occlusion	28/30 (93)	76/82 (93)	20/21 (95)	22/26 (85)

*Thirty patients had temporal arteritis, and 82 patients had been given other diagnoses.

†Twenty-one patients had positive histologic findings, and 26 patients had negative histologic findings (4 in the temporal-arteritis group, 7 in the group with polymyalgia rheumatica, and 15 with other diagnoses).

DISCUSSION

Of the 30 patients we studied with temporal arteritis, 73 percent had a hypoechoic halo around the perfused lumen that disappeared with treatment. At first we thought that this halo might be due to cell infiltrates, but cell infiltrates may be absent from the intima, and the hypoechoic region always began right next to the perfused lumen. In nine of our patients with temporal arteritis we found an intima without cell infiltrates, yet seven of these patients had a halo. We believe that the halo is a sign of fluid in the artery wall. Indeed, edema has been described in temporal arteritis,[9] although it is difficult to document histologically. We think that ultrasonography identifies a different aspect of the disease (edema) than histologic analysis (cell infiltrates), yet both types of findings may be signs of arteritis.

Certain factors could have increased the frequency of the halo in our group. Only 7 of the 37 patients with polymyalgia rheumatica underwent biopsy; therefore, some additional patients might have been classified as having temporal arteritis if they had had biopsies. Many (43 percent) of our patients with temporal arteritis had visual symptoms, so we might have overestimated the incidence of the halo as a result of studying so many patients with severe disease. The diagnosis was confirmed by biopsy in only 21 of the 30 patients with temporal arteritis. With respect to the clinical diagnoses, bias may have resulted from the fact that those who made the diagnosis were in many cases aware of the results of the ultrasound studies.

Angiography has been used in patients with temporal arteritis to pinpoint the region for biopsy and to detect stenoses and occlusions,[10-13] and conventional Doppler ultrasonography has been studied as well.[14-27] A Doppler ultrasound flow detector was used to identify the anatomical location of the temporal arteries preoperatively,[14] and such an approach can show that there are more stenoses and occlusions in patients with biopsy-proved temporal arteritis than in patients with negative histologic findings.[15,17] In three studies of 30 to 59 patients with temporal arteritis, Doppler ultrasonography identified abnormalities in 77 to 90 percent of the patients.[24,25,27] These findings can be compared with our results (stenosis or occlusion was identified by pulsed-wave Doppler ultrasonography in 80 percent of patients) (Fig. 3). In contrast to conventional Doppler ultrasonography, duplex ultrasonography can distinguish between aplasia and occlusion of the vessel because of the imaging capabilities of the B-mode ultrasound. The superficial temporal artery was found to have an abnormal course in 0 to 8 percent of cadavers.[8,28,29] Stenoses or occlusions may also occur in persons without temporal arteritis, although they are less common. Continuous-wave Doppler ultra-

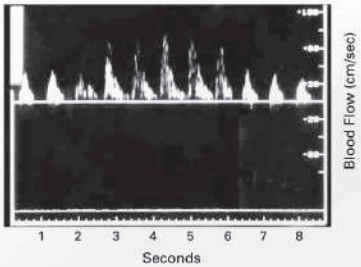


Figure 3. Stenosis of the Superficial Temporal Artery.

On the left-hand side, blood flow before the stenosis is normal. As the pulsed-wave Doppler beam moves through the stenosis, blood flow velocity increases. In the area to the right beyond the stenosis, blood flow velocity again decreases.

sonography identified abnormalities in 9 to 18 percent of control subjects[18,25] and in 20 to 24 percent of patients with negative histologic findings who were suspected of having temporal arteritis.[18,24,27] In our study, pulsed-wave Doppler ultrasonography also identified abnormalities in some of the patients without temporal arteritis (Table 2). The use of duplex ultrasonography can increase the sensitivity and specificity of this approach by its ability to visualize the halo around perfused lumina of inflamed arteries. Orbital duplex ultrasonography can demonstrate reduced blood flow, stenoses, and occlusions, especially in the central retinal and posterior ciliary artery, in patients with temporal arteritis.[30-35]

An investigation of the superficial temporal arteries by duplex ultrasonography requires a linear scanner with a high resolution in the area that is 3 to 10 mm below the surface of the skin. We tried several 7.5-MHz scanners and found that it was very difficult to visualize the artery adequately with them. In most cases it was difficult to find especially narrow rami and to be sure that a halo was actually present, since it is sometimes quite small.

There are reports of the use of histologic analysis of temporal arteries to diagnose Wegener's granulomatosis,[36,37] polyarteritis nodosa,[38] Churg–Strauss syndrome,[39] and primary amyloidosis.[40] Certainly, ultrasonography cannot be used to differentiate between those diagnoses and temporal arteritis, but such syndromes are clinically very rare in comparison with typical temporal arteritis.

We suggest that duplex ultrasonography of the temporal arteries can be used in the diagnosis of temporal arteritis and polymyalgia rheumatica as follows. Patients with typical clinical signs of temporal arteritis and a clear halo on ultrasonography might be treated without a biopsy, unless there is a reason to suspect another form of vasculitis. Patients with strong clinical evidence of temporal arteritis who have only stenoses or occlusions or no abnormalities on ultrasonography should still undergo biopsy. Patients with clinical signs of polymyalgia rheumatica who have no symptoms of temporal arteritis but who have abnormal findings on ultrasonography (a halo, stenosis, or occlusion) should undergo biopsy and be treated with a higher initial dose of corticosteroids to protect against blindness, at least until the biopsy results are known. In patients with equivocal clinical evidence of temporal arteritis, ultrasonography should be particularly helpful.

REFERENCES

1. Hunder GG, Bloch DA, Michel BA, et al. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. Arthritis Rheum 1990;33:1122-8.
2. Scott KR, Tse DT, Kronish JW. Temporal artery biopsy technique: a clinico-anatomical approach. Ophthalmic Surg 1991;22:519-25.
3. Siemssen SJ. On the occurrence of necrotising lesions in temporal arteritis: review of the literature with a note on the potential risk of a biopsy. Br J Plast Surg 1987;40:73-82.
4. Slavin ML. Brow droop after superficial temporal artery biopsy. Arch Ophthalmol 1986;104:1127.
5. Vollrath-Junger C, Gloor B. Warum eine Dopplersonographie vor jeder Biopsie der A. temporalis? Klin Monatsbl Augenheilkd 1989;195:169-71.
6. Schmidt WA, Kraft HE, Völker L, Vorpahl K, Gromnica-Ihle EJ. Colour Doppler sonography to diagnose temporal arteritis. Lancet 1995;345:866.
7. Bird HA, Esselinckx W, Dixon AS, Mowat AG, Wood PH. An evaluation of criteria for polymyalgia rheumatica. Ann Rheum Dis 1979;38:434-9.
8. Stock AL, Collins HP, Davidson TM. Anatomy of the superficial temporal artery. Head Neck Surg 1980;2:466-9.
9. Ashton-Key M, Gallagher PJ. Surgical pathology of cranial arteritis and polymyalgia rheumatica. Baillieres Clin Rheumatol 1991;5:387-404.
10. Gillanders LA, Strachan RW, Blair DW. Temporal arteriography: a new technique for the investigation of giant cell arteritis and polymyalgia rheumatica. Ann Rheum Dis 1969;28:267-9.
11. Hunder GG, Baker HL Jr, Rhoton AL Jr, Sheps SG, Ward LE. Superficial temporal arteriography in patients suspected of having temporal arteritis. Arthritis Rheum 1972;15:561-70.
12. Elliott PD, Baker HL Jr, Brown AL Jr. The superficial temporal artery angiogram. Radiology 1972;102:635-8.
13. Horwitz HM, Pepe PF, Johnsrude IS, McCoy RC, Jackson DC, Farmer JC. Temporal arteriography and immunofluorescence as diagnostic tools in temporal arteritis. J Rheumatol 1977;4:76-85.
14. Kelley JS. Doppler ultrasound flow detector used in temporal artery biopsy. Arch Ophthalmol 1978;96:845-6.
15. Dany F, Liozon F, Adenis JP, Michel JP, Vidal E, Kim M. Doppler et artérite temporale résultats préliminaires. J Mal Vasc 1981;6:273-7.
16. Menkes CJ, Branche I, Feldmann JL, Chauveau M, Delbarre F. Application de l'effet Doppler au dépistage de l'artérite temporale de Horton. Nouv Presse Med 1981;10:2371.
17. Barrier J, Potel G, Renaut-Hovasse H, et al. The use of Doppler flow studies in the diagnosis of giant cell arteries: selection of temporal artery biopsy site is facilitated. JAMA 1982;248:2158-9.
18. Liozon F, Dany F, Catanzano R. Exploration des artères temporales par effet Doppler dans la maladie de Horton. Presse Med 1983;12:2113-4.
19. Bienfang DC. Use of the Doppler probe to detect the course of the superficial temporal artery. Am J Ophthalmol 1984;97:526-7.
20. Liozon F, Weinbreck P, Vidal E, et al. Sténoses des artères des membres supérieurs dans l'artérite temporale de Horton. Ann Med Interne 1986;137:307-12.
21. Dörnberger V, Dörnberger G, Rössler R, Eggstein M. Dopplersonographie in der Diagnostik und Therapie der Arteriitis temporalis. Med Welt 1987;38:1302-8.

22. Brunhölzl C, Möller HR. Dopplersonographische Befunde bei der Arteriitis temporalis. Ultraschall Med 1988;9:232-6.
23. Friedman G, Friedman B, Beer G. Role of Doppler ultrasound flow studies in the diagnosis of temporal arteritis. Isr J Med Sci 1988;24:686-7.
24. Dany F, Liozon F, Chaumont P, et al. Sensibilité et spécificité du Doppler temporal dans la maladie de Horton. J Mal Vasc 1989;14:Suppl C:104-8.
25. Vinckier L, Hatron PY, Gadenne C, Catteau MH, Gosset D, Devulder B. Intérêt du Doppler artériel dans la maladie de Horton: étude prospective de 59 observations. J Mal Vasc 1989;14:Suppl C:116-22.
26. Beckman RL, Hartmann BM. The use of a Doppler flow meter to identify the course of the temporal artery. J Clin Neuroophthalmol 1990; 10:304.
27. Puéchal X, Chauveau M, Menkès CJ. Temporal Doppler-flow studies for suspected giant-cell arteritis. Lancet 1995;345:1437-8. [Erratum, Lancet 1995;346:62.]
28. Marano SR, Fischer DW, Gaines C, Sonntag VH. Anatomical study of the superficial temporal artery. Neurosurgery 1985;6:786-90.
29. Daumann C, Putz R, Schmidt D. Der Verlauf der Arteria temporalis superficialis: Anatomische Untersuchungen als Voraussetzung für eine Arterien-Biopsie. Klin Monatsbl Augenheilkd 1989;194:37-41.
30. Lieb WE, Flaherty PM, Ho A, Sergott RC. Color Doppler imaging of the eye and orbit: a synopsis of a 400 case experience. Acta Ophthalmol Suppl 1992;204:50-4.
31. Williamson TH, Baxter G, Paul R, Dutton GN. Colour Doppler ultrasound in the management of a case of cranial arteritis. Br J Ophthalmol 1992;76:690-1.
32. Aburn NS, Sergott RC. Orbital colour Doppler imaging. Eye 1993;7: 639-47.
33. Ho AC, Sergott RC, Regillo CD, et al. Color Doppler hemodynamics of giant cell arteritis. Arch Ophthalmol 1994;112:938-45.
34. Steigerwalt RD Jr, Cesarone MR, Laurora G, et al. Doppler ultrasonography in giant cell arteritis. Int Angiol 1994;13:286-9.
35. Ghanchi FD, Williamson TH, Lim CS, et al. Colour Doppler imaging in giant cell (temporal) arteritis: serial examination and comparison with non-arteritic anterior ischaemic optic neuropathy. Eye 1996;10:459-64.
36. Small P, Brisson ML. Wegener's granulomatosis presenting as temporal arteritis. Arthritis Rheum 1991;34:220-3.
37. Lie JT. When is arteritis of the temporal arteries not temporal arteritis? J Rheumatol 1994;21:186-9.
38. Morgan GJ Jr, Harris ED Jr. Non-giant cell arteritis: three cases and a review of the literature. Arthritis Rheum 1978;21:362-6.
39. Lie JT, Nagpal S. Churg-Strauss syndrome with nongiant cell eosinophilic temporal arteritis. J Rheumatol 1994;21:366-7.
40. Salvarani C, Gabriel SE, Gertz MA, Bjornsson J, Li CY, Hunder GG. Primary systemic amyloidosis presenting as giant cell arteritis and polymyalgia rheumatica. Arthritis Rheum 1994;37:1621-6.

APPENDIX 2

THE EFFECT OF FECAL OCCULT-BLOOD SCREENING ON THE INCIDENCE OF COLORECTAL CANCER

JACK S. MANDEL, PH.D., M.P.H., TIMOTHY R. CHURCH, PH.D., JOHN H. BOND, M.D., FRED EDERER, M.A., MINDY S. GEISSER, M.S., STEVEN J. MONGIN, M.S., DALE C. SNOVER, M.D., AND LEONARD M. SCHUMAN, M.D.

ABSTRACT

Background Both annual testing for fecal occult blood and biennial testing significantly reduce mortality from colorectal cancer. However, the effect of screening on the incidence of colorectal cancer remains uncertain, despite the diagnosis and removal of precancerous lesions in many persons who undergo screening.

Methods We have followed the participants in the Minnesota Colon Cancer Control Study for 18 years. A total of 46,551 people, most of whom were 50 to 80 years old, were enrolled between 1975 and 1978 and randomly assigned to annual screening, biennial screening, or usual care (the control group). Those assigned to the screening groups were asked to prepare and submit two samples from each of three consecutive stools for guaiac-based testing. Those with at least one positive slide in the set of six were offered a diagnostic examination that included colonoscopy. Screening was conducted between 1976 and 1982 and again between 1986 and 1992. Study participants have been followed with respect to newly diagnosed cases of colorectal cancer and deaths. Follow-up has been more than 90 percent complete.

Results During the 18-year follow-up period, we identified 1359 new cases of colorectal cancer: 417 in the annual-screening group, 435 in the biennial-screening group, and 507 in the control group. The cumulative incidence ratios for colorectal cancer in the screening groups as compared with the control group were 0.80 (95 percent confidence interval, 0.70 to 0.90) and 0.83 (95 percent confidence interval, 0.73 to 0.94) for the annual-screening and biennial-screening groups, respectively. For both screening groups, the number of positive slides was associated with the positive predictive value both for colorectal cancer and for adenomatous polyps at least 1 cm in diameter.

Conclusions The use of either annual or biennial fecal occult-blood testing significantly reduces the incidence of colorectal cancer. (N Engl J Med 2000; 343:1603-7.)

THREE randomized, controlled clinical trials have shown that both annual screening and biennial screening for occult blood in the stool significantly reduce the rate of death from colorectal cancer.[1-4] Several observational studies have had similar results.[5-13] The reduction in mortality is evidently a consequence of the earlier detection and surgical removal of malignant colorectal tumors. Fecal occult-blood screening may yield an additional benefit: a reduction in the incidence of colorectal cancer, resulting from the detection and removal of premalignant adenomatous polyps. Fecal occult-blood tests are not very sensitive for the presence of polyps but will detect some, particularly large polyps, which are more likely to bleed[14-16] and also to become cancerous.[17]

Subjects in the National Polyp Study had a lower incidence of colorectal cancer than subjects in three historical control groups, which suggests that the incidence of this cancer is reduced by colonoscopic polypectomy and surveillance of persons with adenomatous polyps.[18] In the trials of fecal occult-blood screening, many polyps were detected and removed during bowel examination after a positive screening test. Thus, a reduction in the incidence of colorectal cancer in the screened groups might be expected. After 13 years of follow-up in the Minnesota Colon Cancer Control Study, the incidence of colorectal cancer was 12 percent lower in the screened groups than in the control groups (P not significant).[1] We now present data on the cumulative incidence of colorectal cancer after 18 years of follow-up.

METHODS

Participants

The Minnesota study was a randomized, controlled clinical trial conducted to evaluate the effectiveness of fecal occult-blood testing in reducing the rate of death from colorectal cancer.[1,19] Briefly, 46,551 people, the majority of whom were 50 to 80 years old when recruited between 1975 and 1978, were randomly assigned to annual screening, biennial screening, or usual care (control group). The annual-screening and biennial-screening groups will be referred to collectively as the screening groups. For each screening, participants assigned to the screening groups were asked to prepare two guaiac-impregnated paper slides (Hemoccult, Beckman Coulter, Palo Alto, Calif.) for each of three consecutive stools and submit them to the investigators. Screening was conducted between 1976 and 1982 and, after a hiatus, resumed in 1986. All screening was completed in 1992.

The institutional review board of the University of Minnesota reviewed and approved the study, and all participants gave written informed consent.

Study Design

Participants with at least one positive slide were invited to undergo a diagnostic evaluation, which initially included history taking and a physical examination, single-column barium enema, rig-

From Exponent, Menlo Park, Calif. (J.S.M.); the Divisions of Environmental and Occupational Health (T.R.C., M.S.G., S.J.M.), Biostatistics (F.E.), and Epidemiology (L.M.S.), School of Public Health, University of Minnesota, Minneapolis; the Minneapolis Veterans Affairs Hospital, Minneapolis (J.H.B.); the Emmes Corp., Potomac, Md. (F.E.); and Fairview Hospital, Minneapolis (D.C.S.). Address reprint requests to Dr. Mandel at Exponent, 149 Commonwealth Dr., Menlo Park, CA 94025, or at jmandel@exponent.com.

id proctosigmoidoscopy, urinalysis, a complete blood count, routine tests of blood chemistry, an upper gastrointestinal x-ray series, chest radiography, electrocardiography, and colonoscopy. Single-column barium enema was discontinued in 1978. Double-contrast barium enemas were administered to participants who had at least one positive screening test and whose colonoscopic examinations were incomplete or suboptimal (about 5 percent of the subjects). Rigid proctosigmoidoscopy and the upper gastrointestinal series were discontinued in 1982. Throughout the trial, colonoscopy was the dominant diagnostic procedure used to identify colorectal polyps and cancers. During colonoscopy, all polyps detected were routinely resected.

