

## بررسی همپوشانی سنتی، نسبی و درجه‌ی آزادی مرکب در دو پایگاه داده‌ی

### PubMed و Scopus در زمینه‌ی موضوعی بیماری‌های قلبی - عروقی\*

محمد رضا علی بیک<sup>۱</sup>، روح‌انگیز جمشیدی اورک<sup>۲</sup>، لیلیا اصغری هینه‌آباد<sup>۳</sup>

#### چکیده

**مقدمه:** تعیین میزان همپوشانی، روشی است که به مقایسه‌ی محتوای پایگاه‌های داده از جنبه‌ی کمی می‌پردازد. در این نوشتار میزان همپوشانی سنتی، نسبی و درجه‌ی آزادی مرکب پایگاه داده‌ی رایگان PubMed و پایگاه داده‌ی غیر رایگان Scopus در مورد بیماری‌های قلبی - عروقی ارایه می‌گردد.

**روش بررسی:** این پژوهش توصیفی - کاربردی در فاصله‌ی زمانی بهمن ماه ۱۳۸۷ تا تیر ماه ۱۳۸۸ صورت گرفت. ابتدا توصیف‌گرهای مرتبط با سه بیماری قلبی - عروقی شامل سکته‌ی قلبی - مغزی، حوادث عروقی - مغزی و پرفشاری خون از اصطلاح‌نامه Emtree انتخاب شدند. سپس با استفاده از آن‌ها مقالات مرتبط با سه بیماری مذکور در دو پایگاه داده‌ی PubMed و Scopus بازبایی گردیدند. داده‌های اولیه که برای محاسبه‌ی درصد‌های همپوشانی مورد نیاز بودند، به کمک نرم‌افزار Reference Manager به دست آمدند. تجزیه و تحلیل داده‌های به دست آمده با نرم‌افزار Excel، انجام شد.

**یافته‌ها:** در مجموع ۹۸۹۹ عنوان مقاله در هر دو پایگاه داده در زمینه‌ی موضوعی بیماری‌های قلبی - عروقی به دست آمد. ۸۵۹۳ عنوان مقاله در پایگاه داده‌ی PubMed و ۷۷۱۳ عنوان مقاله در پایگاه داده‌ی Scopus بازبایی شدند. تعداد ۶۴۰۷ عنوان مقاله مشترک در هر دو پایگاه داده، ۲۱۸۶ عنوان مقاله منحصر به پایگاه داده‌ی PubMed و ۱۳۰۶ عنوان مقاله منحصر به پایگاه داده‌ی Scopus بودند. درصد همپوشانی سنتی میان دو پایگاه داده برای بیماری پرفشاری خون ۶۸ درصد، برای بیماری حوادث عروقی - مغزی ۶۶/۷ درصد و در خصوص بیماری سکته‌ی قلبی - مغزی ۶۰/۷ درصد محاسبه گردید. همپوشانی‌های نسبی پایگاه داده‌ی PubMed و Scopus در مورد بیماری‌های سکته‌ی قلبی - مغزی، پرفشاری خون و حوادث عروقی - مغزی به ترتیب (۷۵/۸ درصد در مقابل ۷۵/۳ درصد)، (۸۹/۲ درصد در مقابل ۷۴/۱ درصد) و (۸۶/۷ درصد در مقابل ۷۴/۳ درصد) به دست آمد. مقدار شاخص درجه‌ی آزادی مرکب در هر سه بیماری مورد بررسی، کمتر از ۰/۵ حاصل شد.

**نتیجه‌گیری:** درصد‌های همپوشانی نسبی بالاتر پایگاه داده‌ی PubMed نسبت به پایگاه داده‌ی Scopus، کاربران و کتابداران را به سمت استفاده از این پایگاه داده هدایت می‌کنند. تعداد مقالات بازبایی شده منحصر به فرد بیشتر برای هر بیماری در پایگاه داده‌ی PubMed و عدم نیاز به پرداخت حق اشتراک، استفاده از این پایگاه داده را مرجح می‌کند.

**واژه‌های کلیدی:** پایگاه‌های اطلاعاتی؛ PubMed؛ بیماری‌های قلب و عروق.

#### نوع مقاله: تحقیقی

پندیرش مقاله: ۹۰/۴/۱۲

اصلاح نهایی: ۸۸/۸/۱۷

دریافت مقاله: ۸۸/۵/۶

**ارجاع:** علی بیک محمد رضا، جمشیدی اورک روح‌انگیز، اصغری هینه‌آباد لیلیا. بررسی همپوشانی سنتی، نسبی، و درجه‌ی آزادی مرکب در دو پایگاه داده‌ی PubMed و Scopus در زمینه‌ی موضوعی بیماری‌های قلبی عروقی. مدیریت اطلاعات سلامت ۱۳۹۰؛ ۸ (۳): ۳۴۵-۳۵۳.

#### مقدمه

\* این مقاله حاصل پایان‌نامه‌ی دانشجویی در مقطع کارشناسی ارشد می‌باشد.  
۱. مربی، کتابداری و اطلاع‌رسانی پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران.  
۲. استادیار، آمار زیستی، دانشگاه علوم پزشکی تهران، تهران، ایران.  
۳. دانشجوی کارشناسی ارشد، کتابداری و اطلاع‌رسانی پزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران. (نویسنده‌ی مسؤول)

Email: tabrizleila@gmail.com

امروزه علاوه بر متون چاپی، پایگاه‌های داده نیز به طور وسیعی در میان جوامع علمی کشور مورد استفاده قرار می‌گیرند. این

روان‌شناسی آموزشی بررسی کردند. این پژوهشگران با توجه به هزینه‌ی کمتر پایگاه داده‌ی Eric استفاده از آن را مقرون به صرفه یافتند. آن‌ها در مقاله‌ی خود راهکارهایی برای پرهیز از دوباره‌کاری هنگام انتخاب مقالات مرتبط با موضوع، در انجام تجسس ارایه کردند (۳).

