آمار ارجاع

رابرت ادلر، جان ایوینگ، پیتر تیلر[۱]

ترجمه:

شیرین گلچی[۲]

خلاصه:

این گزارش در رابطه با استفاده و سوء استفاده از داده‌های ارجاع برای ارزیابی پژوهش می‌باشد. امروزه این عقیده که ارزیابی پژوهش باید از طریق روش‌های ساده و عینی انجام شود بسیار رایج و متداول است. روش‌های ساده و عینی به‌طور کلی بیبلیومتریک‌ها یعنی داده‌های ارجاع و آماره‌های به‌دست آمده از آن‌ها تفسیر می‌شوند. این باور وجود دارد که آمار ارجاعات ذاتاً صحیح‌ترند زیرا اعداد ساده را جایگزین قضاوت‌های پیچیده می‌کنند و لذا بر شخصی بودن احتمالی داوری توسط همکاران[۳] فائق می‌آیند.

- تکیه بر آماره‌ها زمانی که آنها را به صورت نامناسب مورد استفاده قرار دهیم، صحیح نمی‌باشد. در واقع آماره‌ها در صورتی که به درستی به‌کار گرفته نشوند و یا به درستی درک نشوند می‌توانند گمراه کننده باشند. به نظر می‌رسد که بسیاری از بیبلیومتریک‌های جدید برای تفسیر آمار ارجاعات بر تجربه و شهود تکیه دارند.
- در حالی‌که اعداد عینی به نظر می‌رسند عینی بودن آنها می‌تواند موهومی باشد. معنی ارجاع می‌تواند حتی شخصی‌تر از داوری توسط همکاران باشد. از آنجا که این شخصی بودن در ارجاعات مشهود نمی‌باشد، افرادی که داده‌های ارجاع را مورد استفاده قرار می‌دهند با احتمال کمتری محدودیت‌های خود را درک می‌کنند.
- تکیه کامل بر داده‌های ارجاع در بهترین حالت درکی ناقص و اغلب سطحی از پژوهش به دست می‌دهد. درکی که تنها زمانی معتبر است که توسط داوری‌های دیگر تکمیل گردد. اعداد ذاتاً آنقدر برتر نیستند که داوری محسوب شوند.

استفاده از داده‌های ارجاع برای ارزیابی پژوهش نهایتاً به این معنی است که از آماره‌های مبتنی بر ارجاعات برای رتبه‌بندی مجلات، مقالات، افراد، برنامه‌ها و نظام‌ها استفاده شود. ابزارهای آماری مورد استفاده برای رتبه‌بندی اغلب

[۱] این مقاله ترجمه گزارشی از کار مشترک اتحادیه بین المللی ریاضیات (IMU)، انجمن بین المللی ریاضیات صنعتی و کاربردی (ICIAM) و موسسه آمار ریاضی (IMS) است که توسط رابرت آدلر (Rober Adler)، جان ایوینگ (John Ewing) و پیتر تیلور (Peter Taylor) نوشته شده است.

[۲] دانشجوی کارشناسی ارشد آمار، پژوهشکده آمار

[۳] peer review

در معرض درک نادرست و استفاده نادرست می‌باشند.

- برای رتبه‌بندی مجلات اغلب اوقات ضریب تأثیر[۴] مورد استفاده قرار می‌گیرد. ضریب تأثیر یک میانگین ساده است که از توزیع ارجاعات برای مجموعه‌ای از مقالات داخل مجله به دست می‌آید. میانگین، میزان کمی از اطلاعات این توزیع را در بر می‌گیرد و آماره نسبتاً خامی است. علاوه بر این مسائل ابهام آمیز بسیاری در داوری مجلات توسط ارجاعات وجود دارند و هر مقایسه‌ای از مجلات با استفاده از ضریب تأثیر باید با احتیاط صورت بگیرد. استفاده از ضریب تأثیر به تنهایی برای داوری یک مجله مانند استفاده از وزن به تنهایی برای قضاوت در مورد سلامتی یک فرد است.