Diagnostic evaluation of subjects with positive screening tests was completed in 1993. Follow-up has continued to the present time in order to ascertain the incidence of colorectal cancer and all deaths.

Beginning in 1976 and continuing to the present, all study participants received an annual questionnaire designed to ascertain vital status and identify newly diagnosed colorectal polyps and cancers in the control group and in members of the screening groups whose lesions were not detected by screening. Colorectal cancers reported in all three groups were confirmed by a review of relevant medical records from the diagnosing or treating physicians or hospitals, with the written consent of the participants. For colorectal cancers and for polyps reported in conjunction with colorectal cancers, slides of tissue were obtained for independent confirmation and staging by the study pathologist.

To validate the method of ascertainment, new cases of colorectal cancer were also identified by linkage of the cohort, from 1988 (the first year for which data were available) through 1994 (the last year for which data were available), to the Minnesota Cancer Surveillance System,[20] which collects data on all newly diagnosed and pathologically confirmed cancers in Minnesota.

For deaths occurring in the first 15 years after enrollment in the study, a deaths review committee, whose members had no knowledge of the group assignments, reviewed all records of selected subjects who had died to determine the underlying cause of death. Included in this review were all such subjects in whom a history of gastrointestinal disease or cancer could not be definitively ruled out.[21]

Statistical Analysis

For the analysis of incidence, which included all cases diagnosed in the first 18 years of the study that were logged into the data base before January 1, 1999, the cumulative probability of survival free of colorectal cancer was estimated by the life-table method.[22] The cumulative incidence was obtained by subtracting this probability from 1. The 106 subjects in whom colorectal cancer was diagnosed before randomization were omitted from the analysis of incidence, leaving 46,445 subjects. The ratios of the 18-year cumulative incidence of colorectal cancer in the two screening groups to the incidence in the control group and the 95 percent confidence intervals for these ratios were computed as a measure of the extent to which screening affected incidence. Reported P values are two-sided, and no correction was made for multiple tests.

To evaluate the sensitivity of the fecal occult-blood test for the detection of large polyps, we used the following indirect method.[23] For all returned slide sets that contained all six specimens, the positive predictive value (for colorectal cancer, the fraction of subjects with a positive screening test who had colorectal cancer; for polyps, the fraction who had polyps) was estimated for each given number of positive slides. The positive predictive value was calculated in the screening groups separately and for each of two outcomes: colorectal cancer or polyps at least 1 cm in diameter but no colorectal cancer. Only positive tests that were followed by adequate examinations, defined as colonoscopy or a combination of flexible sigmoidoscopy and barium-enema radiography, were used in the computations.

To account for the potential correlation between repeated screenings in the same subject, we estimated the positive predictive value of each number of positive slides and for each screening group in a single, generalized linear model with a logit-link function, using the generalized-estimating-equation approach.[24-27] Accounting for correlation in this manner corrects for the excessive weight that would be given to participants with multiple screenings if the positive predictive value were estimated as the simple ratio of the number of true positives to the total number of positive screening tests. Spearman's rho test[28,29] was used to measure the size and statistical significance of the association between the number of positive slides and the positive predictive value.

RESULTS

A total of 46,445 people (22,323 men and 24,122 women) without a prior diagnosis of colorectal cancer were randomly assigned to one of the three study groups (Table 1). During the screening period, participants in the annual-screening and biennial-screening groups were offered 11 and 6 opportunities for screening, respectively. The mean rate of compliance per screening opportunity was 75 percent in the annual-screening group and 78 percent in the biennial-screening group. Of the subjects with a positive test, 83 percent in the annual-screening group and 84 percent in the biennial-screening group underwent diagnostic follow-up, including a complete examination of the large bowel by colonoscopy or the combination of double-contrast barium enema and flexible sigmoidoscopy. In each group, about 11 percent of the subjects with positive screening tests underwent flexible sigmoidoscopy or barium enema or underwent another fecal occult-blood test. Five percent of the subjects with positive tests declined to consult a physician. Overall, 75 percent of the subjects with positive screening tests were examined at the University of Minnesota Hospital and Clinics.

Follow-up for vital status through year 18 was complete for 91.3 percent, 91.7 percent, and 91.2 percent of participants in the annual-screening, biennial-screening, and control groups, respectively, and it was 95 percent complete through year 17 for all groups. Death certificates were obtained for all but 3 of the 15,873 people who were known to have died during the 18-year follow-up period. Because active surveillance has continued since the cessation of screening, including a successful linkage with the Minnesota Cancer Surveillance System (performed without knowledge of group assignment), the ascertainment of the incidence of colorectal cancer was nearly complete and did not vary significantly between the groups.

During the 18 years of follow-up, approximately 235,000 person-years were accrued for each study group, and 1359 new cases of colorectal cancer were ascertained: 417 in the annual-screening group, 435 in the biennial-screening group, and 507 in the control group (Table 1 and Fig. 1). The ratios of the cumulative incidence rates in the screening groups to that in the control group were 0.80 (95 percent confidence interval, 0.70 to 0.90; $P<0.001$) for the annual-screening group and 0.83 (95 percent confidence interval, 0.73 to 0.94; $P=0.002$) in the biennial-screening group (Table 1).

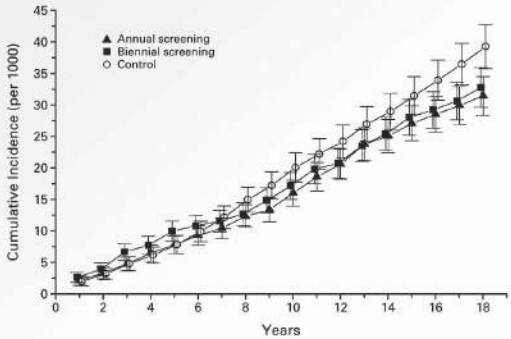
Table 1. Results of Fecal Occult-Blood Testing According to Study Group after 18 Years of Follow-up.*

Variable	Annual Screening	Biennial Screening	Control
Subjects — no.			
Total	15,532	15,550	15,363
Men	7,474	7,430	7,419
Women	8,058	8,120	7,944
Person-years of observation — no.	235,584	235,513	232,612
Cases of colorectal cancer — no.	417	435	507
Cumulative incidence/1000	32 (29–35)	33 (30–36)	39 (36–43)
Cumulative incidence ratio†	0.80 (0.70–0.90)	0.83 (0.73–0.94)	1.0
P value‡	<0.001	0.002	—

*Values in parentheses are 95 percent confidence intervals.

†Values are the ratios of cumulative incidence in each screening group to that in the control group.

‡P values are for the comparison with the control group.



No. of Subjects										
Annual screening	15,532	15,364	14,991	14,538	14,022	13,451	12,763	11,983	11,157	9494
Biennial screening	15,550	15,386	14,984	14,507	13,989	13,407	12,716	11,988	11,155	9554
Control	15,363	15,202	14,835	14,382	13,851	13,218	12,598	11,827	11,021	9342

Figure 1. Cumulative Incidence of Colorectal Cancer, According to Study Group, during 18 Years of Follow-up.

Table 2 shows the positive predictive value of a positive test, according to the number of positive slides, for colorectal cancer and for adenomatous polyps at least 1 cm in diameter but no colorectal cancer. For both screening groups, there was a strong association between the number of positive slides and the percentage of subjects with colorectal cancer. For the annual-screening group, the positive predictive value ranged from 0.87 percent for one positive slide to 4.53 percent for six positive slides (Spearman's rho=0.94, P=0.02). For the biennial-screening group, the positive predictive value ranged from 1.12 percent for one positive slide to 6.13 percent for six positive slides (Spearman's rho=0.94, P=0.02). For adenomatous polyps at least 1 cm in diameter, there was also an association between the number of positive

Table 2. Predictive Value of Fecal Occult-Blood Testing According to the Number of Positive Slides.*

No. of Positive Slides in Set of Six	No. of Screening Tests	Positive Predictive Value for Colorectal Cancer	Positive Predictive Value for Adenomatous Polyps ≥1 cm
		percent (95 percent confidence interval)	
Annual-screening group			
1	2874	0.87 (0.59–1.29)	5.99 (5.18–6.92)
2	2448	1.55 (1.13–2.13)	6.08 (5.18–7.13)
3	1118	1.16 (0.68–1.99)	7.03 (5.69–8.66)
4	883	1.93 (1.20–3.07)	7.49 (5.92–9.43)
5	368	2.72 (1.47–4.97)	8.25 (5.85–11.50)
6	972	4.53 (3.38–6.04)	7.87 (6.29–9.79)
Biennial-screening group			
1	1695	1.12 (0.72–1.75)	6.86 (5.74–8.16)
2	1412	1.13 (0.70–1.84)	8.07 (6.73–9.64)
3	632	2.69 (1.68–4.29)	9.96 (7.79–12.65)
4	526	2.28 (1.30–3.97)	10.87 (8.47–13.86)
5	220	5.00 (2.79–8.82)	10.99 (7.49–15.84)
6	685	6.13 (4.56–8.20)	10.08 (8.03–12.59)

*Only positive tests that were followed by adequate examinations, defined as colonoscopy or a combination of flexible sigmoidoscopy and barium-enema radiography, were used in the computations.

slides and the positive predictive value; this value ranged from 5.99 percent to 7.87 percent for the annual-screening group (Spearman's rho=0.94, P=0.02) and from 6.86 percent to 10.08 percent for the biennial-screening group (Spearman's rho=0.83, P=0.06).

DISCUSSION

Our findings demonstrate a significant reduction in the incidence of colorectal cancer after fecal occult-blood testing. This reduction occurred after both annual and biennial screening. The most plausible explanation is the identification and removal of the precursor lesions for colorectal cancer — that is, adenomatous polyps. In study participants who were examined at the University of Minnesota, polyps visualized during the examination were routinely removed. For patients with a positive slide who were examined elsewhere, the information on whether polyps were diagnosed and removed during follow-up was complete. However, the decision to remove polyps was made less consistently than at the University of Minnesota, especially during the early years of the study. Since the focus of the study was on colorectal cancer, information was abstracted only on polyps that were removed in the course of a diagnosis of cancer or a workup after a positive fecal occult-blood test. Thus, the attribution of the reduced incidence of colorectal cancer to the removal of adenomatous polyps is based on the not unreasonable assumption that the removal of polyps was much more common in the screening groups than in the control group. The National Polyp Study found that the removal of adenomatous polyps, followed by colonoscopic surveillance, significantly reduced the incidence of colorectal cancer below the rates in the general population.[18] The sensitivity of fecal occult-blood tests for the detection of adenomatous polyps has been reported to be moderate but is greater for larger polyps, which are more likely to bleed, than for smaller polyps.[14-17] The positive predictive value increased with the number of positive slides in our study, which further supports the results of earlier studies[14-17]; our findings also provide evidence that polyp removal is the likely explanation for the decrease in incidence.

The failure of screening for fecal occult blood to lead to a clinically significant reduction in the incidence of colorectal cancer after early follow-up in some studies might be interpreted as an argument against the hypothesis that most cancers arise from benign adenomatous polyps.[30,31] Our study supports the theory of the adenoma–carcinoma sequence and emphasizes the importance of detecting and resecting advanced adenomas. To date, however, the European trials have involved less follow-up, making it difficult to show a reduction in incidence with screening. Lang and Ransohoff argued that the reduction in mortality from colorectal cancer in our trial was largely the consequence of chance detection of cancers as a result of many colonoscopies performed, rather than a result of the sensitivity of the fecal occult-blood test.[32] However, the reverse has been shown to be true; that is, the reduced mortality resulted primarily from the sensitivity of the fecal occult-blood test, with chance detection of cancer playing only a minor part.[33] It may similarly be argued that the reduced incidence of colorectal cancer in our trial is largely the result of the chance detection of nonbleeding adenomatous polyps by the many colonoscopies performed, rather than a result of the sensitivity of the fecal occult-blood test for the detection of bleeding polyps. However, the fecal occult-blood test was sensitive for larger polyps, as shown by the statistically significant association between the positive predictive value of a test and the number of positive slides.

The reductions in the incidence of colorectal cancer that we observed may underestimate the true reductions that could be achieved among persons who elect to participate fully in screening. Compliance with screening in our trial was less than complete, averaging about 75 percent for each screening occasion. About 50 percent of the screening subjects participated in all the screening tests they were offered; 10 percent participated in none.[1] In addition, subjects in the control group were not prevented

from undergoing screening through their personal physicians. Thus, noncompliance in the screening groups and participation in screening in the control group are likely to have attenuated the true effect. A hiatus in the screening program, which averaged 4.5 years for the annual-screening group and 3.6 years for the biennial-screening group, is likely to have further attenuated the effect.[4] Although subjects in the screening groups may have participated in some screening beyond that offered in the study, the data that were collected indicate that this occurred infrequently (data not shown). Finally, screening advances the date of diagnosis and thus leads to a temporary, artifactual increase in incidence. After screening ceases, this effect is diminished.

The use of either annual or biennial fecal occult-blood testing significantly reduces the incidence of colorectal cancer. Preventing colorectal cancers reduces morbidity and is likely to improve the cost effectiveness of screening for colorectal cancer by fecal occult-blood testing.

Supported by research contracts (N01-CB-95613, N01-CB-61005, N01-CB-53862, and R01CA65728) with the National Cancer Institute and by a cooperative agreement with the Centers for Disease Control and Prevention through the Association of Schools of Public Health.

We are indebted to G. Mary Bradley, D. Engelhard, J. Cordes, and G. Watt for their contributions to this study.

REFERENCES

1. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. N Engl J Med 1993; 328:1365-71. [Erratum, N Engl J Med 1993;329:672.]
2. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472-7.
3. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet 1996;348:1467-71.
4. Mandel JS, Church TR, Ederer F, Bond JH. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 1999;91:434-7.
5. Zappa M, Castiglione G, Grazzini G, et al. Effect of faecal occult blood testing on colorectal mortality: results of a population-based case-control study in the district of Florence, Italy. Int J Cancer 1997;73:208-10.
6. Bertario L, Russo A, Crosignani P, et al. Reducing colorectal cancer mortality by repeated faecal occult blood test: a nested case-control study. Eur J Cancer 1999;35:973-7.
7. Eddy DM, Nugent FW, Eddy JR, et al. Screening for colorectal cancer in a high-risk population: results of a mathematical model. Gastroenterology 1987;92:682-92.
8. Faivre J, Tazi MA, El Mrini T, Lejeune C, Benhamiche AM, Dassonville F. Faecal occult blood screening and reduction of colorectal cancer mortality: a case-control study. Br J Cancer 1999;79:680-3.
9. Lazovich D, Weiss NS, Stevens NG, White E, McKnight B, Wagner EH. A case-control study to evaluate efficacy of screening for faecal occult blood. J Med Screen 1995;2:84-9.
10. Nakama H. A study on the efficacy of a screening program for colorectal cancer in a small Japanese village. Clin Invest 1994;72:117-21.
11. Winawer SJ, Flehinger BJ, Schottenfeld D, Miller DG. Screening for colorectal cancer with fecal occult blood testing and sigmoidoscopy. J Natl Cancer Inst 1993;85:1311-8.
12. Wahrendorf J, Robra BP, Wiebelt H, Oberhausen R, Weiland M, Dhom G. Effectiveness of colorectal cancer screening: results from a population-based case-control evaluation in Saarland, Germany. Eur J Cancer Prev 1993;2:221-7.
13. Selby JV, Friedman GD, Quesenberry CP Jr, Weiss NS. Effect of fecal occult blood testing on mortality from colorectal cancer: a case-control study. Ann Intern Med 1993;118:1-6.
14. Zhou DY, Feng FC, Zhang YL, et al. Comparison of Shams' test for rectal mucus to an immunological test for fecal occult blood in large intestinal carcinoma screening: analysis of a check-up of 6480 asymptomatic subjects. Chin Med J (Engl) 1993;106:739-42.
15. Hope RL, Chu G, Hope AH, Newcombe RG, Gillespie PE, Williams SJ. Comparison of three faecal occult blood tests in the detection of colorectal neoplasia. Gut 1996;39:722-5.
16. St John DJ, Young GP, Alexeyeff MA, et al. Evaluation of new occult blood tests for detection of colorectal neoplasia. Gastroenterology 1993; 104:1661-8.
17. Demers RY, Stawick LE, Demers P. Relative sensitivity of the fecal occult blood test and flexible sigmoidoscopy in detecting polyps. Prev Med 1985;14:55-62.
18. Winawer SJ, Zauber AG, O'Brien MJ, et al. Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. N Engl J Med 1993;328:901-6.
19. Gilbertsen VA, Church TR, Grewe FJ, et al. The design of a study to assess occult-blood screening for colon cancer. J Chronic Dis 1980;33:107-14.
20. Bushhouse S, Punyko J, Soler J, Cords J. The occurrence of cancer in Minnesota, 1988-1996: incidence, mortality, and trends. Minneapolis: Minnesota Department of Health, 1999.
21. Ederer F, Geisser MS, Mongin SJ, Church TR, Mandel JS. Colorectal cancer deaths as determined by expert committee and from death certificate: a comparison: the Minnesota Study. J Clin Epidemiol 1999;52:447-52.
22. Cutler SJ, Ederer F. Maximum utilization of the life table method in analyzing survival. J Chronic Dis 1958;8:699-712.
23. Church TR, Ederer F, Mandel JS. Fecal occult blood screening in the Minnesota study: sensitivity of the screening test. J Natl Cancer Inst 1997; 89:1440-8.
24. Zeger S, Liang K-Y. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988;44:1049-60. [Erratum, Biometrics 1989;45:347.]
25. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13-22.
26. SAS/STAT software: changes and enhancements through release 6.12. Cary, N.C.: SAS Institute, 1997.
27. Data Analysis Products Division. S-PLUS 5 for UNIX guide to statistics. Seattle: MathSoft, September 1998.
28. Conover WJ. Practical nonparametric statistics. 2nd ed. New York: John Wiley, 1980.
29. Lehmann EL. Nonparametrics: statistical methods based on ranks. Oakland, Calif.: Holden-Day, 1975.
30. Ahlquist DA. Fecal occult blood testing for colorectal cancer: can we afford to do this? Gastroenterol Clin North Am 1997;26:41-55.
31. Simon JB. Should all people over the age of 50 have regular fecal occult-blood tests? N Engl J Med 1998;338:1151-2.
32. Lang CA, Ransohoff DF. Fecal occult blood screening for colorectal cancer: is mortality reduced by chance selection for screening colonoscopy? JAMA 1994;271:1011-3.
33. Ederer F, Church TR, Mandel JS. Fecal occult blood screening in the Minnesota study: role of chance detection of lesions. J Natl Cancer Inst 1997;89:1423-8.