Tonosaki به نقل از Mychko-Megrin در مطالعه‌ی منابع موجود در مورد موضوع «جلوگیری از خودکشی» را در Embase، Health Care and Administration، Medline، Hospital Literature Index HLTH و PsycInfo، MEDB (BRS full-text medical books) جستجو و مقایسه نمود. از میان ۱۰۳ منبع یافت شده، ۴ کتاب (۳/۹ درصد) در MEDB و ۹۹ مقاله در پایگاه‌های داده‌ی دیگر بازیابی شدند که در این میان، تعداد ۶۳ مقاله (۶۱/۲ درصد) غیر مشترک بودند. ۳۶ مقاله در دو یا چند پایگاه داده، همپوشانی داشتند. از میان مقالات غیر مشترک، ۵ مقاله در HLTH، ۹ مقاله در Embase، ۲۴ مقاله در Medline و ۲۵ مقاله در PsycInfo دیده شدند. جستجو در Medline، PsycInfo، EMCS و HLTH به ترتیب ۵۹ (۵۷/۳ درصد)، ۴۳ (۴۱/۷ درصد)، ۳۲ (۳۱/۱ درصد) و ۲۳ (۲۲/۳ درصد) از مقالات بازیابی شده را به دست می‌دادند (۴).

Haafkens و همکاران در مطالعه‌ی، ۵ پایگاه داده‌ی Medline، Embase، PsycInfo، Cinahl و OSHROM را بررسی کردند. آن‌ها دریافتند که از بین مدارک مرتبط بازیابی شده در مورد بیماری‌های کرونری و مشارکت کاری در Medline، Embase و PsycInfo به ترتیب ۳۶، ۴۰ و ۲۳ درصد از مدارک بازیابی می‌شوند. OSHROM و Cinahl درصد پایین‌تری را به خود اختصاص می‌دهند. جالب این که ۸۴ درصد از مدارک تنها در یک پایگاه داده، ۲۴ درصد در Medline، ۲۲ درصد در Embase، ۱۹ درصد در PsycInfo، ۱۳ درصد در OSHROM و ۶ درصد در Cinahl یافت شدند. بنابراین هر پایگاه داده مدارک منحصری به فردی را ارایه می‌دهند که در پایگاه‌های داده‌ی دیگر یافت

پایگاه‌ها از لحاظ قیمت و حق اشتراک متفاوت هستند. بعضی از آن‌ها را می‌توان مشترک شد. بعضی دیگر را می‌توان داخل یک مجموعه خریداری کرد. بعضی دیگر رایگان در اختیار عموم قرار می‌گیرند. پایگاه‌های داده‌ای که برای استفاده از آن‌ها باید حق اشتراک پرداخت کرد، ممکن است همیشه در دسترس نباشند. این امر ممکن است به دلیل کمبود بودجه و یا عدم توافق بین طرفین قرارداد پیش بیاید.

سؤالی که در این شرایط مطرح می‌شود این است که آیا پایگاه‌های داده‌ای که نیاز به حق اشتراک ندارند و به صورت رایگان در دسترس عموم قرار می‌گیرند، می‌توانند جایگزین پایگاه‌های نیازمند اشتراک شوند یا حداقل بخشی از نیازهای کاربر را برطرف کنند. آیا در زمینه‌ی موضوعی معین این جایگزینی، مجاز است؟

برای پاسخ به این سؤالات باید پایگاه‌های داده را از نظر عناوین مقالات نمایه شده و همچنین مجلاتی که مقالات آن‌ها نمایه شدند، مورد مقایسه قرار داد. چنین مقایسه‌ای از طریق شاخصی به نام درصد همپوشانی انجام می‌گیرد (۱).

Wilkinson به نقل از Gluck در یک بررسی پوشش مقالات دارویی Index Medicus، Pharmaceutical Index و موضوعات فرعی مربوط به اثرات سمی ۸ داروی خاص را که در درمان بیماری‌های گرمسیری به کار می‌روند، در ایندکس‌های ۷۱-۱۹۶۹ مورد مقایسه قرار داد. نتایج بررسی نشان دادند که Index Medicus ۱۹۸ مقاله و Pharmaceutical Index ۳۰۹ مقاله دارد. ۷۰ مقاله یا ۱۶ درصد کل مقاله‌ها در هر دو مشترک بودند. بالاترین درصد همپوشانی نیز برای این داروها ۳۸/۹ درصد گزارش شد. در پایان بررسی اضافه شده است که پوشش مطالعات سمیت دارویی در Index Medicus و Pharmaceutical Index، تکمیل‌کننده‌ی یکدیگرند و برای جستجوهای جامع در این زمینه بایستی هر دو منبع مورد استفاده قرار گیرند (۲).

Caldwell و Ellingson در یک پژوهش تطبیقی، همپوشانی بین دو پایگاه داده‌ی Eric و Psychological Abstracts را در زمینه‌ی موضوع

انتخاب پایگاه داده‌ی مناسب، به اطلاعات مرتبط دسترسی پیدا کرد. همپوشانی نسبی نشان می‌دهد که PubMed به همراه Biological Abstract می‌تواند اطلاعات پوشش داده شده توسط SCI را تأمین کند. بازیابی نسبتاً بالا و همپوشانی نسبی بالای SCI، نه تنها دسترسی به اطلاعات جامع‌تر را فراهم می‌سازد بلکه در وقت و انرژی کاربر نیز صرفه‌جویی به عمل می‌آورد (۸).