- برای مقایسه‌ی مقالات به جای تکیه بر تعداد واقعی ارجاعات هر مقاله اغلب ضریب تأثیر مجله‌ای که مقاله در آن چاپ می شود قرار داده می‌شود. این باور وجود دارد که ضریب تأثیر بالاتر به معنی تعداد ارجاعات بیشتر است. اما اغلب اوقات این طور نیست. این یک استفاده غلط فراگیر از آمار است که در هر مکان و زمانی که رخ دهد باید با آن برخورد صورت بگیرد.

- برای افراد (دانشمندان) مقایسه سابقه کامل ارجاعات امری دشوار است. در نتیجه، تلاش‌های زیادی برای یافتن آماره ساده‌ای که سابقه پیچیده ارجاعات یک دانشمند را در یک عدد در بر بگیرد، صورت گرفته است. مهم‌ترین این آماره‌ها شاخص h[۵] است که به نظر می‌رسد محبوبیت زیادی کسب کرده است. اما حتی بررسی سطحی این شاخص و اَشکال دیگر آن نشان می‌دهد که این تلاش‌ها برای درک سابقه پیچیده ارجاعات کامل نیستند. این شاخص‌ها در حالیکه میزان اندکی از اطلاعات توزیع ارجاعات دانشمند را در بر می‌گیرند، اطلاعات حساسی را که برای ارزیابی پژوهش مورد نیاز است از دست می‌دهند.

اعتبار آماره‌هایی از قبیل ضریب تأثیر و شاخص h نه به درستی درک شده و نه به خوبی مورد مطالعه قرار گرفته است. ارتباط این آماره‌ها با کیفیت پژوهش گاهی بر اساس «تجربه» پایه‌گذاری می‌شود. توجیه تکیه بر آنها این است که «در دسترس» می‌باشند. مطالعات اندکی که در مورد این آماره‌ها انجام شده بیشتر بر روی نشان دادن همبستگی آنها با دیگر معیارهای کیفیت متمرکز شده‌اند، تا روی تعیین چگونگی بهترین شکل استخراج اطلاعات مفید از داده‌های ارجاع.
ما آمار ارجاعات را به‌عنوان ابزاری برای ارزیابی کنار نمی‌گذاریم؛ داده‌ها و آماره‌های ارجاع می‌توانند اطلاعات ارزشمندی فراهم سازند. ما تصدیق می‌کنیم که ارزیابی باید کاربردی باشد و به همین دلیل آماره‌های ارجاع که به‌سادگی به‌دست می آیند مطمئناً بخشی از این فرایند خواهند بود. اما داده‌های ارجاع تنها تصویری محدود و ناقص از کیفیت پژوهش به دست می‌دهد و آماره‌های به‌دست آمده از آن گاهی به درستی مورد درک و استفاده قرار نمی‌گیرند.

[۴] Impact Factor
[۵] h-index

پژوهش مهم‌تر از آن است که ارزش آن تنها با استفاده از یک ابزار معمولی سنجیده شود.
آرزو داریم افرادی که دستی در ارزیابی دارند مجموع و جزئیات این گزارش را به منظور نه تنها درک محدودیت‌های آمار ارجاع بلکه حتی استفاده بهتر از آن مورد مطالعه قرار دهند. اگر استانداردهای بالایی برای علم قرار می‌دهیم مطمئناً باید برای ارزیابی کیفیت آن نیز استانداردهایی به همان اندازه بالا در نظر بگیریم.

۱ مقدمه

پژوهش علمی از اهمیت زیادی برخوردار است. پژوهش اساس پیشرفت در دنیای مدرن است و امید حل مشکلات به نظر لاینحلی را که بشر با آنها رو به روست، از محیط زیست تا جوامع در حال گسترش، فراهم می‌کند. به همین دلیل دولت‌ها و مؤسسات در سراسر جهان از پژوهش علمی حمایت مالی قابل توجهی به عمل می‌آورند. طبیعتاً آنها می‌خواهند بدانند که پولشان هوشمندانه سرمایه‌گذاری می‌شود؛ می‌خواهند کیفیت پژوهشی را که برای آن هزینه می‌کنند ارزیابی کنند تا تصمیمات آگاهانه‌ای در مورد سرمایه‌گذاری‌های بعدی اتخاذ نمایند.