APPENDIX 3

The New England
Journal of Medicine

VOLUME 339 DECEMBER 24, 1998 NUMBER 26

SYMPTOMATIC BENEFIT FROM ERADICATING *HELICOBACTER PYLORI* INFECTION IN PATIENTS WITH NONULCER DYSPEPSIA

KENNETH MCCOLL, M.D., LILIAN MURRAY, PH.D., EMAD EL-OMAR, M.D., ANTHEA DICKSON, B.A., ADIL EL-NUJUMI, M.D., ANGELA WIRZ, B.SC., ANDREW KELMAN, PH.D., CHRISTINE PENNY, M.B., CH.B., ROBIN KNILL-JONES, M.D., AND THOMAS HILDITCH, PH.D.

ABSTRACT

Background The eradication of *Helicobacter pylori* infection is beneficial in patients with gastric or duodenal ulcers. The value of eradicating the infection in patients with dyspepsia and no evidence of ulcer disease is not known.

Methods We performed a randomized, placebo-controlled trial comparing the efficacy of treatment for two weeks with 20 mg of omeprazole orally twice daily, 500 mg of amoxicillin three times daily (with 500 mg of tetracycline three times daily substituted for amoxicillin in patients allergic to penicillin), and 400 mg of metronidazole three times daily (160 patients) with that of omeprazole alone (158 patients) for resolving symptoms of dyspepsia in patients with *H. pylori* infection but no evidence of ulcer disease on upper gastrointestinal endoscopy. Symptoms were assessed with the Glasgow Dyspepsia Severity Score, with resolution of symptoms defined as a score of 0 or 1 in the preceding six months (maximal score, 20). One year later the patients were assessed to determine the frequency of the resolution of symptoms.

Results One month after the completion of treatment, 132 of 150 patients (88 percent) in the group assigned to receive omeprazole and antibiotics had a negative test for *H. pylori*, as compared with 7 of 152 (5 percent) in the group assigned to receive omeprazole alone. One year later, dyspepsia had resolved in 33 of 154 patients (21 percent) in the group given omeprazole and antibiotics, as compared with 11 of 154 (7 percent) in the group given omeprazole alone (95 percent confidence interval for the difference, 7 to 22 percent; $P<0.001$). Among the patients in the group given omeprazole and antibiotics, the symptoms resolved in 26 of the 98 patients (27 percent) who had had symptoms for five years or less, as compared with 7 of the 56 patients (12 percent) who had had symptoms for more than five years ($P=0.03$).

Conclusions In patients with *H. pylori* infection and nonulcer, or functional, dyspepsia, treatment with omeprazole and antibiotics to eradicate the infection is more likely to resolve symptoms than treatment with omeprazole alone. (N Engl J Med 1998;339:1869-74.)

DYSPEPSIA affects 20 to 40 percent of the population of the Western world.[1-4] Upper gastrointestinal endoscopy of patients with dyspepsia reveals no abnormality in the majority; such patients are considered to have nonulcer, or functional, dyspepsia.[5] The cause of nonulcer dyspepsia is unclear but is thought to be heterogeneous.[6-8] The management of the disorder is unsatisfactory, and there are few studies showing that active treatment is superior to a placebo.[7]

The recognition of the pathogenic role of *Helicobacter pylori* infection in patients with peptic ulcer and the benefits of eradicating the infection[9-11] has led to suggestions that the infection may also be the cause of dyspepsia in some patients with nonulcer dyspepsia. A higher prevalence of *H. pylori* infection has been reported in patients with dyspepsia than in those without it, but in some of these studies the patients were not matched for age or excluded if they had underlying ulcer disease.[12-14] Several studies have examined the effect of eradicating *H. pylori* infection on dyspeptic symptoms in patients with nonulcer dyspepsia, but the results have been conflicting and all such studies have been criticized because of design flaws.[15] In a recent review of these trials, Talley and Hunt concluded that "*H. pylori* has not been established, as yet, to play a definitive role in dyspepsia."[16]

We report the results of a large, randomized, placebo-controlled trial that assessed the effect of eradicating *H. pylori* in patients with nonulcer dyspepsia.

From the Departments of Medicine and Therapeutics (K.M., L.M., E.E.-O., A.D., A.E.-N., A.W., A.K.) and Clinical Physics (T.H.), Western Infirmary; Victoria Infirmary (C.P.); and the Department of Public Health, University of Glasgow (R.K.-J.) — all in Glasgow, United Kingdom. Address reprint requests to Dr. McColl at the Department of Medicine and Therapeutics, Western Infirmary, Glasgow G11 6NT, United Kingdom.

METHODS

Recruitment and Evaluation of Patients

We studied patients who were referred to one dyspepsia clinic by their primary care physician because they had had dyspepsia for at least four months. All patients were found to have *H. pylori* infection but no endoscopic evidence of current or previous peptic ulcer disease. Dyspepsia was defined as intermittent or persistent pain or discomfort in the upper abdomen or lower part of the chest, heartburn, nausea, a feeling of postprandial fullness, or any other symptoms thought to be related to the upper gastrointestinal tract.[16] Patients were excluded if they had previously been found to have peptic ulcer disease, had endoscopic evidence of esophagitis, were taking nonsteroidal antiinflammatory drugs (other than low-dose aspirin), had undergone gastric resection, were pregnant, or had previously been treated for *H. pylori* infection.

The patients had been asked to stop taking any antisecretory drug before the initial clinic visit. At this visit, a standardized interview was used to determine each patient's symptoms and the duration of symptoms (less than 2 years, 2 to 5 years, more than 5 to 10 years, or more than 10 years) and a physical examination was conducted.

The severity of the dyspepsia during the six months preceding the visit was assessed with the Glasgow Dyspepsia Severity Score.[17] This scale evaluates the frequency of symptoms (maximal score, 5); the effect of dyspepsia on normal activities (2); the number of days of work missed because of dyspepsia (2); and the frequency of medical consultations (2), home visits by a physician (2), tests for dyspepsia (2), use of over-the-counter medications (2), and use of prescribed medications (3). Scores can range from 0 to 20, with higher scores indicating more severe dyspepsia. The quality of life was assessed with the 36-item Medical Outcomes Study Short-Form General Health Survey (SF-36), which examines eight aspects of the quality of life: general and mental health, physical function, social function, physical and emotional health, pain, and vitality. Scores on each of the eight aspects can range from 0 (worst) to 100 (best).[18]

After the clinical assessment, the patients underwent a carbon-14 urea breath test to determine their *H. pylori* status.[19] The results of the test were considered positive if the value at 20 minutes was more than 30 (units equal the percentage of the dose administered per millimole of carbon dioxide expired times the body weight in kilograms times 100). A blood sample was also taken to determine whether IgG antibodies to the *CAGA* gene product of *H. pylori* were present (test kindly performed by Dr. Jean Crabtree). The presence of these antibodies indicates that patients are infected with a more virulent strain of the bacterium.[20]

Two weeks later the patients underwent upper gastrointestinal endoscopy. Before endoscopy, those with a positive urea breath test were invited to undergo randomization to either treatment to eradicate *H. pylori* or placebo treatment if the endoscopy showed no evidence of ulcer disease or esophagitis and to undergo studies of gastric acid secretion after the endoscopic examination. The randomization was carried out independently by the pharmacy department, which used tables of random numbers to assign half the patients to receive active treatment with omeprazole, amoxicillin, and metronidazole and half to receive omeprazole and placebo. The active and placebo tablets for each antibiotic were identical in appearance, and the trial was conducted in a double-blind fashion. The study was approved by the Western Infirmary ethics committee, and all patients gave written informed consent.

During the upper gastrointestinal endoscopy, biopsy samples were obtained from the antrum and body of the stomach. The samples were sectioned and stained with hematoxylin and eosin and examined for *H. pylori*. They were also examined for bacterial urease with a urease slide test (CLO test, Delta West, Bentley, Australia). The patients who had agreed to undergo acid secretory studies had a nasogastric tube inserted immediately after the removal of the endoscope, and they were then taken to the recovery area and allowed to rest for one hour. Acid secretion was then measured first in response to a one-hour intravenous infusion of gastrin-releasing peptide (Calbiochem Novabiochem, Nottingham, United Kingdom) at a dose of 40 pmol per kilogram of body weight per hour and then in response to a one-hour intravenous infusion of pentagastrin (Zeneca, Macclesfield, United Kingdom) at a dose of 0.6 μg per kilogram per hour.

After these tests, the patients received a two-week course of either 20 mg of omeprazole (Losec, Astra Hässle, Mölndal, Sweden) twice daily, 500 mg of amoxicillin three times daily, and 400 mg of metronidazole three times daily or 20 mg of omeprazole twice daily and placebo antibiotics. Patients with a history of allergy to penicillin were given 500 mg of tetracycline three times daily rather than amoxicillin.

The patients were given an appointment for a urea breath test and an assessment of side effects four weeks after treatment was completed and an appointment for a full reassessment one year after the start of treatment. They were also told that they could take any medication necessary, including a proton-pump inhibitor, if they had recurrent or persistent symptoms after the completion of treatment. However, their doctors were requested by letter not to prescribe any treatment to eradicate *H. pylori*. At the one-year visit, dyspeptic symptoms during the preceding six months were assessed with the Glasgow Dyspepsia Severity Score, the quality of life was assessed with the SF-36 questionnaire, and *H. pylori* status was assessed with a urea breath test. The results of the breath tests performed one month after the completion of treatment and one year after the start of treatment were not divulged to the patients or investigators until after the completion of the one-year assessment.

Statistical Analysis

The main end point was the resolution of symptoms, defined as a score of 0 or 1 on the Glasgow Dyspepsia Severity Score, because validation studies indicated that this cutoff point was the most appropriate indication of the resolution of dyspepsia.[17] The analysis was based on the intention-to-treat principle. Two-sample t-tests were used to compare the mean values of variables considered continuous in the active-treatment and placebo groups. Chi-square tests were used to analyze categorical variables. A logistic-regression analysis was carried out to determine which characteristics of the patients were predictive of the resolution of symptoms. All statistical tests were two-sided.

RESULTS

Between October 1994 and October 1996, we assessed 916 patients with dyspepsia, of whom 330 were enrolled in the study. The reasons for exclusion were endoscopic evidence of duodenal or gastric ulcer or esophagitis (209 patients), a negative test for *H. pylori* (306 patients), previous treatment for *H. pylori* (21 patients), inability to tolerate endoscopy (11 patients), use of nonsteroidal antiinflammatory drugs (9 patients), and pregnancy or a serious medical condition (21 patients). Of the eligible patients, only nine declined to participate.

Of the 330 patients who enrolled in the study, 12 were excluded from the final analysis. Eleven did not have at least two positive tests[21] for *H. pylori* infection, because they had entered the study immediately after the endoscopic examination when the results of the urea breath test were available but the results of the urease test and histologic analysis were not available; these subsequently proved to be negative. One patient did not meet the criteria for nonulcer dyspepsia because his base-line Glasgow Dyspepsia Severity Score in the six months preceding the study was 0.

Of the remaining 318 patients, 160 were assigned

to receive therapy to eradicate *H. pylori* (omeprazole and antibiotics) and 158 were assigned to receive omeprazole alone. The two groups were well matched with respect to age, sex, initial results on the Glasgow Dyspepsia Severity Score, prevalence of previously prescribed antisecretory-drug therapy, prevalence of smoking, predominant symptom at presentation, duration of symptoms, and quality of life (Table 1).

A total of 308 patients (97 percent) were reassessed one year after randomization, 154 in each group. Of the 154 patients in the group assigned to receive omeprazole and antibiotics, 150 had a urea breath test one month after the completion of treatment. The test was negative in 132 patients (88 percent), of whom 125 (95 percent) also had a negative test at one year. Of the 154 patients in the group assigned to receive omeprazole alone, 152 had a urea breath test one month after the end of treatment. The test was positive in 145 patients (95 percent), of whom 134 (92 percent) also had a positive test at one year.

The distribution of Glasgow Dyspepsia Severity Scores in the two groups at one year is shown in Figure 1. On an intention-to-treat basis, the symptoms resolved (defined as a score of 0 or 1) in 33 of 154 patients (21 percent) in the group assigned to receive omeprazole and antibiotics, as compared with 11 of 154 patients (7 percent) in the group assigned to receive omeprazole alone (95 percent confidence interval for the difference between groups, 7 to 22 percent; $P<0.001$).

The mean (±SD) dyspepsia score at one year was 5.4±4.0 in the group treated with omeprazole and antibiotics, as compared with 6.2±3.6 in the group treated with omeprazole alone ($P=0.07$) (Table 2). The scores in both groups were lower than those at

TABLE 1. BASE-LINE CHARACTERISTICS OF 318 PATIENTS WITH NONULCER DYSPEPSIA.*

CHARACTERISTIC	OMEPRAZOLE AND ANTIBIOTICS (N=160)	OMEPRAZOLE ALONE (N=158)
Male sex (%)	51	47
Age (yr)		
Mean ±SD	42±12	42.2±13
Range	18–70	17–70
Glasgow Dyspepsia Severity Score†		
Mean ±SD	11.4±2.2	11.5±2.5
Range	6–17	4–18
Quality-of-life score‡		
Mean ±SD	520±176	507±164
Range	25–786	93–764
Duration of symptoms (%)§		
<2 yr	39	38
2–5 yr	25	21
>5–10 yr	11	16
>10 yr	24	25
Predominant symptom (%)		
Epigastric pain	57	54
Retrosternal pain	21	16
Reflux	12	13
Other	10	17
History of smoking in past 6 mo (%)	34	33
Medication prescribed in past 6 mo (%)	84	80
Histamine H_2-receptor antagonists	59	54
Proton-pump inhibitors	13	11
Both	8	11
Other	4	4
Medication prescribed for >3 of the past 6 mo (%)	44	45

*There were no significant differences between the two groups.

†Scores can range from 0 (no symptoms or limitations) to 20 (severe symptoms most days, requiring regular medication).

‡Quality of life was measured with the SF-36; scores can range from 0 (worst) to 800 (best).

§Because of rounding, percentages may not total 100.

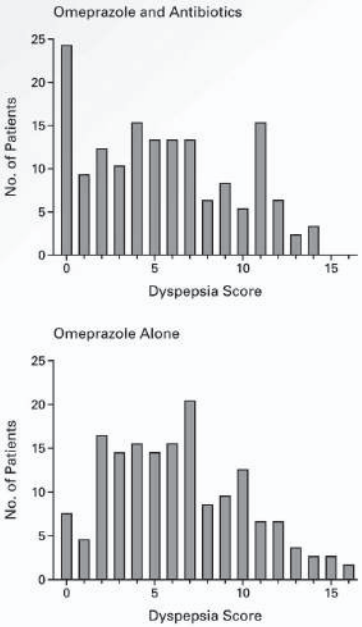


Figure 1. Distribution of Glasgow Dyspepsia Severity Scores One Year after the Start of Treatment.

The resolution of symptoms (defined as a score of 0 or 1) was significantly more common in the group assigned to receive omeprazole and antibiotics ($P<0.001$). Data were missing for one patient in the omeprazole-alone group.

Table 2. Outcomes of the Patients at One Year.*

Characteristic	Omeprazole and Antibiotics (N=154)	Omeprazole Alone (N=154)	P Value
Resolution of symptoms (%)	21	7	<0.001
Glasgow Dyspepsia Severity Score†			
Mean ±SD	5.4±4.0	6.2±3.6	0.07
Range	0–14	0–16	
Quality-of-life score‡			
Mean ±SD	566±179	556±175	0.61
Range	50–786	113–776	
H. pylori–negative (%)	85	12	<0.001
Medication prescribed in past 6 mo (%)	43	53	0.09
Histamine H_2-receptor antagonists	25	28	
Proton-pump inhibitors	16	22	
Both	1	3	
Medication prescribed for >3 of the past 6 mo (%)	20	30	0.05
Eradication therapy prescribed during follow-up (no.)	3	9	0.08

*Ten patients were lost to follow-up: six patients in the group assigned to omeprazole and antibiotics and four patients in the group assigned to omeprazole alone.

†Total scores can range from 0 (no symptoms or limitations) to 20 (severe symptoms most days, requiring regular medication).

‡Quality of life was measured with the SF-36; scores can range from 0 (worst) to 800 (best).

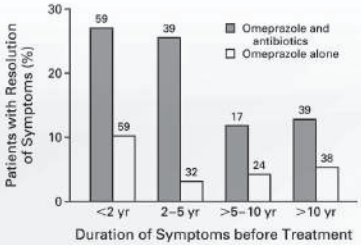


Figure 2. Percentages of Patients with Resolution of Symptoms in Relation to the Duration of Symptoms before Treatment.

The resolution of symptoms was defined as a Glasgow Dyspepsia Severity Score of 0 or 1. The numbers of patients in each group are shown above the bars.

base line; however, the higher scores at entry were due in part to the fact that all patients received a total of three points for visits to their general practitioner, visits to the hospital, and undergoing endoscopy.

At base line, 84 percent of the patients in the group given omeprazole and antibiotics and 80 percent of those in the group given omeprazole alone had been taking prescription drugs for dyspepsia during the preceding six months (Table 1). At one year, the respective values were 43 percent and 53 percent (Table 2). At base line, among patients who were taking prescription drugs for dyspepsia, 53 percent of the patients in the group given omeprazole and antibiotics had taken such a drug for more than three of the preceding six months, as compared with 55 percent of the patients in the group given omeprazole alone; at one year the respective values were 47 percent and 57 percent. There was no significant difference between groups in the quality of life after treatment (Table 2).