ابراهیمی در پژوهشی با عنوان «همپوشانی مقالات مجلات پایگاه‌های اطلاعاتی نمایه و مقالات فرهنگی فارسی (نما) در سال ۱۳۸۴» نشان داد که بیشترین تعداد رکوردها (۵۶/۸ درصد) متعلق به پایگاه داده‌ی نمایه و بیشترین تعداد مجلات (۵۷/۹ درصد) نیز متعلق به این پایگاه داده است. انتخاب پایگاه داده‌ی مورد پژوهش بستگی به نیاز جستجوگر در زمینه‌ی موضوعی خاص دارد. البته با توجه به این که دو پایگاه داده‌ی نمایه و نما تعداد پیشینه‌های زیادی دارند و میزان همپوشانی آن‌ها بسیار اندک است (به جز در مورد موضوع زبان انگلیسی که پایگاه داده‌ی نمایه با پایگاه داده‌ی نما دارای ۸۰ درصد همپوشانی است. در بقیه موضوعات مورد بررسی درصد همپوشانی این دو پایگاه داده از ۲۰ درصد کمتر می‌باشد)، در یک جستجوی جامع، جستجو در هر دو پایگاه داده الزامی می‌باشد (۹).

نتیجه‌ی مشترک پژوهش‌هایی که در مورد همپوشانی پایگاه‌های داده انجام شده است، نشان می‌دهد که همپوشانی، مشخصه‌ی عمومی این پایگاه‌ها می‌باشد (۸). هدف اصلی از نگارش این مقاله، مقایسه‌ی دو پایگاه داده‌ی PubMed و Scopus از نظر پوشش مقالات مرتبط با بیماری‌های قلبی-عروقی است. پایگاه داده‌ی PubMed که دسترسی به Medline را از طریق کتابخانه‌ی ملی پزشکی آمریکا به صورت رایگان فراهم می‌سازد، بیش از ۱۹ میلیون رکورد در رشته‌های پزشکی، پرستاری، دندان‌پزشکی، دامپزشکی و علوم پایه‌ی منتشر شده در آمریکا و ۷۰ کشور جهان که از سال ۱۹۵۵ به بعد انتشار یافتند را تحت پوشش خود دارد. هم‌اکنون ۵۳۹۲ نشریه‌ی جاری در این پایگاه داده، نمایه می‌شوند (۱۰).

نمی‌شوند. آن‌ها ضمن شناسایی بهترین عبارت‌ها برای بازیابی "کار"، در نتیجه‌گیری خود بیان کردند که Medline، Embase و PsycInfo برای جستجوهای سریع مناسب هستند. برای جستجوهای اضافی‌تر، پایگاه‌های داده‌ی Cinahl و OSHROM را پیشنهاد نمودند (۵).

Falagas و همکاران در مطالعه‌ی، پایگاه داده‌ی PubMed را به دلیل آسانی دسترسی به آن، استفاده‌ی راحت و همچنین به دلیل رایگان بودن به عنوان منبعی که برای به دست آوردن اطلاعات زیست پزشکی، بسیار مورد استفاده واقع می‌شوند، معرفی کردند. پایگاه داده‌ی Scopus را به دلیل شامل بودن طیف وسیعی از مجلات نسبت به پایگاه داده‌ی PubMed و WoS توصیه کردند (۶).

Bosman و همکاران در مطالعه‌ی، در بررسی پوشش موضوعی مقالات و مجلات چنین می‌نویسند که علی‌رغم ادعای Elsevier در پوشش صد در صدی پایگاه‌های داده‌ی Embase و Compendex، این پوشش صد در صدی در پایگاه داده‌ی Scopus در زمینه‌های موضوعی مورد بررسی مشاهده نمی‌گردد. در این بررسی مشخص گردید که ۶۰/۳ درصد از مقالات پایگاه داده‌ی PubMed از طریق جستجو در پایگاه داده‌ی Scopus قابل بازیابی هستند (۷).

هاشم‌نیا و ستوده در مطالعه‌ی به بررسی همپوشانی میان سه پایگاه داده‌ی (Biological Abstract) BA، PubMed و (Science Citation Index) SCI پرداختند. آن‌ها به این نتیجه رسیدند که PubMed در دقت و صحت بازیابی اطلاعات بهتر است در حالی که BA و SCI درصد همپوشانی نسبی و بازیابی مرتبط بالایی را نشان می‌دهند. گستره‌ی بالای رکوردهای منحصر به فرد در پایگاه‌های داده‌ی مورد مطالعه، تأیید می‌کند که هیچ یک از پایگاه‌های داده به تنهایی مدارک مرتبط به موضوعات را نمایه نمی‌کنند. برای داشتن یک جستجوی جامع با بیشترین منابع مرتبط، دسترسی به بیش از یک پایگاه داده ضروری به نظر می‌رسد. همچنین وجود همپوشانی بالا در بعضی از موضوعات نشان می‌دهد که در غیاب دیگر پایگاه‌های داده، می‌توان با

پایگاه داده‌ی PubMed و Scopus، پایگاه داده‌ی ارجح شناسایی و معرفی می‌گردد.