این امر امری تازه نیست: سال‌های زیادی است که پژوهش مورد ارزیابی قرار می‌گیرد. آنچه تازه است این تصور است که ارزیابی باید «ساده و عینی» باشد و اینکه چنین ارزیابی‌هایی می‌توانند با تکیه بر متریک (آماره)های به‌دست آمده از داده‌های ارجاع به جای استفاده از روش‌های متنوع، شامل قضاوت‌هایی که خود دانشمندان انجام می‌دهند، حاصل شود. بند بعد که از یک گزارش جدید انتخاب شده،این نظریه را کاملاً اظهار می‌دارد:

قصد دولت (انگلیس) این است که روش کنونی برای تعیین کیفیت پژوهش دانشگاهی روش ارزیابی پژوهش انگلستان[۶] باید پس از کامل شدن چرخه بعدی در سال ۲۰۰۸ جایگزین گردد. تمرکز سیستم جدید، متریک‌ها به جای داوری توسط همکاران خواهد بود و انتظار می‌رود بیبلیومتریک‌ها با استفاده از تعداد مقالات مجلات و ارجاعات آنها شاخص کیفیت مرکزی در این سیستم باشند[۱۲].

افرادی که برای این عینیت ساده بحث می‌کنند، باور دارند که پژوهش خیلی مهم‌تر از آن است که بر داوری‌های شخصی تکیه شود. آنها باور دارند که متریک‌های مبتنی بر ارجاعات، فرایند رتبه‌بندی را شفاف می‌سازند و ابهامات اصلی سایرشکل‌های ارزیابی را حذف می‌کنند. همچنین باور دارند که متریک‌هایی که با دقت انتخاب شوند مستقل و فارغ از اریبی‌اند. آنها بیش از همه اعتقاد دارند که چنین متریک‌هایی امکان مقایسه‌ی کلیه‌ی اجزای پژوهش (مجلات، مقالات، افراد، برنامه‌ها و حتی نظام‌ها) را به‌صورت ساده و مؤثر، بدون استفاده از داوری توسط همکاران فراهم می‌کنند. اما صحت این اعتقاد به دقت، استقلال و سودمند بودن

[۶] The UK Research Assessment Exercise(RAE)

متریک‌ها اثبات نشده است.

- اولاً، دقت این متریک‌ها موهومی است. این یک مثل متعارف است که در صورتی که آمار به شکل نامناسب مورد استفاده قرار بگیرد می تواند دروغ بگوید. سوء استفاده از آمار ارجاعات امری شایع و فاحش است. علی‌رغم تلاش‌های مکرر برای اخطار علیه چنین سواستفاده‌هایی (به عنوان مثال استفاده‌ی غلط از ضریب تأثیر)، دولت‌ها، مؤسسات و خود دانشمندان به نتیجه‌گیری‌های غیر قابل توجیه و حتی غلط از کاربرد نادرست آمار ارجاعات ادامه می‌دهند.
- ثانیاً، تکیه‌ی کامل بر متریک‌های مبتنی بر ارجاعات نوعی داوری را جایگزین نوع دیگری می‌کند. به‌جای داوری شخصی توسط همکاران، تفسیر شخصی از مفهوم ارجاع صورت می‌گیرد. کسانی که تکیه صرف بر متریک‌های مبتنی بر ارجاعات را ترویج می‌کنند، تلویحاً فرض می‌کنند که هر ارجاع معنی یکسانی در مورد پژوهش مرجع می‌دهد (تأثیر آن). این فرضیه اثبات نشده و احتمالاً نادرست است.
- ثالثاً، با اینکه آمار برای درک دنیایی که در آن زندگی می‌کنیم ارزشمند است، تنها درکی جزئی از آن فراهم می‌کند. در دنیای امروز گاهی دفاع از این باور ابهام آمیز که اندازه‌گیری‌های عددی برتر از انواع دیگر ادراک است، رایج می‌باشد. کسانی که استفاده از آمار ارجاعات را به عنوان جایگزینی بر ادراک کامل‌تر ترویج می‌کنند، تلویحاً چنین باوری دارند. ما نه تنها باید آمار را به‌درستی مورد استفاده قرار دهیم بلکه نیازمندیم آن را هوشمندانه به‌کار بریم.