We used univariate logistic-regression analysis to examine whether any characteristics of the patients at base line could be used to predict a response to treatment. Only the duration of symptoms was predictive ($P=0.03$). None of the following factors were predictive: age ($P=0.81$), sex ($P=0.20$), smoking status ($P=0.80$), initial Glasgow Dyspepsia Severity Score ($P=0.35$), predominant symptom at presentation ($P=0.50$), *H. pylori* CagA status ($P=0.78$), acid secretion in response to pentagastrin ($P=0.92$), or acid secretion in response to gastrin-releasing peptide ($P=0.35$). In a multivariate model, treatment ($P<0.001$) and the duration of symptoms ($P=0.03$) were independently related to the resolution of symptoms. There was no evidence of an interaction between treatment and the duration of symptoms ($P=0.90$). The odds ratio for the resolution of symptoms in the group assigned to receive omeprazole and antibiotics, as compared with the group assigned to receive omeprazole alone, was 3.9 (95 percent confidence interval, 1.8 to 8.3). The percentage of patients with resolution of symptoms decreased with increasing duration of symptoms (Fig. 2). Among the patients who had had symptoms for five years or less, 7 of 91 patients (8 percent) in the group given omeprazole alone had a response to treatment, as compared with 26 of 98 patients (27 percent) in the group given omeprazole and antibiotics. Among the patients who had had symptoms for more than five years, 3 of 62 patients (5 percent) in the group given omeprazole alone had a response to treatment, as compared with 7 of 56 patients (12 percent) in the group given omeprazole and antibiotics.

Routine repeated endoscopy was not part of the study protocol. However, nine patients were referred for endoscopy by their primary care physician during the year after randomization because of severe, persistent dyspepsia. Three of the patients were in the group given omeprazole and antibiotics, and all three had normal findings on endoscopy. In contrast, six of the patients were in the group given omeprazole alone, and four had ulcers. Three of these four patients subsequently received therapy to eradicate *H. pylori*; the symptoms resolved in only one.

DISCUSSION

We found that therapy to eradicate *H. pylori* resulted in more frequent resolution of symptoms than did placebo in patients with nonulcer dyspepsia and documented *H. pylori* infection. Treatment with antibiotics and omeprazole resulted in the resolution of dyspepsia in 21 percent of patients; in contrast, symptoms resolved in 7 percent of patients who were treated with omeprazole alone. These results confirm and extend those of an earlier randomized trial showing symptomatic benefit from eradicating *H. pylori* in nonulcer dyspepsia.[22] The beneficial effect of this approach is also reflected by the findings in our study of a trend toward reduced use of prescription-drug therapy for dyspepsia ($P=0.09$) and a shorter duration of such therapy ($P=0.05$) in the group given omeprazole and antibiotics.

We found no evidence that factors associated with duodenal ulcer disease, including smoking, epigastric pain as the predominant symptom, male sex, a high acid output, or a positive test for *H. pylori* CagA,[19,23-25] were predictive of the success of treatment with omeprazole and antibiotics. Only the duration of dyspeptic symptoms before treatment was predictive of the response. The longer the history of dyspepsia, the less likely the benefit from this therapy. The mechanism of this association is not clear. It is possible that prolonged, symptomatic *H. pylori*–associated gastritis induces changes that become irreversible, or only slowly reversible, resulting in persistent dyspepsia despite eradication of the infection.

Our study also suggests that patients with *H. pylori* infection and nonulcer dyspepsia are at risk for ulcers. Of the six patients in the group given omeprazole alone who were referred for further evaluation because of severe, persistent dyspepsia, four had peptic ulcers. More ulcers might have been detected if all the patients in this group who had persistent symptoms had undergone follow-up endoscopy. A high incidence of peptic ulcers in patients with *H. pylori* infection and nonulcer dyspepsia has been reported in other studies. In one randomized study, ulcers developed in 14 percent of 50 *H. pylori*–positive patients with nonulcer dyspepsia in the placebo group during the one-year follow-up period.[22] In another, similar study, peptic ulcers developed in 21 percent of 41 patients during a mean follow-up period of three years.[26]

It is tempting to assume that the subgroup of patients with nonulcer dyspepsia in whom dyspepsia resolved on treatment of *H. pylori* corresponds to the subgroup of patients in the group given omeprazole alone in whom peptic ulcers developed. However, of the three patients in this group in whom ulcers developed and who subsequently underwent treatment for *H. pylori*, only one had resolution of dyspepsia. This finding is consistent with previous reports that the symptomatic response to the eradication of *H. pylori* varies in patients with peptic ulcers.[27-29]

Our findings have clinical implications for the treatment of patients with dyspepsia. Previous studies have shown that eradicating *H. pylori* is beneficial in patients with confirmed peptic ulcers; thus, eradication of the infection has been recommended only in such patients. We found that this approach was beneficial in a proportion of *H. pylori*–infected patients with dyspepsia who had no evidence of current or previous ulcers at the time treatment was initiated. Because it is not yet possible to identify prospectively patients whose symptoms will resolve with such treatment, it will be necessary to treat all patients with *H. pylori* infection and nonulcer dyspepsia in order to cure a minority. However, the potential benefit — curing a chronic disorder with a single course of antibiotic treatment — would seem to justify this approach.

Our study also has implications for the use of endoscopy in younger patients with uncomplicated dyspepsia. Endoscopy has been used to identify patients with *H. pylori* infection and associated peptic ulcer disease and thus to identify those who should be treated for this infection. Our findings that the eradication of *H. pylori* is also beneficial for patients with no peptic ulcers and that negative endoscopic results do not rule out the possibility of an ulcer in the future raise questions about the appropriateness of endoscopy in determining management. They support a strategy that includes noninvasive testing of patients with dyspepsia for *H. pylori* and eradicating *H. pylori* infection in those with evidence of infection.

Funded by the U.K. Medical Research Council.

We are indebted to Drs. Frank Wilson (Govanhill Health Center), Norman Gow (Woodside Health Center), Kenneth Cochran (Victoria Infirmary), and Booth Danesh (Stobhill General Hospital), all of Glasgow, Scotland.

REFERENCES

1. Jones RH, Lydeard SE, Hobbs FDR, et al. Dyspepsia in England and Scotland. Gut 1990;31:401-5.
2. Bernersen B, Johnsen R, Straume B, Burhol PG, Jenssen TG, Stakkevold PA. Towards a true prevalence of peptic ulcer: the Sorreisa gastrointestinal disorder study. Gut 1990;31:989-92.
3. Talley NJ, Zinsmeister AR, Schleck CD, Melton LJ III. Dyspepsia and dyspepsia subgroups: a population-based study. Gastroenterology 1992; 102:1259-68.
4. Nyrén O, Lindberg G, Lindstrom E, Marke L-A, Seensalu R. Economic costs of functional dyspepsia. Pharmacoeconomics 1992;1:312-24.
5. Johannessen T, Petersen H, Kleveland PM, et al. The predictive value of history in dyspepsia. Scand J Gastroenterol 1990;25:689-97.
6. Jebbink HJA, Smout AJPM, Van Berg-Henegouwen GP. Pathophysiology and treatment of functional dyspepsia. Scand J Gastroenterol Suppl 1993;200:8-14.
7. Talley NJ. Functional dyspepsia — should treatment be targeted on disturbed physiology? Aliment Pharmacol Ther 1995;9:107-15.
8. Mearin F, Cucala M, Azpiroz F, Malagelada J-R. The origin of symptoms on the brain-gut axis in functional dyspepsia. Gastroenterology 1991; 101:999-1006.
9. O'Riordan T, Mathai E, Tobin E, et al. Adjuvant antibiotic therapy in

duodenal ulcers treated with colloidal bismuth subcitrate. Gut 1990;31: 999-1002.
10. Marshall BJ, Goodwin CS, Warren JR, et al. Prospective double-blind trial of duodenal ulcer relapse after eradication of *Campylobacter pylori*. Lancet 1988;2:1437-42.
11. Hentschel E, Brandstätter G, Dragosics B, et al. Effect of ranitidine and amoxicillin plus metronidazole on the eradication of *Helicobacter pylori* and the recurrence of duodenal ulcer. N Engl J Med 1993;328:308-12.
12. Rauws EAJ, Tytgat GN. Cure of duodenal ulcer associated with eradication of *Helicobacter pylori*. Lancet 1990;335:1233-5.
13. Armstrong D. *Helicobacter pylori* infection and dyspepsia. Scand J Gastroenterol Suppl 1996;215:38-47.
14. Talley NJ. *Helicobacter pylori* and non-ulcer dyspepsia. Scand J Gastroenterol Suppl 1996;220:19-22.
15. Talley NJ, Hunt RH. What role does *Helicobacter pylori* play in dyspepsia and nonulcer dyspepsia? Arguments for and against *H. pylori* being associated with dyspeptic symptoms. Gastroenterology 1997;113:Suppl: S67-S77.
16. Management of dyspepsia: report of a working party. Lancet 1988;1: 576-9.
17. El-Omar EM, Banerjee S, Wirz A, McColl KEL. The Glasgow Dyspepsia Severity Score — a tool for the global measurement of dyspepsia. Eur J Gastroenterol Hepatol 1996;8:967-71.
18. Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? BMJ 1993;306:1440-4.
19. McColl KEL, el-Nujumi A, Murray L, et al. The *Helicobacter pylori* breath test: a surrogate marker for peptic ulcer in dyspeptic patients. Gut 1997;40:302-6.
20. Xiang Z, Bugnoli M, Ponzetto A, et al. Detection in an enzyme immunoassay of an immune response to a recombinant fragment of the 128 kilodalton protein (CagA) of *Helicobacter pylori*. Eur J Clin Microbiol Infect Dis 1993;12:739-45.
21. Megraud F. Advantages and disadvantages of current diagnostic tests for the detection of *Helicobacter pylori*. Scand J Gastroenterol Suppl 1996; 215:57-62.
22. Gilvarry J, Buckley MJM, Beattie S, Hamilton H, O'Morain CA. Eradication of *Helicobacter pylori* affects symptoms in non-ulcer dyspepsia. Scand J Gastroenterol 1997;32:535-40.
23. El-Omar EM, Penman ID, Ardill JES, Chittajallu RS, Howie C, McColl KEL. *Helicobacter pylori* infection and abnormalities of acid secretion in patients with duodenal ulcer disease. Gastroenterology 1995;109: 681-91.
24. Gillen D, el-Omar EM, Wirz AA, Ardill JES, McColl KEL. The acid response to gastrin distinguishes duodenal ulcer patients from *Helicobacter pylori*-infected healthy subjects. Gastroenterology 1998;114:50-7.
25. Cover TL, Glupczynski Y, Lage AP, et al. Serologic detection of infection with cagA+ *Helicobacter pylori* strains. J Clin Microbiol 1995;33: 1496-500.
26. Matysiak-Budnik T, Poniewierka E, Gasciniak G, Jelen M, Knapik Z. A 5-year follow-up study of chronic gastritis patients. Ir J Med Sci 1992; 161:Suppl 10:37. abstract.
27. Reilly TG, Ayres RCS, Poxon V, Walt RP. *Helicobacter pylori* eradication in a clinical setting: success rates and the effect on the quality of life in peptic ulcer. Aliment Pharmacol Ther 1995;9:483-90.
28. Phull PS, Halliday D, Price AB, Jacyna MR. Absence of dyspeptic symptoms as a test for *Helicobacter pylori* eradication. BMJ 1996;312:349-50.
29. McColl KEL, el-Nujumi A, Murray LS, et al. Assessment of symptomatic response as predictor of *Helicobacter pylori* status following eradication therapy in patients with ulcer. Gut 1998;42:618-22.

APPENDIX 4

Title: *CALCIUM-CHANNEL BLOCKADE AND INCIDENCE OF CANCER IN AGED POPULATIONS* , By: Pahor, Marco, Guralnik, Jack M., Lancet, 0099-5355, August 24, 1996, Vol. 348, Issue 9026
Database: *Biomedical Reference Collection: Comprehensive*

CALCIUM-CHANNEL BLOCKADE AND INCIDENCE OF CANCER IN AGED POPULATIONS

Summary

Background Calcium-channel blockers can alter apoptosis, a mechanism for destruction of cancer cells. We examined whether the long-term use of calcium-channel blockers is associated with an increased risk of cancer.

Methods Between 1988 and 1992 we carried out a prospective cohort study of 5052 people aged 71 years or more and who lived in three regions of Massachusetts, Iowa, and Connecticut USA. Those taking calcium-channel blockers (n=451) were compared with all other participants (n=4601). The incidence of cancer was assessed by survey of hospital discharge diagnoses and causes of death. These outcomes were validated by the cancer registry in the one region where it was available. Demographic variables, disability, cigarette smoking, alcohol consumption, blood pressure, body-mass index, use of other drugs, hospital admissions for other causes, and comorbidity were all assessed as possible confounding factors.

Findings The hazard ratio for cancer associated with calcium-channel blockers (1549 person-years, 47 events) compared with those not taking calcium-channel blockers (17225 person-years, 373 events) was 1.72 (95% CI 1.27-2.34, p=0.0005), after adjustment for confounding factors. A significant dose-response gradient was found. Hazard ratios associated with verapamil, diltiazem, and nifedipine did not differ significantly from each other. The results remained unchanged in community-specific analyses. The association between calcium-channel blockers and cancer was found with most of the common cancers.

Interpretation Calcium-channel blockers were associated with a general increased risk of cancer in the study populations, which suggested a common mechanism. These observational findings should be confirmed by other studies.

Introduction

Calcium-channel blockers are prescribed primarily for treatment of hypertension and coronary heart disease but lately their long-term safety has been questioned.[ ] Excess all-cause mortality has been associated with the use of short-acting nifedipine.[,] Calcium-channel blockers are potent drugs that affect various organ systems. Calcium-channel blockers are used to treat oesophageal diseases,[ ] and cause constipation,[ ] increase risk

of haemorrhage,[7,8] and can impair differentiation during embryogenesis.[9] Furthermore, cases of lupus after use of diltiazem have been reported.[10]

We have hypothesised that calcium-channel blockers increase the risk of cancer by interfering with physiological mechanisms that regulate cancer cell growth.[4,11] Evidence is emerging that calcium-channel blockers can block apoptosis,[12-14] an efficient mechanism for limiting cancer growth.[15-18] Calcium-channel blockers might affect cancer risk generally or be limited to specific sites where calcium mechanisms predominate. For example, colon cancer has been related to reduced calcium ingestion.[19]

In this prospective study, the initial analyses done in individuals receiving treatment for hypertension[11] have been extended to the general older population. The aim was to assess whether individuals taking calcium-channel blockers, for any indication were at higher risk of developing cancer than those not taking those drugs. The study focused on specific cancers, additional potential confounding factors, and the effects of dose.

Methods

The study was based on the Established Populations for Epidemiologic Studies of the Elderly (EPESE)--a collaborative, prospective cohort study of older persons supported by the US National Institute on Aging.[20] The participants surveyed were from three regions. Between 1982 and 1983, a regional survey was carried out on all persons aged 65 years or older living in East Boston, Massachusetts, and in the counties of Iowa and Washington in the state of Iowa. During the same period of time another survey was done in New Haven, Connecticut, where a random sample stratified by housing type and sex was taken.

Over 80% of the eligible population was interviewed. At the initial EPESE interview, and after 3 and 6 years of follow-up, trained interviewers administered a 90-min questionnaire that covered a wide range of psychosocial and health-related issues. Telephone follow-up interviews were done 1, 2, 4, and 5 years after the initial interview. Cancer events were assessed for the time between the sixth EPESE annual follow-up in 1988 and the end of 1992.

Information on date of hospital admission and up to five discharge diagnoses for each person in the survey was gathered from the Health Care Financing Administration Medicare Provider Analysis and Review (MEDPAR) files. The match of EPESE participants with MEDPAR files was complete from Jan 1, 1985, until Dec 31, 1992.

Vital status on Dec 31, 1992, was assessed at the seventh follow-up (Iowa and Connecticut) by contact with relatives, examination of obituaries in local newspapers, and linkage with the National Death Index.[21] One nosologist studied death certificates and then used ICD-9 codes to record the underlying, immediate, and contributing causes of death. Hospital discharge diagnoses or causes of death with ICD-9 codes 140-208 recorded at any time in the MEDPAR files were defined as cancer events. Specific codes were used to identify site of cancer.

In previous studies, hospital chart reviews provided validation of cancer as a hospital discharge diagnosis.[22] In this study, in the individuals from the Iowa regions cancer outcomes were validated by matching of the EPESE participants with the State Health Registry of Iowa's cancer registry, which is part of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program.[23] The kappa statistic[24] for the agreement of the MEDPAR files and death certificates with the SEER registry in ascertaining prevalent (before baseline), incident, and any prevalent or incident cancer

cases was 0.77, 0.83, and 0.86, respectively.

During the section of the baseline interview on the use of medications, participants were asked to show all containers for all prescription and non-prescription drugs taken over the past 2 weeks. The interviewer transcribed the drug product name from the container label. If the label was not seen, the interviewer asked the participant the name of the drug. These ascertainment methods are similar to those of other studies and have proven valid and reliable.[25,26] Participants taking calcium-channel blockers were compared with participants who were not. In addition, separate analyses were done for verapamil, nifedipine, and diltiazem.

For participants living in Massachussetts and Iowa the daily dose of calcium-channel blocker was calculated; the strength of the drug product was multiplied by the number of times the medication was taken per day. In Connecticut the dose could not be calculated. The median dose was established for verapamil, nifedipine, and diltiazem separately (240,30,and 180 mg/day, respectively), then the calcium-channel blocker dose was ranked as low (below median), median, or high (over median).

The following factors potentially associated with cancer were assessed at the baseline interview: age; sex; ethnic origin; coronary heart disease (use of nitrates, self-report of heart attack, any hospital discharge diagnosis coded 410-414 in the preceding 3 years, or a positive Rose questionnaire for angina assessed at any interview giving in person[27]); heart failure (use of both digoxin and a diuretic, or any hospital discharge diagnosis coded 428 in the preceding 3 years); hypertension (systolic blood pressure >/=140 mm Hg or diastolic >/=90 mm Hg measured at any visit, self-report of a medical diagnosis of hypertension, or any hospital discharge diagnosis coded 401 in the preceding 3 years); stroke (self-report of stroke, or any hospital discharge diagnosis coded 430-434 in the preceding 3 years); diabetes (use of antidiabetic drugs, self-report of diabetes, or any hospital discharge diagnosis coded 250 in the preceding 3 years); use of b-blockers, angiotensin-converting enzyme (ACE) inhibitors, diuretics, digoxin, nitrates, non-steroidal anti-inflammatory drugs, aspirin, corticosteroids, oestrogens, or coumarin; current smoking status (non-smoker or smoking 1-19 or >/=20 cigarettes per day); alcohol intake (>/=28.35 g/day, assessed by self-reported frequency and number of consumed drinks of beer, wine, and spirits); and physical disability (limitation in one or more of the following: walking half a mile, climbing a flight of stairs, doing heavy housework, walking across a room, bathing, dressing, eating, or transferring from bed to chair).[28,29] Trained interviewers measured blood pressure two or three times according to the Hypertension Detection and Follow-up Program protocol.[30] The mean of the last two measurements was calculated. Body-mass index was calculated from reported weight and height. The number of hospital admissions before the qualifying event or before the end of follow-up was assessed.