### روش بررسی

این پژوهش، از نوع توصیفی- کاربردی است. گردآوری اطلاعات لازم برای انجام این پژوهش از بهمن ماه سال ۱۳۸۷ تا تیر ماه ۱۳۸۸ به طول انجامید. برای دستیابی به داده‌های پژوهش یعنی مقالات مرتبط با زمینه‌ی بیماری‌های قلبی- عروقی انگلیسی زبان منتشر شده در سال ۲۰۰۸ که در پایگاه داده‌های PubMed و Scopus نمایه گردیدند، نیاز به توصیفگرهایی بود که با استفاده از آن‌ها باید به بازیابی این مقالات مبادرت ورزیده می‌شد. بدین منظور ابتدا بیماری‌های قلبی- عروقی خود به بیماری‌های خاص‌تری تجزیه گردید. این بیماری‌ها شامل سکته‌ی قلبی- مغزی، پرفشاری خون و حوادث عروقی- مغزی بودند. آنگاه برای هر یک از این بیماری‌ها تمامی توصیفگرهایی که در شاخه‌های موضوعی این بیماری‌ها بودند، از اصطلاح‌نامه Emtree استخراج شدند (۱۳). انتخاب توصیفگرها از اصطلاح‌نامه Emtree به دلایل زیر انجام شد (۱۴). در اصطلاح‌نامه Emtree، توصیفگرها تا جایی که ممکن است با استفاده از زبان طبیعی ساخته می‌شوند به گونه‌ای که در عناوین مقالات می‌توان آن‌ها را به همان صورت یافت. اصطلاح‌نامه‌ی Emtree شامل اکثر توصیفگرهای Mesh است.

بعد از مشخص شدن توصیفگرها با توجه به ویژگی‌های در نظر گرفته شده برای مقالات، به جستجو در دو پایگاه داده پرداخته شد. به منظور پرهیز از بازیابی حجم زیاد مقالات و همچنین یافتن مقالات مرتبط با موضوعات مورد نظر، توصیفگر به کار رفته در عنوان مقاله مورد جستجو قرار گرفت. از طرف دیگر در این جستجوها تنها مقالاتی بازیابی شدند که Original Article بودند. به عبارت دیگر مقالات مروری (Review)، یادداشتی (Comment) یا تاریخچه‌ای (Historical) و مقالاتی از این نوع حذف شدند. داده‌های کتاب‌شناختی مقالات بازیابی شده برای استفاده در مراحل

این در حالی است که پایگاه داده‌ی Scopus با پوشش دادن ۳۸ میلیون رکورد از بیش از ۱۸۰۰۰ عنوان از ۵۰۰۰ ناشر بین‌المللی که ۱۶۴۰۰ عنوان نشریه علمی را هم در برمی‌گیرند، بزرگترین پایگاه داده‌ی نمایه‌سازی و چکیده‌نویسی دنیا به شمار می‌رود. این پایگاه داده در علوم پزشکی فراتر از ۶۸۰۰ عنوان (۱۰۰ درصد عناوین مدلاین را شامل می‌گردد). در علوم زیستی در حدود ۴۳۰۰ عنوان، در علوم فیزیک ۷۲۰۰ عنوان و در علوم اجتماعی در حدود ۵۳۰۰ عنوان را از سال ۱۹۶۶ تحت پوشش قرار می‌دهد. استفاده از این پایگاه داده، مستلزم پرداخت حق اشتراک است (۱۱).

با به دست آوردن میزان همپوشانی بین این دو پایگاه داده در زمینه‌ی بیماری‌های قلبی- عروقی می‌توان راجع به استفاده از پایگاه داده‌ی Scopus و یا استفاده‌ی رایگان از پایگاه داده‌ی PubMed بهتر تصمیم‌گیری کرد. به این ترتیب معیاری برای انتخاب پایگاه داده‌ی مناسب تعیین می‌شود. علاوه بر ارتقای کیفیت تحقیقات به سبب دسترسی به ادبیات بیشتر در زمینه‌ی پژوهش‌ها، استفاده از پایگاه داده‌ی مناسب منجر به صرفه‌جویی در وقت محققین در هنگام بازیابی پژوهش‌های مرتبط با موضوعات پژوهششان می‌گردد.

بیماری‌های قلبی- عروقی در کشورهای توسعه یافته و در حال توسعه اصلی‌ترین عامل مرگ و میر و ناتوانی هستند (۱۲). همین امر نیاز به پژوهش در زمینه‌ی این بیماری را برای دانشجویان، متخصصان و محققان رشته‌های پزشکی ضروری می‌سازد. لازمه‌ی انجام این پژوهش‌ها، کسب اطلاعاتی در مورد پیشینه‌ی آن‌هاست. این مهم با فراهم آوردن متون، منابع چاپی و الکترونیکی مناسب در زمینه‌ی این بیماری‌ها محقق می‌گردد. وظیفه‌ای که بر عهده‌ی کتابداران و متخصصان اطلاع‌رسانی است، این است که با تشخیص منبع مناسب ضمن تأمین نزدیک‌ترین و مرتبط‌ترین مطالب، هم در وقت کاربران و هم در هزینه‌های مصرفی پژوهشگران تا حد امکان صرفه‌جویی به عمل آورند.

از این رو در این نوشتار پس از بررسی همپوشانی منابع اطلاعاتی موجود در زمینه‌ی بیماری‌های قلبی- عروقی در دو

از پایگاه‌های داده اقدام نمود.

در سنجش همپوشانی نسبی حجم هر یک از پایگاه‌های داده مد نظر قرار می‌گیرد. همپوشانی نسبی بر مبنای اندازه‌ی هر یک از پایگاه‌های داده با استفاده از فرمول‌های زیر محاسبه می‌شود (۱۶، ۱۵، ۲).

$$\%overlap_{BinA} = 100 * \frac{n(A \cap B)}{n(A)}$$

$$\%overlap_{AinB} = 100 * \frac{n(A \cap B)}{n(B)}$$

بدین ترتیب در مثال فوق همپوشانی نسبی پایگاه داده‌ی

A نسبت به پایگاه داده‌ی B برابر با  $\frac{10}{15} * 100 = 66.7\%$  یعنی ۶۶/۷ درصد است در حالی که همپوشانی پایگاه داده‌ی B نسبت به پایگاه داده‌ی A برابر با  $\frac{10}{45} * 100 = 22.2\%$  یعنی ۲۲/۲ درصد است.