ما بحثی با تلاش برای ارزیابی پژوهش نداریم اما در عوض در ارتباط با این تقاضا که ارزیابی باید به طور عمده بر متریک‌های ساده و عینی مبتنی بر ارجاعات تکیه داشته باشد بحث می‌کنیم. این تقاضا اغلب به صورت نیاز برای اعدادی که به‌سادگی قابل محاسبه باشند، برای رتبه‌بندی نشریات، افراد یا برنامه‌ها تفسیر می‌شود. پژوهش معمولاً اهداف متعددی شامل اهداف کوتاه مدت و بلند مدت دارد. لذا منطقی است که ارزش آن باید توسط معیارهای متعددی داوری شود. ریاضیدانان می‌دانند که اشیاء زیادی، حقیقی یا موهومی وجود دارند که به راحتی نمی‌توان آن‌ها را به صورتی که هر دو شیء قابل مقایسه باشند مرتب کرد. مقایسه اغلب نیازمند تحلیل‌های پیچیده‌تر است و گاهی انسان را مردد باقی می‌گذارد که کدام شیء «بهتر» است. گاهی جواب صحیح به این سؤال این است که «بستگی دارد»! پیشنهادهای زیادی در ارتباط با استفاده از روش‌های متعدد برای ارزیابی پژوهش ارائه شده است (برای مثال [۲۵] یا [۹]). نشریات می‌توانند به طرق بسیاری و نه تنها توسط ارجاعات مورد داوری قرار بگیرند. معیارهای اعتبار از قبیل دعوت‌ها، عضویت در هیئت تحریریه‌ها و پاداش‌ها اغلب کیفیت را می‌سنجند. در برخی نظام‌ها و برخی کشورها، کمک هزینه‌ها نقش مهمی دارند. همچنین داوری توسط همکاران مؤلفه مهمی در ارزیابی

است (نباید داوری توسط همکاران را تنها به خاطر اینکه گاهی توسط اریبی خدشه‌دار می شود کنار بگذاریم. همانگونه که آمار ارجاعات را به خاطر استفاده‌های نادرست دور نمی‌ریزیم). این یک نمونه از طرق متعددی است که ارزیابی می‌تواند انجام شود. راه‌های بسیاری برای ارزیابی خوب وجود دارد، و اهمیت نسبی آنها در نظام‌های مختلف متفاوت است. علیرغم این موضوع آماره‌های «عینی» مبتنی بر ارجاعات مکررا روش ارجح برای ارزیابی‌اند. به نظر می‌رسد که فریبندگی یک فرایند ساده و اعداد ساده (ترجیحاً یک عدد) بر مفهوم متعارف و قضاوت درست غلبه می‌کند.

این گزارش توسط ریاضیدانان برای تفهیم استفاده‌های نادرست از آمار در ارزیابی پژوهش علمی نوشته شده است. البته این استفاده‌های غلط گاهی به سوی نظام ریاضیات معطوف‌اند و این خود یکی از دلایل اصلی نگارش این گزارش است. فرهنگ خاص ارجاع در ریاضیات، با میزان کم ارجاع به مجلات، مقالات و نویسندگان، آن را به صورتی ویژه در برابر استفاده‌های نادرست از آمار ارجاعات آسیب پذیر می‌سازد. هر چند ما معتقدیم که کلیه دانشمندان و حتی عموم مردم باید نگران استفاده از روش‌های علمی معتبر برای ارزیابی پژوهش باشند.