Of the original population of more than 10 000 interviewed in 1982, 6566 participants were still alive and were interviewed in 1988, the baseline period for this study. We excluded 298 participants who could not be matched with MEDPAR files, and 1216 participants who reported a cancer at any of the interviews, who had a hospital discharge diagnosis of cancer in the 3 years before the study baseline, or who were taking anticancer drugs such as tamoxifen. The remaining population at risk was 5052 participants (1741 from East Boston, 2020 from Iowa, and 1291 from New Haven).

The first cancer event that occurred during follow-up was used as a primary endpoint. Participants with no events, by the time of the final follow-up in 1992 were censored on Dec 31, 1992, or at the time of death, whichever occurred first. The Kaplan-Meier method was used to plot survival free of cancer. Cox proportional hazards regression models were used to estimate the hazard ratio and 95% CI for the association of variables of interest with

cancer. The assumption of proportionality of hazard was assessed with log minus log plots and by tests of the interaction of exposure with time. In multivariate models, additional indicator variables were created to adjust for the effects of missing data for cigarette smoking, alcohol intake, and disability (1.1%, 0.8%, and 6.8% with missing data, respectively). In analyses summarising results for all three regions, the regression models were stratified by region.

Results

Compared with participants who were not taking calcium-channel blockers those taking these drugs were significantly more likely to have cardiovascular diseasesand diabetes, to use cardiovascular drugs, to be disabled, to have a lower diastolic blood pressure, and to be admitted to hospital (table 1). Age, sex, ethnic origin, cigarette smoking, alcohol intake, systolic blood pressure, body-mass index, and use of non-cardiovascular drugs did not differ significantly between the two groups.

During 18774 person-years of follow-up (mean follow-up time 3.7 years), 420 participants had a cancer event (140 in East Boston, 185 in Iowa, and 95 in New Haven; overall rate 22.4 per 1000 person-years; table 2). The most frequent cancers were those of the colon, prostate, lung, lymphatic and haemopoietic organs, urinary tract, and breast (table 3). 169 people died from cancer (9.0 per 1000 person-years).

Users of calcium-channel blockers had significantly higher crude rates of cancer than non-users (table 2, figure 1). Patients taking verapamil had the highest rate of cancer followed by patients taking nifedipine and diltiazem.

To assess independent risk factors for cancer that might have confounded the association of calcium-channel blockers with cancer, separate proportional hazard regression models (adjusted for age and sex) were calculated for groups of variables, including demographic characteristics, chronic diseases, medications, and general health indicators (cigarette smoking, alcohol intake, physical disability, blood pressure, body-mass index, and number of hospital admissions). In the initial models there were significant associations between age, being male, being black, cigarette smoking, and alcohol consumption and increased risk of cancer. An inverse relation with cancer was found for heart failure and number of hospital admissions ($p<0.1$). In a summary multivariate model adjusted for these potential confounding factors, calcium-channel blocker users had a significantly increased risk of cancer compared with all other participants (hazard ratio=1.72, $p=0.0005$, table 2). The association was significant for verapamil and nifedipine, but not for diltiazem. However, the difference in hazard ratio among specific calcium-channel blockers was not significant. In the fully adjusted models the hazard ratios were greater than those calculated without adjustment. In separate models adjusted for confounding factors and either systolic and diastolic blood pressure or any past diagnosis of hypertension, the hazard ratio of cancer associated with calcium-channel blockers was 1.67 (1.21-2.29), and 1.71 (1.25-2.32), respectively. The hazard ratio associated with use of b-blockers, ACE inhibitors, diuretics, digoxin, nitrates, non-steroidal anti-inflammatory drugs, aspirin, corticosteroids, or coumarin was not significant and was close to 1 (not shown).

In community-specific multivariate analyses, the hazard ratio of cancer for calcium-channel blocker users was 1.55 (0.89-2.71), 1.82 (1.12-2.98), and 1.84 (1.04-3.26), compared with all other participants, in East Boston, Iowa, and New Haven, respectively. Daily dose of calcium-channel blocker was assessed in East Boston and Iowa and increasing dose was associated with a significant gradient of increased hazard ratio of cancer when compared

with those who did not use calcium-channel blockers (figure 2).

Separate multivariate models stratified according to potential risk factors of cancer were calculated. The hazard ratio of cancer for calcium-channel blockers was 1.77 (1.20-2.63) in participants aged 71-79 years (n=2982) and 1.55 (0.93-2.59) in those aged 80 years and older (n=2070); 1.64 (1.09-2.46) in men (n=1796) and 1.81 (1.13-2.90) in women (n=3256); 1.60 (1.15-2.22) in white people (n=4757) and 3.42 (1.25-9.39) in black people (n=264); 1.78 (1.29-2.47) in those not currently cigarette smokers (n=4559) and 1.44 (0.55-3.73) in current cigarette smokers (n=439); 1.76 (1.27-2.42) in those drinking <28.35 g of alcohol per day (n=4762) and 1.26 (0.42-3.73) in those drinking >/=28.35 g of alcohol per day (n=252); 1.71 (1.23-2.37) in those with any lifetime diagnosis of hypertension (n=4163) and 1.99 (0.78-5.07) in those with no hypertension (n=889); 1.80 (1.17-2.76) in those with diastolic and systolic blood pressure below 90 mm Hg and below 140 mm Hg, respectively (n=2313); and 1.57 (0.97-2.54) in those with elevated (>/=90/140 mm Hg) blood pressure (n=1969).

In multivariate models stratified by number of hospital admissions, and in analyses that excluded cancer events in which a circulatory disease (ICD 9 codes 390-459) was the principal discharge diagnosis, the results for calcium-channel blockers were unchanged (summary hazard ratio for the stratified models 1.70 [1.25-2.31], and 1.62, [1.18-2.24], respectively). When individual cancer sites were analysed separately, calcium-channel blocker use was associated with a significantly increased hazard ratio for cancers of the uterus and adnexa uteri, and lymphatic and haemopoietic organs (table 3). The hazard ratio was increased, but did not reach significance for stomach, colon, rectum, breast, prostate, and urinary tract cancers. For liver, gallbladder and pancreas, and skin cancers, the hazard ratio was increased, but close to 1. A non-significant inverse relation was found for lung cancer (table 3).

Separate analyses used the Iowa SEER cancer registry data rather than the MEDPAR files or death certificates to define the population that was free of cancer at baseline and incident cases of cancer. In these analyses, the incidence rate of cancer was 12.1 per 1000 person-years among participants not taking calcium-channel blockers (120 events, 1829 participants) and 30.4 per 1000 person-years among those taking calcium-channel blockers (16 events, 106 participants, p<0.001). After adjustment for confounding factors, the hazard ratio of cancer associated with calcium-channel blocker use was 3.09 (1.80-5.31), p<0.0001.

Discussion

We found that after adjustment for confounding factors, older people taking calcium-channel blockers for any indication had a significant 70% excess hazard ratio of developing cancer during 3.7 years (mean) of follow-up compared with those not taking these drugs. This association was consistent in all three regions in this study, and for most types of cancer. The dose-response gradient was significant. The excess risk of cancer associated with use of calcium-channel blockers was even more evident when the SEER cancer registry information was used to define the events. No association with cancer was found for other cardiovascular medications.

If calcium-channel blockers truly promote cancer, what are the potential mechanisms? Apoptosis, or programmed cell death, is an efficient physiological mechanism for elimination of cells with genetic lesions, and, therefore, for preventing cancer.[15] Calcium-channel blockers inhibit apoptosis in various experimental models by interfering with calcium-triggered signals.[12,14,18] At therapeutic concentrations, verapamil and nifedipine suppress gene activity associated with involution by apoptosis of prostate and breast after the removal of physiological hormone stimulation.[12,31] Such mechanisms may also prevent apoptosis of neoplastic cells, thereby promoting cancer growth.[11] Inhibition of

calcium-mediated cell differentiation may also play a part.[18]

The consistent association of calcium-channel blockers with increase in most types of cancer found in this study suggests that calcium-channel blockers might promote cancer through a common underlying biological process, such as interference with apoptosis or cell differentiation. Calcium-channel blockers may act as cancer promoters but standard carcinogenicity bioassays, which are designed to identify chemical mutagens, may fail to recognise such promoters. Further research is needed to assess whether the inverse relation of calcium-channel blockers with lung cancer was produced by chance alone or by other specific mechanisms.

Calcium-channel blocker use was ascertained at baseline and the latent period required for producing a clinically relevant cancer could not be calculated. Interestingly the Kaplan-Meier curves start diverging at about 2 years after the start of the study. The latent period for the underlying mechanism by which calcium-channel blockers might promote cancer may be less than that needed for chemical mutagens.

Calcium-channel blockers can potentiate anticancer drugs and facilitate apoptosis. However, such effects were observed in animal and in-vitro models at toxic concentrations not tolerated in man, and were accompanied by a paradoxical increase in intracellular calcium.[32]

We have identified six potential limitations of the study. First, although several confounding factors were examined, observational data such as these cannot conclusively establish causal links. Whether calcium-channel blockers truly promote cancer needs to be confirmed in long-term clinical trials. Physical disabilities and the consequent incapacity of a number of participants in this study would probably preclude their attendance for follow-up visits. Thus, these individuals would probably be excluded from clinical trials.

Second, users of calcium-channel blockers had higher rates than the non-users of hospital admission for cardiovascular disease; this difference could have led to a bias in cancer detection. However, the significant calcium-channel blocker cancer association did not change after adjustment or stratification on number of hospital admissions, and after exclusion of cancer hospital admissions in which a circulatory disease was the principal diagnosis. Use of other antihypertensive agents, nitrates, digoxin, corticosteroids, and anticoagulants, which are used for treating diseases requiring frequent medical surveillance, was not associated with cancer. Patients with heart failure, who are frequently admitted to hospital, tended to have a lower hazard ratio of cancer. Patients who are elderly and have an important cardiovascular condition might not undergo cancer screening as often as those who are healthy.

A third possible limitation was that exposure and outcome may have been misclassified. Drug use was assessed only at baseline and some participants may have stopped taking calcium-channel blockers during follow-up. Furthermore, cancers that did not lead to hospital admission or death were not ascertained, and the diagnoses in the MEDPAR files and death certificates may not be accurate. However, such potential misclassifications would probably dilute any associations found and introduce a conservative bias, leading to an underestimation of the association of calcium-channel blocker use with cancer.

Fourth, our findings in older people who have higher rates of cancer may not be applicable to younger individuals. Fifth, mainly short-acting calcium-channel blockers were used in this study. Theoretically, with new calcium-channel blockers and the slow-release formulations the plasma concentrations may remain below a harmful threshold. Finally, hypertension has been associated with an increased risk of cancer mortality in middle-aged people.[33,34]

However, in this study blood pressure was not related to cancer and did not confound the calcium-channel blocker-cancer association.

Evaluation of long-term safety and efficacy of medications such as calcium-channel blockers, which exert multiple and incompletely identified actions, needs to focus on assessment of major outcomes and survival.[1-3,8,35,36] A judgment about the clinical implications of our findings is currently premature. The evidence provided by this single study is not sufficient to recommend withdrawal of treatment from current users. However, these observational data should stimulate further experimental, epidemiological, and clinical research to assess all the cardiovascular and non-cardiovascular effects of calcium-channel blockers.

This study was supported by contracts N01-AG-0-2105, N01-AG-0-2106, and N01-AG-0-2107 from the National Institute on Aging, Bethesda, MD, USA. MP was supported by a grant from Ministero per l'Universita e Ricerca Scientifica e Tecnologica, 60% N7020532, and from Consiglio Nazionale delle Ricerche, Italy, N95000959PF40.

Table 1: Population baseline characteristics according to use of calcium-channel blockers

Characteristic	Not taking calcium-channel blockers (n=4601)	Taking calcium-channel blockers (n=451)	p*
Demographic			
Mean (SE) age	79.3 (0.1)	79.0 (0.3)	p>0.1
Women	64.5%	64.1%	p>0.1
Ethic origin			
White	94.2%	93.6%	
Black	5.1%	6.4%	p>0.1
Other	0.7%	0	
Comorbid conditions			
Coronary heart disease	21.9%	71.2%	p<0.0001
Heart failure	9.7%	29.3%	p<0.0001
Hypertension	81.7%	90.0%	p<0.0001
Stroke	9.2%	12.6%	p=0.016
Diabetes	15.7%	26.2%	p<0.0001
Medications			
beta-blockers	11.6%	23.3%	p<0.0001
ACE inhibitors	4.4%	8.6%	p<0.0001
Diuretics	33.5%	50.6%	p<0.0001
Digoxin	11.6%	25.1%	p<0.0001
Nitrates	6.2%	38.8%	p<0.0001
NSAIDs	12.5%	14.6%	p>0.1
Aspirin	29.7%	37.3%	p<0.001
Oestrogens	0.8%	0.2%	p>0.1
Corticosteroids	3.2%	3.8%	p>0.1
Coumarin	2.6%	6.0%	p<0.0001
Mean (SE) blood			

pressure (mm Hg)			
Systolic	136 (0.4)	138 (1.3)	p>0.1
Diastolic	73 (0.2)	72 (0.6)	p=0.033
Mean (SE) body-mass index (kg/m2)	24.4 (0.1)	24.8 (0.3)	p>0.1
Smoking status			
Current non-smoker	90.1%	91.8%	p>0.1
1-19 cigarettes per day	4.9%	5.3%	
>/=20 cigarettes per day	3.8%	2.7%	
Alcohol intake >/=28.35 g per day	5.0%	4.4%	p>0.1
Health indicators			
Physical disability	45.4%	62.7%	p<0.0001
Mean (SE) number of hospital admissions	1.2 (0)	2.2 (0.1)	p<0.0001

Percentages may not total 100% because of missing data.
*chi^2 test for categorical variables, ANOVA or Mann-Whitney tests for comparison of means.
ACE=angiotensin converting enzyme.
NSAIDs=non-steroidal anti-inflammatory drugs.

Table 2: Use of calcium-channel blockers and their relation to cancer

LINE 1: Number of events; Person-years; Rate per 1000 person-years

	Unadjusted model				Adjusted model*	
LINE 2:	Hazard ratio (95% CI)			p	Hazard ratio (95% CI)	p
All calcium-channel blockers						
Non-users (n=4601)	373	17225	21.7			
1	. .		1		. .	
Users (n=451)	47	1549	30.3			
1.42 (1.05-1.92)	0.032		1.72 (1.27-2.34)		0.0005	
Individual calcium-channel blockers						
Verapamil (n=118)	18	420	42.9			
2.03 (1.26-3.25)	0.0035		2.49 (1.54-4.01)		0.0002	
Nifedipine (n=146)	16	480	33.3			
1.56 (0.94-2.51)	0.083		1.74 (1.05-2.88)		0.031	
Diltiazem (n=184)	13	641	20.3			
0.94 (0.54-1.63)	>0.1		1.22 (0.70-2.12)		>0.1	

*Adjusted for age, sex, ethnic origin, heart failure, number of hospital admissions, cigarette smoking, and alcohol intake.
Three participants were excluded from the analyses of individual calcium-channel blockers because they were taking two

calcium-channel blockers.

Table 3: Relation of use of calcium-channel blockers with specific types of cancer

Type of cancer (ICD-9 code)	Number of events	Hazard ratio (95% CI)
Stomach (151)	13	3.64 (0.96-13.76)
Colon (153)	65	1.98 (0.90-4.38)
Rectum (154)	23	1.32 (0.31-5.74)
Liver, gallbladder, pancreas (155-157)	24	1.15 (0.26-4.96)
Lung (162)	56	0.21 (0.03-1.52)
Skin (172-173)	14	1.11 (0.14-8.62)
Breast (174)	31	1.65 (0.49-5.55)*
Uterus, adnexa (182-183)	23	3.69 (1.22-11.14)*
Prostate (185)	58	1.99 (0.93-4.27)
Bladder, ureter, kidney (188-189)	38	1.57 (0.55-4.47)
Lymphatic, haemopoietic (200-208)	46	2.57 (1.13-5.83)

Regression models stratified by community and adjusted for age, sex, ethnic origin, heart failure, number of hospital admissions, cigarette smoking, and alcohol intake.
*Adjusted for oestrogen use.

GRAPH: Figure 1: Probability of remaining free of cancer during follow-up according to use of calcium-channel blockers

GRAPH: **Figure 2: Hazard ratio and 95% CI of incident cancer according to daily dose of calcium-channel blocker compared with no calcium-channel blocker use** Estimates are adjusted for age, sex, ethnic origin, heart failure, number of hospital admissions, smoking, and alcohol intake. Median daily dose for verapamil, nifedipine, and diltiazem 240, 30, and 180 mg.

References

1 *Pahor M, Guralnik JM, Corti C, Foley DJ, Carbonin PU, Havlik RJ. Long-term survival and use of antihypertensive medications in older persons. J Am Geriatr Soc 1995; 43: 1191-97.*

2 *Furberg CD, Psaty BM, Meyer JV. Nifedipine. Dose-related increase in mortality in patients with coronary heart disease. Circulation 1995; 92: 1326-31.*

3 *Psaty BM, Heckbert SR, Koepsell TD, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. JAMA 1995; 274: 620-25.*

4 *Furberg CD, Pahor M, Psaty BM. The unnecessary controversy. Eur Heart J 1996; 17: 1142-47.*

5 *Konrad Dalhoff I, Baunack AR, Ramsch KD, et al. Effect of the calcium antagonists nifedipine, nitrendipine, nimodipine and nisoldipine on oesophageal motility in man. Eur J Clin Pharmacol 1991; 41: 313-16.*

6 *Pahor M, Guralnik JM, Chrischilles EA, Wallace RB. Use of laxative medication in older persons and*

associations with low serum albumin. J Am Geriatr Soc 1994; 42: 50-56.