درجه‌ی آزادی مرکب برای توصیف میزان تفاوت و پراکندگی میان اجزا منابع ردیف دوم به کار می‌رود. منابع ردیف دوم به منابعی گفته می‌شود که به جای پاسخ‌گویی مستقیم به سوالات، کاربر را به سمت اطلاعات مانند نمایه‌نامه‌ها و چکیده‌نامه‌ها هدایت می‌کنند (۱۷). این مفهوم به تعیین اندازه‌ی نسبی منابع و در نتیجه به تغییر جهت مفهوم همپوشانی سنتی می‌پردازد. Gluck این روش را جایگزین روش سنتی کرد. وی با استفاده از فرمول‌های زیر، همپوشانی را برای دو یا چند منبع ثانوی مختلف محاسبه نمود (۱۴، ۲).

موقعی که A زیر مجموعه‌ای از B نباشد (۱۵، ۱۴).

$$SF(A, B) = (1 - \frac{n(A \cap B)}{n(A \cup B)}) * (\frac{MIN}{MAX})$$

موقعی که A زیر مجموعه‌ای از B باشد.

$$SF(A, B) = (1 - \frac{MIN}{MAX})$$

MIN: تعداد اعضای مجموعه‌ای که کمتر عضو دارد.

MAX: تعداد اعضای مجموعه‌ای که بیشتر عضو دارد.

نکته: برای تمام مقادیر A و B  $0 \leq SF(A, B) \leq 1$

بعدی به صورت فایل‌های قابل خواندن برای نرم‌افزار Reference Manager ذخیره گردیدند. بعد از وارد ساختن فایل‌های داده‌های کتاب‌شناختی در نرم‌افزار ذکر شده، داده‌های اولیه شامل تعداد مقالات مشترک و تعداد مقالات منحصر به هر پایگاه داده که برای محاسبه‌ی درصدهای همپوشانی مورد نیاز بودند، برای هر بیماری به صورت جداگانه شمارش شدند و جدول داده‌ها تکمیل گردید. این داده‌ها جهت انجام تجزیه و تحلیل‌های بعدی در نرم‌افزار Excel به صورت یک فایل کامپیوتری ذخیره گردیدند. از این فایل برای محاسبه‌ی درصدهای همپوشانی دو پایگاه داده‌ی مورد بررسی، استفاده شد. همپوشانی میان این دو پایگاه داده از سه نقطه نظر همپوشانی سنتی، همپوشانی نسبی و درجه‌ی آزادی مرکب مورد بررسی قرار گرفت.

همپوشانی در مفهوم عام آن به موجود بودن یک منبع اطلاعاتی خاص در دو مکان متفاوت یا در قالب‌های مختلف اطلاق می‌گردد. دو شاخص همپوشانی مهم که در ارزیابی پایگاه‌های داده‌ها از آن‌ها استفاده می‌شوند، همپوشانی سنتی (Traditional Overlap) و همپوشانی نسبی (Relative Overlap) می‌باشند.

همپوشانی سنتی به محاسبه تعداد منابع مشترک میان دو یا چند پایگاه داده با استفاده از فرمول زیر می‌پردازد.

$$\%overlap = 100 * \frac{n(A \cap B)}{n(A \cup B)}$$

اگر دو پایگاه داده‌ی A و B، هر یک به ترتیب دارای ۴۵ و ۱۵ رکورد باشند که از آن میان ۱۰ رکورد در هر دو مشترک باشد، میزان همپوشانی آن‌ها بر اساس فرمول فوق ۲۰ درصد می‌باشد.

$$\%Overlap = 100 * (\frac{10}{50}) = 20\%$$

این مثال عدم حساسیت همپوشانی سنتی را نسبت به اندازه‌ی پایگاه‌های داده نشان می‌دهد. بورن و بیرمن از این موضوع آگاه بودند و می‌دانستند که میزان همپوشانی سنتی، اندازه‌ی نسبی پایگاه‌های داده را نشان نمی‌دهد. بیرمن برای جبران این نقیصه به محاسبه‌ی همپوشانی نسبی برای هر یک

## یافته‌ها

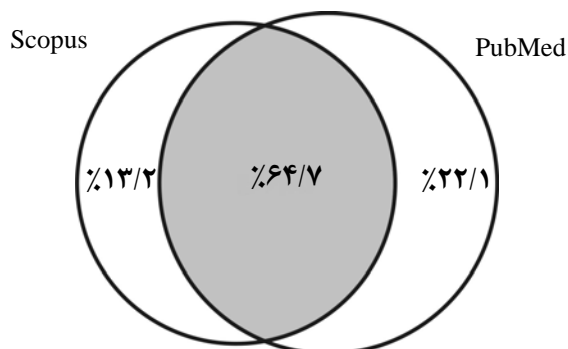
زمینه‌ی موضوعی بیماری‌های قلبی- عروقی بر اساس توصیفگرهای استخراج شده از اصطلاح‌نامه‌ی Emtree در سه بیماری سکنه قلبی- مغزی، پرفشاری خون و حوادث عروقی- مغزی مورد بررسی قرار گرفت.