برخی در جامعه علمی طی یک عکس العمل عیب‌جویانه به سوءاستفاده‌های گذشته، آمار ارجاعات را به کلی رد می‌کنند. اما این کار به معنای دورانداختن یک ابزار با ارزش است. آماره‌های مبتنی بر ارجاعات به شرطی که به شکل مناسب مورد استفاده قرار بگیرند، با احتیاط تفسیر شوند و تنها بخشی از فرایند ارزیابی را تشکیل دهند، می‌توانند در ارزیابی پژوهش نقش داشته باشند. ارجاعات در مورد مجلات، مقالات و افراد اطلاعاتی فراهم می‌کنند. ما نمی‌خواهیم این اطلاعات را پنهان کنیم بلکه می‌خواهیم آن را واضح‌تر سازیم.

هدف این گزارش همین است. در سه بخش اول راه‌های استفاده (و سوء استفاده) از داده‌های ارجاع برای ارزیابی مجلات، مقالات و افراد توضیح داده شده است. در بخش بعد معانی متفاوت ارجاع و محدودیت‌های حاصل روی آماره‌های مبتنی بر ارجاعات مورد بحث قرار داده شده است. در بخش انتهایی در رابطه با استفاده هوشمندانه از آمار توصیه‌هایی ارائه گردیده و بر استفاده از آمار ارجاعات در ترکیب با دیگر روش‌های داوری حتی اگر از سادگی ارزیابی می‌کاهد، تأکید شده است.

آلبرت اینشتین می‌گوید: همه چیز باید تا حد امکان ساده شود؛ اما نه ساده‌تر از آن. این توصیه، از یکی از دانشمندان سرآمد جهان، به‌ویژه به هنگام ارزیابی پژوهش علمی، شایان توجه است.

۲ رتبه‌بندی مجلات: ضریب تأثیر

ضریب تأثیر در دهه ۱۹۶۰ به عنوان روشی برای سنجش ارزش مجلات توسط محاسبه میانگین تعداد ارجاعات هر مقاله طی یک دوره زمانی مشخص به وجود آمد [۱۸]. این میانگین از داده‌هایی که توسط *Thompson Scientific* (قبلاً با نام مؤسسه اطلاعات علمی [۷] خوانده می‌شد.) که گزارش ارجاعات مجلات را

[۷] The Institute for Scientific Information

منتشر می‌کند محاسبه می‌شود.

$Thompson\ Scietific$ هر ساله مراجع را از بیش از ۹۰۰۰ مجله استخراج می‌کند و اطلاعات در رابطه با هر مقاله و مراجع آن را به پایگاه داده خود اضافه می‌کند[$THOMPSON : SELECTION$]. با استفاده از آن اطلاعات می‌توان تعداد دفعاتی را که به یک مقاله به‌خصوص توسط مقالات بعدی که در مجموعه مجلات فهرست شده منتشر شده‌اند، ارجاع داده می‌شود، شمارش کرد. (خاطر نشان می‌کنیم که $Thompson\ Scietific$ کمتر از نیمی از مجلات ریاضی پوشش داده شده توسط مرورهای ریاضی [۸] و زنترابلات [۹]، دو مجله بازبینی عمده در ریاضیات را فهرست می‌کند.)

برای یک مجله و یک سال به‌خصوص، ضریب تأثیر مجله توسط محاسبه‌ی میانگین تعداد ارجاعات به مقالات داخل مجله طی دو سال قبل از چاپ کلیه مقالات در سال مربوطه بدست می‌آید. (در مجموعه به خصوصی از مجلات که توسط

$Thompson\ Scietific$ فهرست می‌شوند. اگر ضریب تأثیر مجله درسال ۲۰۰۷، ۱۰۵ باشد به این معنی است که به طور متوسط به مقالاتی که طی ۲۰۰۵ و ۲۰۰۶ منتشر شده‌اند، ۱۰۵ بار توسط مقالاتی که در مجموعه مجلات فهرست شده منتشر شده در ۲۰۰۷، ارجاع داده شده‌اند.