7 Wagenknecht LE, Furberg CD, Hammon JW, Legault C, Troost BT. Surgical bleeding: unexpected effect of a calcium antagonist. BMJ 1995; 310: 776-77.

8 Pahor M, Guralnik JM, Furberg CD, Carbonin PU, Havlik RJ. Risk of gastrointestinal haemorrhage with calcium antagonists in hypertensive persons over 67 years old. Lancet 1996; 347: 1061-65.

9 Stein G, Srivastava MK, Merker HJ, Neubert D. Effects of calcium channel blockers on the development of early rat postimplantation embryos in culture. Arch Toxicol 1990; 64: 623-38.

10 Crowson AN, Magro CM. Diltiazem and subacute cutaneous lupus erythematosus-like lesions. N Engl J Med 1995; 333: 1429.

11 Pahor M, Guralnik JM, Salive ME, Corti MC, Carbonin P, Havlik RJ. Do calcium channel blockers increase the risk of cancer? Am J Hypertens 1996; 9: 695-99.

12 Connor J, Sawczuk IS, Benson MC, et al. Calcium channel antagonists delay regression of androgen-dependent tissues and suppress gene activity associated with cell death. Prostate 1988; 13: 119-30.

13 Juntti Berggren L, Larsson O, Rorsman P, et al. Increased activity of L-type Ca2+ channels exposed to serum from patients with type I diabetes. Science 1993; 261: 86-90.

14 Ray SD, Kamendulis LM, Gurule MW, Yorkin RD, Corcoran GB. Ca2+ antagonists inhibit DNA fragmentation and toxic cell death induced by acetaminophen. FASEB J 1993; 7: 453-63.

15 Carson DA, Ribeiro JM. Apoptosis and disease. Lancet 1993; 341: 1251-54.

16 Trump BF, Berezesky IK. Calcium-mediated cell injury and cell death. FASEB J 1995; 9: 219-28.

17 Martin SJ, Green DR. Apoptosis and cancer: the failure of controls on cell death and cell survival. Crit Rev Oncol Hematol 1995; 18: 137-53.

18 Whitfield JF. Calcium signals and cancer. Crit Rev Oncog 1992; 3: 55-90.

19 Garland C, Shekelle RB, Barrett-Connor E, Criqui MH, Rossof AH, Paul O. Dietary vitamin D and calcium and risk of colorectal cancer: a 19-year prospective study in men. Lancet 1985; i: 307-09.

20 Cornoni-Huntley J, Ostfeld AM, Taylor JO, et al. Established populations for epidemic studies of the elderly: study design and methodology. Aging Clin Exp Res 1993; 5: 27-37.

21 Stampfer MJ, Willett WC, Speizer FE, Dysert DC. Test of the National Death Index. Am J Epidemiol 1984; 119: 837-39.

22 Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Hlth 1992; 82: 243-48.

23 SEER cancer statistics review, 1973-1991: tables and graphs, National Cancer Institute. NIH publication No 94-2789. Bethesda, MD: NIH publication no 94-2789, 1994.

24 Cohen J. Weighted kappa: nominal scale agreement with provisionfor scaled disagreement or partial credit. Psychol Bull 1968; 70:213-20.

25 Psaty BM, Lee M, Savage PJ, Rutan GH, German PS, Lyles M. Assessing the use of medications

in the elderly: methods and initial experience in the Cardiovascular Health Study. J Clin Epidemiol 1992; 45: 638-92.

26 Pahor M, Chrischilles EA, Guralnik JM, Brown SL, Wallace RB, Carbonin PU. Drug data coding and analysis in epidemiologic studies. Eur J Epidemiol 1994; 10: 405-11.

27 Rose GA, Blackburn H, Gillum RF, Prineas RJ. Annex 6--London school of hygiene cardiovascular questionnaire. In: World Health Organisaton. Cardiovascular survey methods, 2nd edn. Geneva: WHO, 1982: 162-65.

28 Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963; 185:94-99.

29 Rosow I, Breslau N. A Guttman health scale for the aged. J Gerontol 1966; 21: 556-59.

30 Hypertension Detection and Follow-up Program Cooperative Group. Blood pressure studies in 14 communities; a two-stage screen for hypertension. JAMA 1977; 237: 2385-91.

31 Tenniswood MP, Guenette RS, Lakins J, Mooibroek M, Wong P, Welsh JE. Active cell death in hormone-dependent tissues. Cancer Metastasis Rev 1992; 11: 197-220.

32 Batra S, Popper LD, Hartley-Asp B. Effect of calcium and calcium antagonists on 45Ca influx and cellular growth of human prostatic tumor cells. Prostate 1991; 19: 299-311.

33 Raynor WJ, Shekelle RB, Rossof AH, Maliza M, Paul O. High blood pressure and 17-year cancer mortality in the Western Electric Health Study. Am J Epidemiol 1981; 118: 371-77.

34 Dyer AR, Stamler J, Berkson DM, Lindberg HA, Stevens E. High blood pressure: a risk factor for cancer mortality? Lancet 1975; i: 1051-56.

35 Furberg CD, Psaty BM. Calcium antagonists: antagonists or protagonists of mortality in elderly hypertensives? J Am Geriatr Soc 1995; 43: 1309-10.

36 Yusuf S. Calcium antagonists in cornary artery disease and hypertension. Time for reevaluation? Circulation 1995; 92:1079-82.

~~~~~~~~

By Marco Pahor, Jack M Guralnik, Luigi Ferrucci, Maria-Chiara Corti, Marcel E Salive, James R Cerhan, Robert B Wallace, Richard J Havlik

Department of Internal Medicine and Geriatrics, Catholic University, Rome, Italy (M Pahor MD); Epidemiology, Demography, and Biometry Program, National Institute on Aging, Bethesda, MD, USA (J M Guralnik MD, M-C Corti MD, M E Salive MD, R J Havlik MD); Geriatric Department, I Fraticini, Istituto Nazionale Ricerca e Cura Per Gli Anziani (INRCA), Florence, Italy (L Ferrucci MD); and Department of Preventive Medicine and Environmental Health, University of Iowa, IA, USA (J R Cerhan MD, R B Wallace MD

Correspondence to: Dr Marco Pahor, Department of Preventive Medicine, University of Tennessee, Memphis, 66 N Pauline,Suite 232, Memphis, TN 38105, USA

Source: Lancet, 08/24/96, Vol. 348 Issue 9026, p493, 5p, 3 charts, 2 graphs.
Item Number: 9608287821
~~~~~~~~

APPENDIX 5

BMJ 1997;315:1582-1587 (13 December)

Papers

Epileptic seizures after a first stroke: the Oxfordshire community stroke project

John Burn, *consultant in rehabilitation medicine*,[a] **Martin Dennis**, *senior lecturer in stroke medicine*,[b] **John Bamford**, *consultant neurologist*,[c] **Peter Sandercock**, *reader in medical neurology*,[b] **Derick Wade**, *consultant neurologist*,[d] **Charles Warlow**, *professor of medical neurology* [b]

[a] Rehabilitation Research Unit, Southampton General Hospital, Southampton SO9 4XY, [b] Department of Clinical Neurosciences, Western General Hospital, Edinburgh EH4 2XU, [c] Department of Neurology, St James's University Hospital, Leeds LS9 7TF, [d] Rivermead Rehabilitation Centre, Oxford OX1 4XD

Correspondence to: Dr Burn rehab@soton.ac.uk

Abstract

Objective: To describe the immediate and long term risk of epileptic seizures after a first ever stroke.
Design: Cohort study following up stroke survivors for 2 to 6.5 years; comparison with age specific incidence rates of epileptic seizures in the general population.
Setting: Community based stroke register.
Subjects: 675 patients with a first stroke, followed up for a minimum of 2 years.
Main outcome measures: Occurrence of single and recurrent seizures.
Results: 52 patients had one or more post stroke seizures; in 25 the seizures were recurrent. The 5 year actuarial risk of a post stroke seizure in survivors (excluding 19 patients with a history of epilepsy and 3 patients in whom the seizure occurred shortly before death from another cause) was 11.5% (95% confidence interval 4.8% to 18.2%). The relative risk of seizures, in comparison with the general population, was estimated at 35.2 in the first year after stroke and 19.0 in year 2. The risk of seizures was increased in survivors of subarachnoid and intracerebral haemorrhage (hazard ratio for intracranial haemorrhage *v* cerebral infarction 10.2 (3.7 to 27.9)). The risk of seizures after ischaemic stroke was substantial only in patients presenting with severe strokes due to total anterior circulation infarction. Only 9 of 295 patients (3%) independent one month after stroke suffered a seizure between 1 month and 5 years (actuarial risk 4.2% (0.1% to 8.3%)).
Conclusion: Stroke patients have about an 11.5% risk of single or recurrent seizures in the first 5 years after a stroke. Patients with more severe strokes or haemorrhagic strokes are at higher risk.

Key messages

- Cerebrovascular disease is an important cause of epilepsy but the risk of seizures after stroke may be overestimated by hospital based studies
- From a community based register, patients with a first ever stroke had a 2% risk of having a seizure at stroke onset and an 11% risk of having a later seizure in the first 5 years of follow up
- Patients with intracerebral and subarachnoid haemorrhage were at higher risk of seizures after stroke
- Survivors who were independent at 1 month were at very low risk of future seizures

Introduction

Cerebrovascular disease is an important cause of epilepsy, particularly in elderly people. When seizures complicate a clinical stroke they have a devastating effect on morale and further impair an already compromised quality of life. Precise estimates of the risk of developing epilepsy would be helpful not only to patients but also to those who give advice on returning to work or driving. Data from hospital based studies may give falsely high estimates of risk since patients with severe strokes, or strokes presenting with seizures, may be more likely to be admitted to hospital. In this study we had three aims: to describe the immediate and long term risk of epileptic seizures after a first stroke among patients in a community stroke register; to analyse the occurrence of seizures in relation to the pathological and clinical subtype of first stroke; and to compare the risk of a first epileptic seizure in these stroke patients with the risk in the general population.

Method

The cohort

The Oxfordshire community stroke project registered 675 patients with a first stroke, 357 (53%) women and 318 (47%) men, from a study population of about 105 000 over 4 years. Patients of all ages were included: 163 (24%) were under 65 years, 195 (29%) 65-74 years, 228 (34%) 75-84 years, and 89 (13%) over 84 years (mean age 72.2 years). Patients were recruited between 1981 and 1986 and were followed up until 1988, when the project finished (minimum follow up 2 years). The methods used to identify and

assess cases have been described elsewhere.[4 5] A total of 545 patients (81%) had cerebral infarction, 66 (10%) primary intracerebral haemorrhage, and 33 (5%) subarachnoid haemorrhage. In 31 (4%) cases the type of stroke could not be determined.

Ten patients had intracranial operations for aneurysm, and five of these patients received prophylactic phenytoin for 18 months to 2 years. Prophylactic anticonvulsant treatment was not offered to any other patients. One patient with primary intracerebral haemorrhage (not complicated by seizures) had intracranial surgery.

All but two of the 545 patients with cerebral infarction could be classified into subtypes according to their clinical presentation[6]: 92 patients had total anterior circulation infarction (cortical and subcortical ischaemia likely to be due to occlusion of the main stem of the middle cerebral artery or of the internal carotid artery); 185 patients had partial anterior circulation infarction (predominantly cortical ischaemia likely to be due to occlusion of a branch of the middle cerebral or anterior cerebral artery); 137 patients had lacunar infarction (a lacunar syndrome likely to be due to intracerebral small vessel occlusion[7]); and 129 patients had posterior circulation infarction (ischaemia in the brainstem, cerebellum or occipital cortex likely to be due to either small or large vessel occlusion).

Follow up

Survivors were interviewed by a study nurse at 1 month, 6 months, 12 months, and then annually around the anniversary of their first stroke. At each follow up the nurse asked the patient and caregivers specifically about whether any possible seizure had occurred. As a check the nurse also reviewed the records kept by the primary healthcare team. Patients with suspected seizures were reassessed clinically by the study neurologist. Long term survivors were visited at the close of the study by one of the study neurologists (JPSB), and all available information about possible seizures was reviewed again with the patient. The hospital and primary care records of patients who died before this final visit were reviewed by one of the study neurologists (JPSB), and data relating to possible seizures were discussed at regular consensus meetings of the study team. Generalised seizures were diagnosed with reference to statements of witnesses. Focal seizures were distinguished clinically from clonus, spasms, and recurrent strokes.[8 9] Electroencephalograms were not routinely performed. Each visit included an assessment of functional ability using the Oxford handicap scale[10]; in this report patients in grades 0-2 were described as independent and patients in grades 3-5 as dependent.

The Oxfordshire community stroke project followed up every patient for 2 years or until death. Patients recruited earlier in the study were followed for longer, and no patient was lost to follow up. Seventeen patients died on day 1, 129 by 30 days, 208 by 1 year, and 254 by 2 years. A total of 421 patients were followed up at 2 years, 274 were seen at 3 years, 182 at 4 years, and 92 at 5 years; 324 patients survived to a final examination by a study neurologist between August 1987 and November 1988.

Classification of seizures
Onset seizures were defined as occurring within 24 hours of the onset of stroke. This definition (see Appendix 1) is based on that used in a population based study of acute seizures after head injury.[11] Seizures occurring later were classified as post stroke seizures, except in two cases where evolution of stroke was prolonged.

Comparison with the general population
The incidence of post stroke seizures was compared with the estimated rate of epileptic seizures in the general population, using data from the VAMP research bank, a large British primary care database.[12 13] Eighty two practices had sent in data on the medical and prescribing history of 369 819 patients. The data were checked for completeness and internal consistency,[13] and practices supplying unreliable data were excluded.

Statistical methods
Actuarial analysis was used for cohort data involving the follow up of patients for varying lengths of time. Kaplan-Meier survival curves were constructed in which patients who died or were no longer being followed up (after 2 years) were censored. The occurrence of a post stroke seizure qualified as an end point, and these were graphically represented as a cumulative seizure rate. Different pathological types, age stratifications, and clinical subtypes of stroke were compared by using the log rank method.[14] The incidence of epileptic seizures was compared between stroke survivors and the general population and expressed as a ratio of the observed to the expected frequency with confidence intervals calculated from the Poisson distribution.[15] Calculations were made for year 1 and year 2 separately; there were insufficient events to continue this comparison in years 3-5. Confidence intervals of small samples were calculated by using an exact test.[15]

Results

Seizures preceding stroke
Nineteen of the 675 patients (3%) registered with the Oxfordshire community stroke project gave a history of one or more seizures before their first stroke or had documentation of past seizures in their medical record. In one patient the seizures were secondary to a craniotomy for a benign tumour; in one they occurred in the context of eclampsia; and in two they were related to alcohol abuse. The remaining seizures were considered idiopathic, and some may have been caused by otherwise asymptomatic cerebrovascular disease. Eleven of the 19 patients had had a seizure in the year before the stroke: in seven this was a first seizure, and four of these seven patients had partial motor seizures in the same limb as was subsequently affected by the stroke.

Seizures at the onset of the first stroke
Fourteen patients (2%) had an onset seizure, none of whom had a previous history of seizures. In seven the seizure was generalised, in six it was a simple partial seizure, and in one it was a complex partial seizure.

The occurrence of an onset seizure was not, in this cohort, associated with a worse outcome. The 30 day case fatality rate of patients with onset seizures was not raised. Two patients (14%) with onset seizures died compared with 127 of the 661 without onset seizures (19.2%, 95% confidence interval 16.2% to 22.2%). There was also no suggestion that patients with onset seizures had a worse functional prognosis. Three of ten previously independent patients with onset seizures were dependent at one month compared with 179 of those without onset seizures (38.3%, 33.9% to 42.7%).

The occurrence of an onset seizure was, however, associated with an increased risk of having further seizures. Five (36%) of the 14 patients with onset seizures went on to have one or more post stroke seizures compared with 43 (6.9%) of 625 patients who survived the first day without a history of pre-stroke or onset seizures (odds ratio 7.52, 2.46 to 22.98).

Patients with subarachnoid haemorrhage had an excess of onset seizures but this was not statistically significant. Four (4%) of the 99 patients with intracranial haemorrhage had an onset seizure compared with 10 (1.8%) of the 545 patients with cerebral infarction (odds ratio 2.25, 0.69 to 7.33) (table 1).

Table 1 Frequency of onset seizures and development of post stroke seizures by type of first stroke

Type of first stroke	No (%; 95% CI) of patients with onset seizures	Number (%) of patients with onset seizures who developed post stroke seizures
Cerebral infarction (n=545)	10 (2; 0.9 to 3.4)	4 (40)
Primary intracerebral haemorhage (n=66)	2 (3; 0.4 to 10.5)	1 (50)
Subarachnoid haemorrhage (n=33)	2 (6; 0.7 to 20.2)	0
Unknown (n=31)	0 (0; 0 to 11.2)	0
Total (n=675)	14 (2; 1.1 to 3.5)	5 (36)

Post stroke seizures

Fifty two (8%) of 658 patients who survived the first day had one or more post stroke seizures. Three patients had their first seizure only a few hours before death as part of the terminal illness and one patient had had seizures before the stroke. Most analyses were restricted to the remaining 48 patients. Twenty three patients had only a single post stroke seizure; 25 patients had recurrent post stroke seizures as defined in the appendix (table 2), but these generally occurred infrequently; only eight patients had more than one seizure a

month over a period of two months or more. Only two of the five patients who had previously had an onset seizure went on to develop recurrent seizures.

Twenty four (50%) of the 48 patients were treated with anticonvulsant drugs; six of these had had only one post stroke seizure. Forty (83%) patients had at least one generalised seizure. Partial seizures occurring alone were less common than at the onset of the stroke.