از ۹۸۹۹ عنوان مقاله‌ی به دست آمده در هر دو پایگاه داده برای زمینه‌ی موضوعی مورد بررسی ۸۵۹۳ عنوان مقاله در پایگاه داده‌ی PubMed و ۷۷۱۳ عنوان مقاله در پایگاه داده‌ی Scopus، ۶۴۰۷ عنوان مقاله مشترک در هر دو پایگاه داده، ۲۱۸۶ عنوان مقاله منحصر به پایگاه داده‌ی PubMed و ۱۳۰۶ عنوان مقاله منحصر به پایگاه داده‌ی Scopus بودند.

نمودار ۱، درصد مقالات یافت شده مشترک دو پایگاه داده (قسمت هاشور خورده) و مقالات منحصر به هر پایگاه داده (قسمت‌های بدون هاشور) را در زمینه‌ی موضوعی مورد بررسی نشان می‌دهد.

بر اساس اطلاعات مندرج در جدول ۱ برای هر سه بیماری مورد بررسی در زمینه‌ی موضوعی بیماری‌های قلبی- عروقی، پایگاه داده PubMed تعداد مقالات بیشتری را ارائه می‌دهد. بیشترین تعداد مقالات بازیابی شده برای بیماری سکنه قلبی- مغزی است. این مطلب در هر دو پایگاه داده PubMed و Scopus صادق است. کمترین تعداد مقالات بازیابی شده در هر دو پایگاه داده، به حوادث عروقی- مغزی تعلق دارد. برای هر سه بیماری مورد جستجو، تعداد مقالات منحصر به فرد بیشتری در

پایگاه داده‌ی PubMed نسبت به پایگاه داده‌ی Scopus به دست آمد. بیشترین تعداد مقالات مشترک بین دو پایگاه داده PubMed و Scopus متعلق به بیماری پرفشاری خون و کمترین تعداد متعلق به حوادث عروقی- مغزی است.



نمودار ۱: همپوشانی میان دو پایگاه داده‌ی PubMed و Scopus در زمینه‌ی موضوعی بیماری‌های قلبی- عروقی

بر اساس اطلاعات جدول ۲ بیشترین درصد همپوشانی سستی در بیماری پرفشاری خون دیده می‌شود. مقایسه‌ی همپوشانی‌های نسبی بیانگر آن است که در کلیه‌ی موارد، درصد همپوشانی نسبی پایگاه داده‌ی PubMed بیشتر از درصد همپوشانی نسبی پایگاه داده‌ی Scopus است. مقدار شاخص درجه‌ی آزادی مرکب در همه موارد کمتر از ۰/۵ است که نشان از آن دارد که فاصله بین دو منبع از یکدیگر زیاد نیست.

جدول ۱: فراوانی مقالات بازیابی شده‌ی منحصر به فرد و مشترک در دو پایگاه داده‌ی PubMed و Scopus در سال ۲۰۰۸ میلادی

بیماری‌های مورد جستجو	فراوانی مقالات بازیابی شده	تعداد مقالات PubMed	تعداد مقالات Scopus	تعداد مقالات منحصر PubMed	تعداد مقالات منحصر Scopus	تعداد مقالات مشترک دو پایگاه داده	تعداد کل مقالات
سکنه‌ی قلبی مغزی	۳۲۷۲	۳۲۵۱	۸۰۸	۷۸۷	۲۴۶۴	۴۰۵۹	
پرفشاری خون	۳۳۸۶	۲۸۱۲	۸۷۸	۳۰۴	۲۵۰۸	۳۶۹۰	
حوادث عروقی- مغزی	۲۱۸۳	۱۸۷۰	۵۶۲	۲۴۹	۱۶۲۱	۲۴۳۲	

جدول ۲: درصد‌های همپوشانی سنتی، نسبی دو پایگاه داده‌ی PubMed و Scopus بر اساس مقالات بازیابی شده‌ی زمینه‌ی موضوعی بیماری‌های قلبی-عروقی در سال ۲۰۰۸ میلادی

درصد‌های همپوشانی	همپوشانی نسبی	همپوشانی نسبی	همپوشانی سنتی	بیماری‌های مورد جستجو
درجه‌ی آزادی مرکب	Scopus	PubMed		
۰/۳۹	۷۵/۳	۷۵/۸	۶۰/۷	سکته‌ی قلبی- مغزی
۰/۲۷	۷۴/۱	۸۹/۲	۶۸	پرفشاری خون
۰/۲۹	۷۴/۳	۸۶/۷	۶۶/۷	حوادث عروقی- مغزی

### بحث

سکته‌ی قلبی- مغزی، پرفشاری خون و حوادث عروقی- مغزی مورد مطالعه در این پژوهش نیز به ترتیب برابر با ۶۰/۷، ۶۸ و ۶۷ درصد به دست آمد (۷).

اکثر قریب به اتفاق عناوین مجلاتی که مقالات آن‌ها در پایگاه داده‌ی PubMed نمایه می‌گردند، مقالات مرتبط با موضوعات پزشکی تشکیل می‌دهند. در حالی که در پایگاه داده‌ی Scopus، عناوین مجلات از تنوع بسیار زیادی بهره‌مند هستند و تنها بخشی از این مجلات مختص به رشته‌های پزشکی هستند. شاید به همین دلیل است که علی‌رغم آن که تعداد مجلات بیشتری در پایگاه داده‌ی Scopus نمایه می‌شوند، تعداد مقالات کمتری در این پایگاه داده در مقایسه با پایگاه داده‌ی PubMed بازیابی گردیدند. اصولاً پایگاه داده‌ی PubMed به عنوان یک پایگاه داده‌ی مطرح و شناخته شده در زمینه‌های مختلف علوم پزشکی محسوب می‌گردد.