$Thompson\ Scietific$ خود ضریب تأثیر را به عنوان عاملی در انتخاب مجلاتی که فهرست می‌کنند مورد استفاده قرار می‌دهد[$THOMPSON : SELECTION$]. از طرف دیگر، $Thompson$ استفاده‌ی کلی‌تر از ضریب تأثیر را برای مقایسه مجلات ترویج می‌کند.

«ضریب تأثیر به‌عنوان ابزاری برای مدیریت مجموعه مجلات کتابخانه، اطلاعاتی در مورد مجلاتی که هم اکنون در مجموعه هستند و مجلاتی که برای استفاده تحت بررسی هستند برای مسئول کتابخانه فراهم می‌کند. این داده‌ها باید با داده‌های هزینه و تیراژ ترکیب شوند تا تصمیمات معقولی در مورد خرید مجلات اتخاذ شود» [$THAMPSON : IMPACT\ FACTOR$].

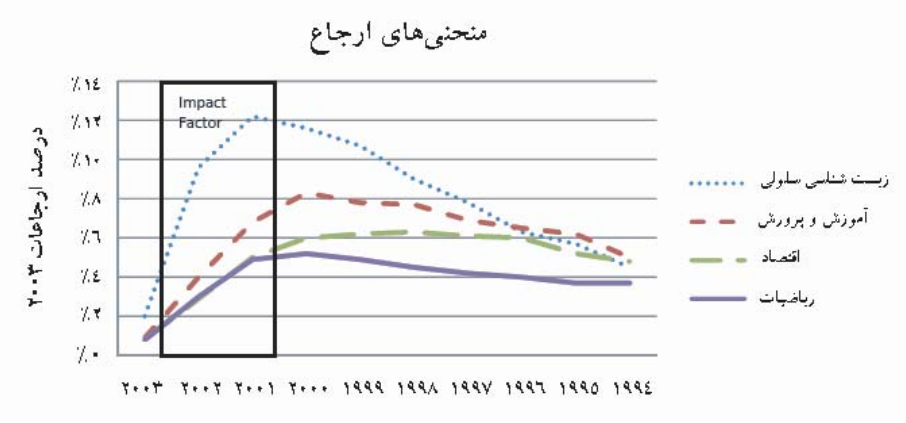
بسیاری از نویسندگان اشاره کرده‌اند که ارزش آکادمیک یک مجله نباید با استفاده از داده‌های ارجاعات به تنهایی مورد قضاوت قرار بگیرد و نویسندگان حاضر با این عیقده موافقند. علاوه بر این دیدگاه کلی، ضریب تأثیر به‌خاطر دلایل دیگری نیز مورد انتقاد قرار گرفته است ([۳۰]، [۲]، [۲۸]، [۱۳]، [۱] و [۲۰] رابببینید).

۱— تعیین ضریب تأثیر به عنوان میانگین، درست نیست. زیرا بسیاری از مجلات قطعات خبری فرعی مانند نامه‌هایامربوط به هیئت تحریریه [۱۰] چاپ می‌کنند که به ندرت به آنها ارجاع داده می‌شود. این اقلام در مخرج ضریب تأثیر به حساب آورده نمی‌شوند. از طرف دیگر هرچند نادر اما این اقلام گاهی مورد ارجاع واقع می‌شوند و این ارجاعات در صورت ضریب تأثیر به حساب آورده می‌شوند. لذا ضریب تأثیر کاملاً میانگین ارجاعات هر مقاله نیست. زمانی که مجلات تعداد زیادی از این قطعات خبری فرعی به چاپ می‌رسانند، این انحراف