Table 2 Number (% of cohort; 95% confidence interval) of patients with single and recurrent seizures after first stroke (four patients who had seizures within a few hours of death as part of a terminal illness, or seizures that started before the stroke, were excluded)

Classification of first stroke	Single post stroke seizure	Recurrent post stroke seizures	Total
Cerebral infarction (n=545):	17 (3)	18 (3)	35 (6; 4 to 9)
Total anterior circulation infarction (n=92)	5 (5)	10 (11)	15 (16)
Partial anterior circulation infarction (n=185)	5 (3)	3 (2)	8 (4)
Lacunar infarction (n=137)	3 (2)	2 (1)	5 (3)
Posterior circulation infarction (n=129)	4 (3)	3 (2)	7 (5)
Primary intracerebral haemorrhage (n=66)	3 (5)	4 (6)	7 (11; 3 to 18)
Subarachnoid haemorrhage (n=33)	3 (9)	3 (9)	6 (18; 5 to 31)
Unknown (n=31)	0	0	0
Total (n=675)	23 (3)	25 (4)	48 (7; 6 to 9)

Risk of post stroke seizure

A Kaplan-Meier survival curve was constructed in which the occurrence of a first post stroke seizure was recorded as an end point and deaths were censored (table 3, fig 1). Patients with a history of epilepsy before the stroke (n=19) or who died in the first day after the stroke (n=17) were excluded from these analyses. The probability of having a post stroke seizure was 5.7% (3.5% to 7.9%) within the first year and 11.5% (4.8% to 18.2%) within 5 years. The incremental risk averaged about 1.5% a year after the first 12 months. Five of the 10 patients who had a first seizure more than 2 years after the stroke had had another stroke more recently, and two had taken prophylactic phenytoin for two years after a subarachnoid haemorrhage.

Table 3 Cumulative actuarial risks (95% confidence intervals) of experiencing a seizure after stroke by type of first stroke (19 patients with a history of prestroke seizures were excluded)

Time after stroke	Cerebral infarction	Primary intracerebral haemorrhage	Subarachnoid haemorrhage	Total
1 year	4.2 (2.2 to 6.2)	19.9 (1.5 to 38.3)	22.0 (2.6 to 41.8)	5.7 (3.5 to 7.9)
2 years	6.7 (4.1 to 9.3)	19.9 (1.5 to 38.3)	27.8 (5.3 to 50.7)	8.2 (5.4 to 11.0)

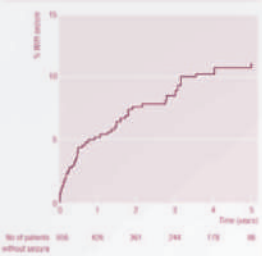


Fig 1 Kaplan-Meier estimate of the risk of post stroke seizures after first stroke; bars indicate 95% confidence intervals. Patients with a history of prestroke seizures (n=19) were excluded and deaths censored. Number of patients still being followed up at 1-5 years who had not had a seizure are indicated. Six of the 92 patients still being followed up at 5 years had had a previous seizure

The probability of a post stroke seizure was higher among patients aged over 84 years (observed/expected 2.0) but there was no significant trend on log rank analysis for an increasing risk with age (χ^2=0.91, df=1, P>0.25). Subsequent analyses were not adjusted for age.

Patients who had had a haemorrhagic first stroke were more at risk of post stroke seizures than patients who had cerebral infarction (hazard ratio 10.2, 3.7 to 27.9) (table 3); fig 2. However, the lower risk after cerebral infarction concealed a significant heterogeneity between different subtypes (χ^2=34.0, df=3, P<0.001) (table 2); fig 3.

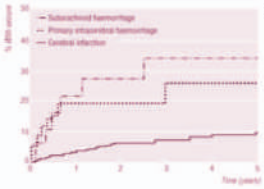


Fig 2 Kaplan-Meier estimate of post stroke seizures subdivided by type of stroke

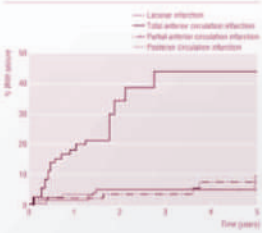


Fig 3 Kaplan-Meier estimates of post stroke seizures after cerebral infarction, subdivided by clinical subtype

Survivors of total anterior circulation infarction had a markedly increased risk of post stroke seizures when compared with the survivors of other subtypes of cerebral infarction.

The good prognosis for survivors of partial anterior circulation infarction, lacunar infarction, and posterior circulation infarction was reflected in the low probability of future post stroke seizures among independent stroke survivors (those with Oxford handicap scale score 0-2). A further Kaplan-Meier curve restricted to patients alive and independent at 1 month revealed an incremental risk of suffering a post stroke seizure of about 0.9% a year and an actuarial risk at 5 years of 4.2% (0.1% to 8.3%). If only survivors independent at 6 months were included the 5 year actuarial risk was 2.7% (0 to 6.0%).

Comparison with the general population

Forty eight patients had at least one post stroke seizure recorded in a total follow up period (for patients without a history of epilepsy) of 1641 patient years—an incidence of 29.3 patients developing post stroke seizures per 1000 patient years, or 23.5 per 1000 patient years when patients with intracerebral infarction, subarachnoid haemorrhage, or unknown pathology were excluded. This incidence was compared year by year for the first 2 years after stroke with the age specific incidence in the general population, as recorded by the VAMP research data bank over the 12 months between 1 February 1989 and 31 January 1990. Stroke survivors had a 35-fold increased risk of seizures in the first year (table 4). The increase in risk was particularly evident for patients under 65 years.

Discussion

An earlier, community based study of seizures after stroke reported that 9.5% of 518 survivors suffered a seizure 2-5 years after stroke, but no mention was made of subjects who were not followed up or who had died before the assessment.[16] Such problems can be addressed by an actuarial analysis which censors patients who die or who are lost to follow up. This method of analysis has been used in hospital based studies of epilepsy after stroke in New York[17] and Sweden[18] and in a population based study of seizures after cerebral infarction in Rochester, Minnesota.[19] The 5% risk of developing recurrent seizures by 5 years after the stroke found in the Swedish study[18] is similar to the risk observed in the Oxfordshire community stroke project (5.4%, 0.3% to 10.5%). The New York study reported a 19% risk of single or recurrent seizures by six years after an

current British licensing regulations which allow stroke patients who are functionally competent to return to driving after 30 days.

Acknowledgements

We gratefully acknowledge support from the Medical Research Council (UK) and the Stroke Association. We also thank Lesley Jones, computer programmer, for her work in data management and statistical analysis, and the study nurses Sue Price, Liz Mogridge, and Claire Clifford. The project would not have been possible without the cooperation of the following general practices (liaison partners): Beaumont Street, Oxford (Dr A MacPherson); East Oxford Health Centre (Dr D Leggate); Thame (Dr A Marcus); Berinsfield (Dr M Agass); Stert Street, Abingdon (Dr N Crossley); Marcham Road, Abingdon (Dr R Pinches); Malthouse, Abingdon (Dr D Otterburn); Kidlington (Dr S Street); Wantage (Dr V Drury); Deddington (Dr H O'Donell). We also thank Mr Jim Slattery for statistical advice.

Funding: MRC and the Stroke Association.

Conflict of interest: None.

Appendix

A *single seizure* includes a single episode of status epilepticus or multiple seizures with less than 2 weeks between events. Seizures which were "provoked" by a stroke recurrence, occurring at the onset of a new cerebrovascular event, were included as post stroke seizures but were not included in the definition of recurrent seizures.

Recurrent seizures—two or more unprovoked seizures with at least two weeks between events.[29]

Prestroke seizure—single or recurrent seizures occurring before the day of the first stroke.

Onset seizures—seizures occurring within 24 hours of stroke onset or, in two cases, at three and four days during the evolution of a progressive stroke while it was still getting worse.

Post stroke seizures—seizures occuring more than 24 hours after stroke onset or after a stable deficit had been established.

References

1. Roberts R, Shorvon S, Cox T, Gilliatt R. Clinically unsuspected cerebral

infarction revealed by CT scanning in late onset epilepsy. *Epilepsia* 1988;29:190-4. [Medline]

2. Roberts M, Godfrey J, Caird F. Epileptic seizures in the elderly: aetiology and type of seizure. *Age Ageing* 1982;11:24-8.

3. Bamford J, Sandercock P, Warlow C, Gray M. Why are patients with acute stroke admitted to hospital. *BMJ* 1986;292:1369-72.

4. Bamford J, Sandercock P, Dennis M, Warlow C, Jones L, McPherson K, et al. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire community stroke project 1981-1986. *J Neurol Neurosurg Psychiatry* 1988;51:1373-80.

5. Bamford J, Sandercock P, Dennis M, Burn J , Warlow C. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire community stroke project 1981-1986. Part 2. Incidence, case fatality rates and overall outcome at one year of cerebral infarction, primary intracerebral haemorrhage and subarachnoid haemorrhage.*J Neurol Neurosurg Psychiatry* 1990;53:16-22.

6. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* 1991;337:1521-6. [Medline]

7. Bamford J, Warlow C. Evolution and testing of the lacunar hypothesis. *Stroke* 1988;19:1074-82. [Medline]

8. Baquis GD, Pessin MS, Scott RM. Limb shaking—a carotid TIA. *Stroke* 1985;16:444-8. [Abstract]

9. Ropper AH. "Convulsions" in basilar artery occlusion. *Neurology* 1988;38:1500-1.

10. Bamford JM, Sandercock PAG, Warlow CP, Slattery J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 1989;20:828. [Medline]

11. Annegers JF, Grabow JD, Groover RV, Laws ER, Elveback LR, Kurland LT. Seizures after head trauma: a population study. *Neurology* 1980;30: 683-9.

12. Tallis R, Hall G, Craig I, Dean A. How common are epileptic seizures in old age? *Age Ageing* 1991;20:442-8.

13. Hall GC, Luscombe DK, Walker SR. Post-marketing surveillance using a computerised general practice database. *Pharm Med* 1988; 2:345-51.

14. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design

and analysis of randomised clinical trials requiring prolonged observation of each patient. 2. Analysis and examples. *Br J Cancer* 1977;35:1-39.

15. Gardner MJ, Altman DG, eds. *Statistics with confidence*. London: BMJ, 1989.

16. Moskowitz E, Lightbody FEH, Freitag NS. Long term follow up of the poststroke patient. *Arch Phys Med Rehabil* 1972;53:167-72.

17. Hauser WA, Ramirez-Lassepas H, Rosenstein R. Risk for seizures and epilepsy following cerebrovascular insults [abstract]. *Epilepsia* 1984;25:666.

18. Viitanen M, Erikson S, Asplund K. Risk of recurrent stroke, myocardial infarction and epilepsy during long term follow up after stroke. *Eur Neurol* 1988;28:227-31. [Medline]

19. So E, Annegers J, Hauser W, O'Brien P, Whisnant M. Population-based study of seizure disorders after cerebral infarction. *Neurology* 1996;46:350-5. [Abstract]

20. Bamford J, Sandercock P, Warlow C, Gray M. Why are patients with acute stroke admitted to hospital? *BMJ* 1986;292:1369-72.

21. Barolin GS, Scherzer E, Schnaberth G. Epileptische Manifestationen als Vorboten von Schlaganfällen. "Vaskuläre Präkursiv Epilepsie." *Fortschr Neurol Psychiatr* 1971;39:199-216.

22. Goodridge DMG, Shorvon SD. Epileptic seizures in a population of 6000. 1. Demography, diagnosis, and classification and role of the hospital services. *BMJ* 1983;287:641-4. [Medline]

23. Shinton RA, Gill JS, Zezulka AV, Beevers DG. The frequency of epilepsy preceding stroke. *Lancet* 1987;1:11-3.

24. Starkey IR, Warlow CP. Epilepsy preceding stroke [letter]. *Lancet* 1987;1:742-3. [Medline]

25. Kase CS, Wolf PA, Chodosh EH, Zacker HB, Kelly-Hayes M, Kannel WB, et al. Prevalence of silent stroke in patients with initial stroke: the Framingham study. *Stroke* 1989;20:850-2.

26. Shinton RA, Gill JS, Melnick SC, Gupta AK, Beevers DG. The frequency, characteristics and prognosis of epileptic seizures at the onset of stroke. *J Neurol Neurosurg Psychiatry* 1988;51:273-6. [Abstract]

27. Davalos A, Cendra E, Genis D, Lopez-Pousa S. The frequency, characteristics and prognosis of epileptic seizures at the onset of stroke [letter]. *J Neurol Neurosurg Psychiatry* 1988;51:1464. [Medline]

28. Aring CD, Merrit HH. Differential diagnosis between cerebral haemorrhage and cerebral thrombosis. *Arch Intern Med* 1935;56:435-56.

29. Kilpatrick CJ, Davis SM, Tress BM, Rossiter SC, Hopper JL, Vandendriesen ML. Epileptic seizures in acute stroke. *Arch Neurol* 1990;47:157-60.

30. Reith J, Jorgensen HS, Nakayama H, Raascha HO, Olsen TS. Early seizures do not worsen rehabilitation outcome. The Copenhagen stroke study [abstract]. *Eur J Neurol* 1996; 3(suppl 2):93.

31. Jennet B. *Epilepsy after non-missile head injury*. London: Whitefriars Press, 1975.

32. Gupta SR, Naheedy MN, Elias D, Rubino FA. Postinfarction seizures: a clinical study. *Stroke* 1988;19:1477-81.

33. De Carolis P, D'Alessandro RD, Ferrara R, Andreoli A, Saquegna T, Lugaresi E. Late seizures in patients with internal carotid artery and middle cerebral artery occlusive disease following ischaemic events. *J Neurol Neurosurg Psychiatry* 1984;47:1345-7.

34. Dodge PR, Richardson EP, Victor M. Recurrent convulsive seizures as a sequel to cerebral infarction: a clinical and pathological study. *Brain* 1954;77:610-38.

35. Evans JH. The significance of early post-traumatic epilepsy. *Neurology* 1963;13:207-12.

36. Olsen TS, Hogenhaven H, Thage O. Epilepsy after stroke. *Neurology* 1987;37:1209-11. [Abstract]

37. French JD, Gernandt BE, Livingstone RB. Regional differences in seizure susceptibility in monkey cortex. *Arch Neurol Psychiatr* 1956;75:260-74.

38. Faught E, Peters D. Seizures in intracerebral haemorrhage [abstract]. *Epilepsia* 1984;25:666.

39. Sung C-Y, Chu N-S. Epileptic seizures in intracerebral haemorrhage. *J Neurol Neurosurg Psychiatry* 1989;52:1273-6.

40. Avrahami A, Drury VE, Rabey MJ, Cohn DF. Generalised epileptic seizures as the presenting symptom of lacunar infarction in the brain. *J Neurol* 1988;235:472-4. [Medline]

(Accepted 8 July 1997)

APPENDIX 6

The prevalence and causes of dementia in people under the age of 65 years

R J Harvey, M Skelton-Robinson and M N Rossor

J. Neurol. Neurosurg. Psychiatry 2003;74;1206-1209
doi:10.1136/jnnp.74.9.1206

PAPER

The prevalence and causes of dementia in people under the age of 65 years

R J Harvey, M Skelton-Robinson, M N Rossor

J Neurol Neurosurg Psychiatry 2003;**74**:1206–1209

See end of article for authors' affiliations

Correspondence to: Dr Harvey, Community Mental Health, The Geelong Hospital, Ryrie Street, Geelong, Victoria, Australia; rj.harvey@virgin.net

Received 3 February 2003
In revised form 24 March 2003
Accepted 25 March 2003

Objectives: To determine the prevalence of dementia in people under the age of 65 in a large catchment area, and use these figures to estimate the number of younger people affected by dementia in the UK.
Design: Epidemiological catchment area prevalence survey.
Setting: The London boroughs of Kensington and Chelsea, Westminster, and Hillingdon with a total population of 567 500 people.
Participants: All residents of the catchment area with dementia, where the illness began before the age of 65 years. Participants were notified to the study by medical and care professionals. The diagnosis and age of onset was established from all available health and social care records. In total, 227 people were identified, of whom 185 fulfilled the inclusion criteria of having a dementia which started before their 65th birthday.
Main outcome measures: Diagnosis of dementia and differential diagnosis of the cause of the dementia.
Results: The prevalence of dementia in those aged 30–64 was 54.0 per 100 000 (95% CI 45.1 to 64.1 per 100 000). For those aged 45–64 years, the prevalence was 98.1 per 100 000 (95% CI 81.1 to 118.0 per 100 000). From the age of 35 onwards, the prevalence of dementia approximately doubled with each 5 year increase in age. Extrapolating these figures nationally suggests that there are 18 319 (15 296–21 758) people with dementia under the age of 65 in the UK.
Conclusions: The study confirms previous "guestimates" of the number of younger people affected by dementia in UK. The prevalence figures generated are robust, and are supported by other smaller and targeted prevalence surveys. The prevalence figures provided by this study will allow health planners to accurately estimate need and plan services.

In the past decade, interest and concern for younger people with dementia have dramatically increased. Following sustained campaigning and support from the Alzheimer's Society[1 2] and other voluntary organisations[3] in the UK, the needs of younger people with dementia and those who care for them have been increasingly addressed. The Royal College of Psychiatrists has recently issued guidelines recommending that every area in the country has a named old age psychiatrist responsible for younger people with dementia.[4] This has been matched by the development of specialised support services, day care, and care homes for this group of patients.

However, delivering effective services relies on accurate epidemiological data to drive the planning and implementation process. In the past, estimates of the prevalence of dementia in younger people have relied on historical data[5–8] drawn from hospital inpatient data and diagnostic registers. The inherent problem of studies based upon known, diagnosed cases is the potential for underestimating the real prevalence, particularly given the rarity and relatively small numbers involved. Unfortunately, because of the relatively low prevalence, conventional "door knocking" would be too costly and time consuming.

The aim of our study was to use a large, geographical catchment area and to use a combination of techniques to identify every known case of dementia starting before the age of 65 years. Having identified cases, we were also interested in exploring the prevalence of the different causes of dementia, and using the results obtained to extrapolate estimates of the numbers of younger people with dementia in the UK.

METHODS

Ethical approval was obtained from the five local research ethics committees covering the catchment areas of the London Borough of Hillingdon (110 km², population 232 000, of whom 81 184 were between 30 and 64 years of age; Jarman Index 8)[9] and the London boroughs of Kensington, Chelsea, and Westminster (25 km², population 335 500 of whom 112 309 were between 30 and 64 years of age; Jarman Index 21).