### نتیجه‌گیری

چنانچه کتابخانه یا مراکز اطلاع‌رسانی دچار چالش‌های مالی و محدودیت‌های بودجه‌ای باشند، ناچار به اتخاذ تصمیمات بهینه در سفارش و تکمیل منابع خود می‌باشند. درصد‌های همپوشانی بین پایگاه‌های مختلف، شاخص‌های کمی مهمی هستند که می‌توانند به مدیران کتابخانه‌ها در تصمیم‌گیری‌های خود کمک کنند. در مورد سه بیماری مورد بررسی در این پژوهش، جستجو در پایگاه داده‌ی PubMed، مقالات بیشتری نسبت به پایگاه داده Scopus به دست می‌دهد. علاوه بر این دسترسی آسان به PubMed و عدم نیاز به پرداخت هزینه‌ی اشتراک برای PubMed، کاربران و کتابداران را به سمت استفاده از این پایگاه داده هدایت می‌کند. عدم نیاز به پرداخت حق اشتراک

یافته‌های پژوهش نشان می‌دهند که درصد‌های همپوشانی سنتی برای هر سه بیماری مطرح شده در این پژوهش، بالای ۶۰ درصد است که همپوشانی بالایی محسوب می‌گردد. همپوشانی بالا بدین معناست که با جستجو در یک پایگاه داده می‌توان به نتیجه مطلوب دست یافت. علاوه بر این همپوشانی بالا نشان‌دهنده‌ی تکرار یک مقوله است. از طرف دیگر تکرار زیاد نیز خود نشان‌دهنده انجام کار کافی در یک رشته است. نقاط دستیابی به یک مطلب را افزایش و تلاش کاربر را کاهش می‌دهد. بر این اساس هزینه‌های غیر مستقیم جستجوی مطالب را کاهش می‌دهد (۱۴). بررسی مقدار شاخص درجه‌ی آزادی مرکب به عنوان یک معیار عدم تشابه این نتیجه‌گیری را تأیید می‌کند، چرا که در هر سه بیماری مورد بررسی، مقدار این شاخص کمتر از ۰/۵ است. این مطلب نشان می‌دهد که فاصله این دو پایگاه داده از یکدیگر زیاد نیست. در هر سه بیماری مورد بررسی درصد‌های همپوشانی نسبی پایگاه داده‌ی PubMed بیشتر از درصد‌های همپوشانی نسبی پایگاه داده‌ی Scopus است.

Falagas و همکاران، پایگاه داده‌ی PubMed را به دلیل دسترسی آسان، استفاده‌ی راحت و همچنین به دلیل رایگان بودن به عنوان منبعی که برای به دست آوردن اطلاعات زیست پزشکی، بسیار مورد استفاده واقع می‌شود، معرفی کردند. یافته‌های این پژوهش نیز استفاده از PubMed را به دلیل دسترسی آسان، رایگان بودن و همچنین در اختیار قرار دادن تعداد مقالات منحصر به فرد بیشتر توصیه می‌کند (۶). Bosman و همکاران میزان همپوشانی مقالات پایگاه داده‌ی PubMed توسط Scopus را در مطالعه‌ی خود، ۶۰/۳ درصد به دست آوردند. میزان همپوشانی برای بیماری‌های

نکنته‌ی دیگری که باید بدان اشاره نمود این است که مقایسه‌ی پایگاه‌های داده با استفاده از شاخص‌های همپوشانی ارجحیت یک پایگاه داده را نسبت به پایگاه داده‌ی دیگر از جنبه‌های کمی (کمیت مقاله‌ها و منابع) تعیین می‌کند. این مطلب لزوماً بدین معنا نیست که پایگاه داده‌ای که بر اساس این شاخص‌ها برتر شناخته شد، از نظر کیفیت مقالات و منابع نیز بر پایگاه داده‌ی دیگر برتری داشته باشد. بررسی کیفیت محتوای پایگاه‌های داده، خود مقوله‌ای است که نیاز به بررسی همه‌جانبه‌ی مقالات و منابع نمایه شده در آن‌ها دارد و انجام تحقیقات گسترده‌تر و دقیق‌تری را لازم دارد.

بدین معنی است که حتی بیرون از محیط‌های دانشگاهی نیز می‌توان به پایگاه داده‌ی PubMed دسترسی پیدا کرد. شایان ذکر است که حکم بر برتری جستجو در پایگاه داده‌ی PubMed نسبت به پایگاه داده‌ی Scopus در این پژوهش تنها برای بازیابی مطالب مرتبط با بیماری‌های قلبی-عروقی است. بدیهی است برای تعیین ارجحیت این دو پایگاه داده در سایر زمینه‌های موضوعی، نیاز به انجام مطالعات دیگری دارد. به طور کلی نویسندگان مقاله‌ی حاضر لزوم انجام مطالعات بیشتری را در سایر زمینه‌های موضوعی پزشکی در این دو پایگاه داده، توصیه می‌نمایند تا مشخص شود که کدام پایگاه داده نتایج بهتر و بیشتری به دست می‌دهند.