[۸] Mathematical Reviews
[۹] Zentralblatt
[۱۰] Editorial

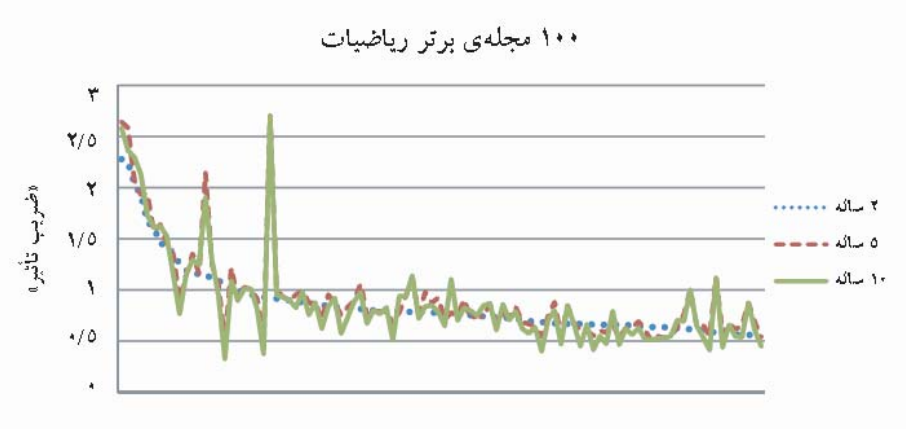
می‌تواند معنی‌دار شود. در بسیاری زمینه‌ها، شامل ریاضیات، این انحراف کمینه است.

۲– دوره دو ساله مورد استفاده در تعریف ضریب تأثیر به منظور به روز ساختن آماره بوده است [۱۸]. برای برخی رشته‌ها مانند علوم زیست پزشکی[۱۱] این تعریف مناسب است زیرا اغلب مقالات منتشر شده بیشتر ارجاعات خود را در زمان کوتاهی پس از انتشار دریافت می‌کنند. در دیگر رشته‌ها، مانند ریاضیات، بیشتر ارجاعات ورای این دوره دو ساله اتفاق می‌افتد.



با آزمون کردن مجموعه‌ای از بیش از سه میلیون ارجاع در مجلات ریاضی (پایگاه داده ارجاعات نشریات ریاضی می‌توان دید که تقریباً ۹۰درصد از ارجاعات به مجله خارج از این پنجره‌ی دو ساله قرار می‌گیرند. در نتیجه، ضریب تأثیر براساس تنها ۱۰ درصد از فرایند ارجاع به دست آمده و بخش بزرگی از ارجاعات را از دست می‌دهد. آیا این فاصله دو ساله به معنای گمراه کننده بودن ضریب تأثیر است؟ برای مجلات ریاضی گواه روشنی وجود ندارد.

Thompson Scietific ضریب تأثیرهای پنج ساله را محاسبه می‌کند و نشان می‌دهد که به خوبی با ضریب تأثیرهای دو ساله همبسته‌اند [۱۷].