Sampling

Stage 1

The aim of the first stage of sampling was to raise awareness of the project, and the need to identify and report any case, or suspected case of dementia, where the illness had started before the age of 65 years.

Personal letters and information were sent out to every general practitioner, psychiatrist, old age psychiatrist, neurologist, geriatrician, and general physician. In addition, information was also provided to departments of social services, voluntary organisations, day centres, home care, and community care teams through senior managers in the relevant organisations. The project ran in parallel with social services/voluntary sector initiatives in both areas,[9 10] providing senior organisational leadership across health and social services and ensuring widespread engagement with the project. Clinicians and practitioners were asked to notify any cases under their care on a defined "census" day for the study.

In parallel, hospital information systems and case registers at the four main hospitals in the catchment areas were searched for cases falling in the target age range who had relevant ICD-9/10 diagnoses.

Personal contact was made with key neurologists and psychiatrists working in the area who allowed hand searching of copies of clinic letters, discharge summaries, and departmental databases.

Stage 2

An experienced psychiatrist (RJH) followed up notification of cases, requesting copies and/or access to all available hospital and GP case notes. The available information included clinical history, and investigations (including neuroimaging) were reviewed.

A hierarchical diagnostic algorithm was then applied. At the top level, subjects achieved a definite diagnosis if the dementia was associated with a known disease-specific genetic mutation, if there were neuropathological results available from a cerebral biopsy, or from autopsy if the patient had died after the "census" date.

At the second level of the algorithm, well recognised and validated clinical diagnostic criteria were applied to make the diagnoses of Alzheimer's disease (NINCDS/ADRDA criteria)[11], vascular dementia (NINDS/AIREN criteria)[12], dementia with Lewy bodies,[13] frontotemporal dementia (Lund and Manchester Criteria)[14], alcohol related dementia (DSM-IV criteria for alcohol induced persisting dementia)[15], or other dementias for which DSM-IV criteria are available. These criteria were not applied hierarchically, but rather on the basis of best clinical opinion.

At the third level of the algorithm, patients fulfilling the DSM-IV criteria for dementia, but not fulfilling criteria for one of the above diagnostic categories were assigned to a Dementia Not Otherwise Specified category.

The age at onset of disease was defined as the age of the patient at which the earliest conclusive dementia symptom was noted or documented in the medical notes or other correspondence.

Stage 3

For half of the patients identified in stage 2, a direct approach was made through the general practitioner seeking consent for a detailed clinical assessment to confirm diagnosis. Those patients and carers willing to participate were visited at home by RJH, who completed a semi-structured interview to collect demographic details and a detailed clinical history. The clinical assessment included a medical and neurological examination and completion of the Modified Hachinski Ischaemia Scale,[16] Clinical Dementia Rating Scale ,[17] the Behavioural Pathology in Alzheimer's Disease Rating Scale,[18] the Interview to Determine Deterioration in Daily Functioning,[19] and the Cambridge Mental Disorders in the Elderly Examination Cognitive Assessment.[20] The data collected were used to confirm the diagnosis.

Any diagnoses that were not clear at either stage 2 or stage 3 were discussed and agreed by consensus between RJH and MNR.

RESULTS

In total, 227 people with suspected young onset dementia were identified to the study. The sources of identification were psychologists (21%), databases and case registers (20%), neurologists (17%), psychiatrists (14%), social workers (9%), GPs (6%), physicians (6%), community nurses (4%), and other (3%).

Of the 227 referrals, 19 were excluded as not fulfilling DSM-IV criteria for dementia (six were "worried well", seven had traumatic brain injuries, two had chronic schizophrenia, and four cases identified from hospital information systems were clearly miscoded when the notes were reviewed).

Of the remaining 208 cases, review of the notes showed an age of onset greater than 65 years of age in 23 cases who were further excluded from the study. The remaining 185 cases formed the population for the prevalence study, and of these, 87 cases (47%) underwent a detailed stage 3 personal assessment.

In the stage 3 assessment group, 69 patients (83%) had undergone neuroimaging with either CT or MRI, with 18 (21%) having undergone both CT and MRI. Twenty one patients (24%) had had an EEG.

The mean age of the population was 58.7 years (95% CI 57.4 to 60.1 years). Fifty eight percent (95% CI 50.9 to 65.6%) were male and 42% (95% CI 34.4 to 49.1%) were female, suggesting a trend towards a higher prevalence of dementia in males than females.

Table 1 summarises the age and gender specific prevalence rates for dementia in the study population on the "census" day; the pattern of prevalence approximately doubles with each five year increase in age.

Pragmatically, mixed diagnoses were avoided and we aimed to make a "highest confidence" single diagnosis in each case. In 168 (90%) of cases a single diagnosis was established using the diagnostic algorithm, 19 cases (10%) fitted more than one set of criteria and required a clinical consensus diagnosis to be made. Differential diagnoses in the population included Alzheimer's disease (62 cases, 34%), vascular dementia (34 cases, 18%), frontotemporal dementia (23 cases, 12%), alcohol related dementia (19 cases, 10%), and dementia with Lewy bodies (12 cases, 7%). Other causes of dementia accounted for 35 cases (19%) and included Huntington's disease (nine

Table 1 Age and gender specific prevalence rates in the study population

Population			All causes of dementia									
			All			Male			Female			
Age range	Male (n)	Female (n)	n	Rate*	95% CI	n	Rate	95% CI	n	Rate	95% CI	Significance†
30–34	23898	23375	6	12.7	(4.7 to 26.7)‡	3	12.6	(2.6 to 36.7)	3	12.8	(2.7 to 37.5)	NS
35–39	18526	19106	3	8.0	(1.6 to 23.3)	1	5.4	(0.1 to 30.1)	2	10.5	(1.3 to 37.8)	NS
40–44	18982	19643	6	15.5	(5.7 to 33.8)	1	5.3	(0.1 to 29.4)	5	25.5	(8.3 to 59.4)	NS
45–49	16549	16799	11	33.0	(16.5 to 59.0)	6	36.3	(13.3 to 78.9)	5	29.8	(9.7 to 69.5)	NS
50–54	15185	15237	19	62.5	(37.6 to 97.5)	10	65.9	(31.6 to 121)	9	59.1	(27 to 112)	NS
55–59	13983	13626	42	152.1	(110 to 206)	28	200.2	(133 to 289)	14	102.7	(56.2 to 172)	NS
60–64	12716	13141	43	166.3	(120 to 224)	26	204.5	(134 to 300)	17	129.4	(75.4 to 207)	NS
30–64	119839	120927	130	54.0	(45.1 to 64.1)	75	62.6	(49.2 to 78.4)	55	45.5	(34.3 to 59.2)	NS
45–64	58433	58803	115	98.1	(81.1 to 118)	70	119.8	(93.4 to 151.0)	45	76.5	(55.8 to 102)	NS
Over 65§			55			33			22			

*Rate per 100 000 people at risk.
†Significance of difference between genders by inference from 95% CI.
‡95% confidence interval for the prevalence rate.
§Subjects who had a dementia starting before the age of 65 years, but were over 65 but still living on the study census day.

Table 2 Age specific prevalence rates for the most common causes of young onset dementia

	Alzheimer's disease			Vascular dementia			Frontotemporal dementia			Alcohol related dementia		
Age range	n	Rate*	95% CI	n	Rate*	95% CI	n	Rate*	95% CI	n	Rate	95% CI
40–44	1	2.6	(0.7 to 14.4)									
45–49	2	6.0	(0.7 to 21.7)				4	12.0	(3.3 to 30.7)	2	6.0	(0.7 to 21.7)
50–54	5	16.4	(5.3 to 38.4)	2	6.6	(0.8 to 24.4)	1	3.3	(0.8 to 18.3)	6	19.7	(7.2 to 42.9)
55–59	14	50.7	(27.7 to 85.1)	9	32.6	(14.9 to 67.9)	7	25.4	(10.2 to 52.2)	5	18.1	(5.9 to 42.3)
60–64	20	77.3	(47.2 to 119)	10	38.7	(18.5 to 71.1)	6	23.2	(8.5 to 50.5)	3	11.6	(2.4 to 33.9)
30–64	42	17.4	(12.6 to 23.6)	21	8.7	(5.4 to 13.3)	18	7.5	(4.4 to 11.8)	16	6.6	(3.8 to 10.8)
45–64	41	35.0	(25.1 to 47.4)	21	17.9	(11.1 to 27.4)	18	15.4	(9.1 to 24.3)	16	13.6	(7.8 to 22.2)

*Rate per 100 000 people at risk.

Table 3 Estimated numbers of cases of dementia in people under the age of 65 in the UK

Age	UK Population 2001	All causes of dementia	Alzheimer's disease	Vascular dementia	Frontotemporal dementia	Alcohol related dementia
30–34	4493585	570 (209 to 1240)*				
35–39	4625810	369 (76 to 1078)				
40–44	4151580	645 (237 to 1403)	107 (27 to 598)			
45–49	3735964	1232 (616 to 2204)	224 (27 to 811)		448 (122 to 1147)	224 (27 to 811)
50–54	4040437	2523 (1519 to 3939)	664 (216 to 1552)	266 (32 to 96)	133 (34 to 739)	797 (293 to 1733)
55–59	3338861	5079 (3673 to 6878)	1693 (925 to 2841)	1088 (497 to 2267)	847 (341 to 1743)	605 (196 to 1412)
60–64	2879948	4789 (3456 to 6451)	2228 (1359 to 3427)	1114 (533 to 2048)	668 (245 to 1454)	334 (69 to 976)
30–64	27266185	18319 (15296 to 21758)	5918 (4254 to 7989)	2959 (1832 to 4526)	2536 (1502 to 4008)	2255 (1290 to 3654)
45–64	13995210	13728 (1134 to 16514)	4894 (3513 to 6634)	2507 (1553 to 3835)	2149 (1274 to 3401)	1910 (1092 to 3107)

*95% confidence interval of the estimate.

cases), dementia in multiple sclerosis (eight cases), corticobasal degeneration (two cases), prion disease (two cases), dementia in Down's syndrome (three cases), dementia in Parkinson's disease (two cases), and dementia not otherwise specified (nine cases). The very low prevalence of dementia in Down's syndrome is likely to be an artefact due to sampling bias. Despite raising awareness of the project with learning disability services, no referrals of cases were made from these sources.

The age specific prevalence rates for the four most common causes of dementia in this group of younger people are summarised in table 2.

Estimates of numbers of cases in the UK

One of the objectives of the study was to generate robust estimates of the numbers of people in the UK with young onset dementia. These figures are important both for service planning and for campaigning for improvements in recognition and provision.

Using data from the 2001 UK Census, we applied the prevalence figures derived from our population to the population figures for the UK. Table 3 summarises the estimated numbers of younger people with dementia, and with specific diagnoses in the UK.

DISCUSSION

This is the first study to report the prevalence of dementia and differential diagnoses in a community based catchment area study of younger people with dementia. The prevalence of dementia in people between the ages of 30 and 64 years was 54 per 100 000 with the prevalence approximately doubling for each five year age band. Based on the prevalence figures derived from this study we estimate that there are 18 319 (15 296–21 758) people under the age of 65 years with dementia in the UK.

The main strength of the study is the combination of a large catchment area and the use of multiple and varied means of case identification, essentially attempting to leave no stone unturned. The very varied sources of case identification confirm the benefits of this methodology. The main weakness of the study is that despite using a catchment area with a population of more than 500 000 people, the rarity of the condition meant that only 185 cases were identified. For more robust estimates with smaller confidence intervals a much larger population would be helpful.

This study was unique in being large enough to consider a range of differential diagnoses with our disease specific figures and overall figures being supported by other studies of individual diseases.

Thus our prevalence figure for young onset Alzheimer's disease of 35.0 per 100 000 in the 45–64 years age group is very similar to estimates for the Northern Region (34.6),[4] Finland (32.7),[21] the Framingham Study (31.8),[20] and Cambridge (15.1).[8]

Similarly, the prevalence rates we found for vascular dementia were close to the estimates derived from an earlier study in Scotland,[3] and the prevalence rate we identified for frontotemporal dementia (15.4 per 100 000) was almost identical to the rate identified in the Cambridge study (15.1).[8]

While none of these studies used an identical methodology, this study is probably most similar to the study from the Northern Region which used case ascertainment from hospital inpatient records, and the Cambridge study which identified cases from memory clinic databases and admission records. The Framingham study and the Finnish study both used ascertainment from large population diagnostic registers, but generate similar prevalence figures. One conclusion from this is that younger people with dementia inevitably come to medical attention and receive a diagnosis—what is less certain is what follow up and support these patients are given after the diagnosis.

Our study underpins and supports initiatives from the Alzheimer's Society and the Royal College of Psychiatrists to secure and improve services for this group of patients. Even though dementia in younger people is uncommon when compared with dementia in older people, every case is a personal

tragedy for the individual and their family, which often includes young children.

The study also highlights the differences between dementia in younger people and dementia in older people. Alzheimer's disease is considerably less common in this group, accounting for only 34% of cases, whereas in the elderly it would be expected to account for up to 80% of cases. Moreover, unusual diseases such as frontotemporal dementia accounts for 12% of cases, although there is increasing evidence that it is a commonly missed diagnosis in older people.[23]

The prevalence data generated by this study can be applied to local population figures to derive reliable estimates of the potential numbers of cases in any area. This should support better planning and delivery of services.

ACKNOWLEDGEMENTS

Project funded by NHS Executive (North Thames) Responsive R&D Grant, RFG045. Richard Harvey is funded by the Alzheimer's Society.

Funding was awarded in open competition and following peer review. The funder had no role in study design, data collection, analysis, interpretation of data, or in the writing of the report. The funder encouraged us to submit the paper for publication as part of good research governance.

Competing interests: none declared

Authors' affiliations

R J Harvey, M N Rossor, Imperial College London and the Institute of Neurology (UCL), London, UK
M Skelton-Robinson, Woodlands Centre, Hillingdon Hospital, Hillingdon, UK

REFERENCES

1 **Alzheimer's Disease Society**. Younger People with Dementia: a review and strategy. London: Alzheimer's Disease Society, 1996.
2 **Alzheimer's Society**. Younger people with dementia: a guide to service development and provision. London: Alzheimer's Society, 2001.
3 **Quinn**, C. The Care Must Be There. London: Dementia Relief Trust, 1996.
4 **Royal College of Psychiatrists.** Services for Younger People with Alzheimer's Disease and other Dementias. CR77. London: Royal College of Psychiatrists, 2000.
5 **McGonigal G**, Thomas B, McQuade C, *et al*. Epidemiology of Alzheimer's presenile dementia in Scotland, 1974–88. *BMJ* 1993;**306**:680–3.
6 **Newens AJ**, Forster DP, Kay DW, *et al*. Clinically diagnosed presenile dementia of the Alzheimer type in the Northern Health Region: ascertainment, prevalence, incidence and survival. *Psychol Med* 1993;**23**:631–44.
7 **Woodburn K**, Johnstone E. Ascertainment of a population of people with early-onset dementia in Lothian, Scotland. *Int J Geriatr Psychiatry* 1999;**14**:362–7.
8 **Ratnavolli E**, Brayne C, Dawson K, *et al*. The prevalence of frontotemporal dementia. Neurol 2002;**58**:1615–21.
9 **Jarman B**. Identification of underprivileged areas. *BMJ* 1983;**286**:1705–9.
10 **Kirk L**, Smith M, Templeton H, *et al*. Early Onset Dementia (Pre-senile Dementia): A Case of Need. Hillingdon: Health and Social Service Purchasing Agencies, 1995.
11 **McKhann G**, Drachman D, Folstein M, *et al*. Clinical diagnosis of Alzheimer's Disease: Report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurol* 1984;**34**:939–44.
12 **Roman GC**, Tatemichi TK, Erkinjuntti T, *et al*. Vascular Dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurol 1993;**43**:250–60.
13 **McKeith IG**, Galasko D, Kosaka K, *et al*. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): Report of the consortium on DLB international workshop. *Neurol* 1996;**47**:1113–24.
14 **The Lund and Manchester Groups**. Clinical and neuropathological criteria for frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 1994;**57**:416–8.
15 **American Psychiatric Association**. Diagnostic and Statistical Manual of Mental Disorders, 4th edn (DSM-IV). Washington DC: APA, 1994.
16 **Rosen WG**, Terry RD, Fuld PA, *et al*. Pathological verification of ischemic score in differentiation of dementias. *Ann Neurol* 1980;**7**:486–8.
17 **Hughes CP**, Berg L, Danziger WL, *et al*. A new clinical scale for the staging of dementia. *Brit J Psychiatry* 1982;**140**:566–72.
18 **Reisberg B**, Borenstein J, Salob SP, *et al*. Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *J Clin Psychiatry* 1987;**48** Suppl:9–15.
19 **Teunisse S**, Mayke M, Van Crevel H. Assessing the severity of dementia. *Arch Neurol* 1991;**48**:274–7.
20 **Blessed G**, Black SE, Butler T, *et al*. The diagnosis of dementia in the elderly. A comparison of CAMCOG (the Cognitive section of CAMDEX), the AGECAT program, DSM-III, the Mini-Mental State Examination and some short rating scales. *Brit J Psychiatry* 1991;**159**:193–8.
21 **Sulkava R**, Wikstrom J, Aromaa A, *et al*. Prevalence of severe dementia in Finland. *Neurol* 1985;**35**:1025–9.
22 **Kokmen E**, Beard CM, Offord KP, *et al*. Prevalence of medically diagnosed dementia in a defined United States population: Rochester, Minnesota, January 1 1975. *Neurol* 1989;**39**:773–6.
23 **Stevens T**, Livingston G, Kitchen G, *et al*. Islington study of dementia subtypes in the community. *Brit J Psychiatry* 2002;**180**:270–6.

University Excellence Center

Clinical Excellence Office

Critical Appraisal

Supervised By: Dr. Shahram Yazdani

Editors:

Dr. Hadi Beyhaghi

Dr. Saeed Hatami

Dr. Armin Shirvani

Contributors:

Dr. Hadi Beyhaghi

Dr. Elnaz Ehteshami Afshar

Dr. Soroush Hasani

Dr. Zahra Rahmani

Dr. Mona Seyed Safizadeh

Dr. Samira Yazdaninejad