## References

1. Markham JW. Bibliographic Database Comparisons [Online]. 2000 [Cited 2008 Dec 5]; Available from: URL: <https://darchive.mblwhoilib.org/bitstream/1912/1662/1/proc91087.pdf>
2. Gluck M. A Review of Journal Coverage Overlap with an Extension to the Definition of Overlap. *Journal of the American Society for Information Science* 1990; 41(1): 43-60.
3. Caldwell J, Ellingson C. A Comparison of Overlap: ERIC and Psychological Abstracts. *Database* 1979; 2(2): 62-7.
4. Mychko-Megrin AY. A comparison of biomedical databases. *Bull Med Libr Assoc* 1991; 79(3): 302-8.
5. Haafkens J, Moerman C, Schuring M, Van Dijk F. Searching bibliographic databases for literature on chronic disease and work participation. *Occup Med (Lond)* 2006; 56(1): 39-45.
6. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 2008; 22(2): 338-42.
7. Bosman J, Mourik IV, Rasch M, Sleverts E, Verhoeff H. Scopus reviewed and compared: The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar [Online]. 2006 [Cited 2009 Jul 1]; Available from: URL: <http://info.scopus.com/news/coverage/utrecht.pdf/>
8. Hashem Nia S, Sotoudeh H. The Study of Coverage of Laccase Enzyme Literature by Databases: Measuring Algorithmic and Cognitive Relevance. *Informology* 2004; 2(1): 175-7.
9. Ebrahimi H. An examination of the degree of overlap in Namayeh and Nama databases in 2005. *Fasname-Ketab* 2007; 18(4): 299-308.
10. PubMed Overview [Online]. 2008 May 1 [Cited 2009 Nov 9]; Available From: URL: <http://www.ncbi.nlm.nih.gov/corehtml/query/static/overview.html#Introduction/>
11. Elsevier Scopus [Online]. 2009 [Cited 2009 Nov 9]; Available From: URL: <http://info.scopus.com/detail/what/>
12. Atarchi MS, Mohammadi S, Aghili Nejad MA, Asghari Roudsari E. Cardiovascular Diseases Impairment (Coronary Artery Diseases, Hypertension, Arrhythmia). *Scientific Journal of Forensic Medicine* 2007; 13(2): 114-22.
13. Khani S. Comparison of Thesaurus Structure of PubMed & Embase Databases with NISO Standard of the Thesauri and Survey of These Databases Indexing Methods, [MSc Thesis] Tehran: School of Management & Medical Information Science, Iran Medical Sciences University; 2006.
14. Elsevier Pharma Development Group. Emtree: The Life Science Thesaurus [Online]. 2009 [Cited 2009 Nov 9]; Available from: URL: [http://embase.com/info/UserFiles/Files/emtree\\_white\\_paper.pdf/](http://embase.com/info/UserFiles/Files/emtree_white_paper.pdf/)
15. Davarpanah MR, Dayani MH. Scientific Information Seeking in Printed and Electronic Resources. Tehran: Dabizesh. Chapar Publication; 2006. p. 72-5.
16. Egghe L, Goovaerts M. A note on measuring overlap. *Journal of Information Science* 2007; 33(2): 189-95.
17. Moradi N. Study Reference: Knowledge Services and Reference Books. Tehran: Farhange-Moaser Publication; 1993.



## A Survey on Traditional Overlap, Relative Overlap and Synthetic Degrees of Freedom between PubMed and Scopus in the Cardiovascular Disease Field\*

Mohammad Reza Alibeig<sup>1</sup>; Roohangiz Jamshidi Orak, PhD<sup>2</sup>; Leila Asghari Heineh Abad<sup>3</sup>

### Abstract

**Introduction:** Determining the overlap rates is a quantitative method to compare database contents. In this survey, traditional overlap, relative overlap and synthetic degrees of freedom between PubMed as a free database and Scopus as a subscription-based one in the cardiovascular disease field have been determined.

**Methods:** This descriptive applied research was done during January to June 2009. First, the descriptors related to three cardiovascular diseases, namely stroke, high Blood pressure and cerebrovascular accidents, were obtained from Emtree Thesaurus. Then, original articles on the three mentioned problems were retrieved from PubMed and Scopus by searching these descriptors. Primary raw data needed for overlap calculation were obtained using Reference Manager Software. Data analysis was done in Microsoft Excel.

**Results:** A total of 9899 articles on cardiovascular diseases were retrieved from both databases, including 8593 and 7713 articles from PubMed and Scopus, respectively. There were 6407 common titles while 2186 and 1306 titles belonged only to PubMed and Scopus, respectively. The traditional overlaps were 68%, 66.7% and 60.7% for high blood pressure, cerebrovascular accidents and stroke, respectively. Relative overlap between PubMed and Scopus for stroke, high blood pressure and cerebrovascular accidents were 75.8% vs. 75.3%, 89.2% vs. 74.1% and 86.7% vs. 74.3%, respectively. Synthetic degrees of freedom were smaller than 0.5 for the three considered diseases.

**Conclusion:** Due to PubMed's higher relative overlaps versus Scopus's, more user and librarians are led to use PubMed. Larger numbers of retrieved unique articles for each cardiovascular disease from PubMed as compared with Scopus, along the fact that it is free, makes PubMed preferable.

**Keywords:** Databases; PubMed; Cardiovascular Diseases.

**Type of article:** Original article

*Received: 28 Jul, 2009*

*Accepted: 3 Jul, 2011*

**Citation:** Alibeig MR, Jamshidi Orak R, Asghari Heineh Abad L. **A Survey on Traditional Overlap, Relative Overlap and Synthetic degrees of Freedom between PubMed and Scopus in the Cardiovascular Disease Field.** Health Information Management 2011; 8(3): 353.

\* This article was derived from an MSc thesis.

1. Lecturer, Medical Library and Information Science, Tehran University of Medical Sciences, Tehran, Iran.

2. Assistant Professor, Biostatistics, Tehran University of Medical Sciences, Tehran, Iran.

3. MSc Student, Medical Library and Information Science, Tehran University of Medical Sciences, Tehran, Iran. (Corresponding Author) Email: tabrizleila@gmail.com