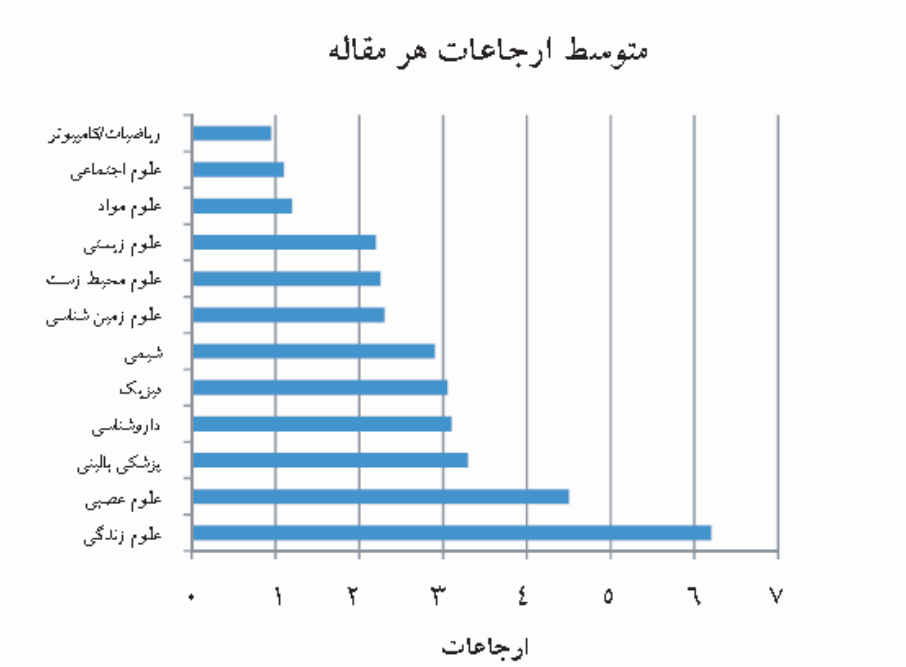


[۱۱] Biomedical

با استفاده از پایگاه داده ارجاعات نشریات ریاضی می‌توان ضریب تأثیرها (یعنی میانگین ارجاعات هر مقاله) را برای مجموعه‌ای از صد مجله ریاضی که بیشترین ارجاعات را داشته‌اند با استفاده از دوره‌های دو، پنج و ده ساله محاسبه نمود. نمودار صفحه‌ی قبل نشان می‌دهد که ضریب تأثیرهای پنج و ده ساله در حالت کلی از ضریب تأثیر دو ساله تبعیت می‌کنند. یک داده پرت بزرگ مجله‌ای است که در بخشی از این دوره زمانی مقاله‌ای منتشر نکرده است؛ داده‌های پرت کوچک‌تر مجلاتی هستند که نسبتاً تعداد کمتری از مقالات را در هر سال منتشر کرده‌اند. نمودار برای ضریب تأثیر چنین مجلاتی تنها یک تغییرپذیری نرمال را نشان می‌دهد. واضح است که تغییر در تعداد «سال‌های مورد نظر» برای محاسبه‌ی ضریب تأثیر، در رتبه‌بندی مجلات تأثیرگذار است و موجب ایجاد تغییراتی در آن می‌شود، اما این تغییرات به جز برای مجلات کوچک که ضریب تأثیر آنها نیز با تغییر «سال مرجع» تغییر می‌کند، نسبتاً کم است.

۳– ضریب تأثیر به شکل قابل توجهی بین نظام‌ها متغیر است [۲]. بخشی از این اختلاف‌ها ریشه در نظریه ۲ دارد: اگر در برخی نظام‌ها تعداد زیادی از ارجاعات خارج از پنجره دو ساله رخ دهند، ضریب تأثیر مجلات خیلی کم خواهد بود. از طرف دیگر بخشی از اختلاف به‌سادگی به این صورت توضیح داده می‌شود که فرهنگ ارجاع از نظامی به نظام دیگر متفاوت است و دانشمندان با نرخ‌های متفاوت و به دلایل متفاوت به مقالات ارجاع می‌دهند (این دیدگاه بعداً با جزئیات بیشتر شرح داده می‌شود زیرا مفهوم ارجاعات از اهمیت بالایی برخوردار است). نتیجه اینکه به هیچ طریقه‌ی معنی‌داری نمی‌توان دو مجله را در دو نظام متفاوت با استفاده از ضریب تأثیر مقایسه کرد.

۴– ضریب تأثیر می‌تواند از سالی به سال دیگر به شکل قابل توجهی تغییر کند و این تغییرات برای مجلات کوچکتر بیشتر است [۲]. به عنوان مثال برای مجلاتی که کمتر از ۵۰ مقاله چاپ می‌کنند، میانگین تغییرات ضریب تأثیر از ۲۰۰۲ تا ۲۰۰۳ تقریباً ۵۰ درصد بوده است. البته این امر دور از انتظار نیست زیرا اندازه نمونه برای مجلات کوچک، کم است. از طرف دیگر مجلات اغلب برای یک سال ثابت، بدون در نظر گرفتن تغییرات بیشتر برای مجلات کوچک، مقایسه می‌شوند.



۵– مجلاتی که به زبان‌های غیر از زبان انگلیسی منتشر می‌شوند، امکان دریافت ارجاع کمتری دارند، زیرا بخش بزرگی از جامعه علمی قادر به خواندن آنها نیستند (یا نمی‌خوانند). همچنین به جز کیفیت مجله، نوع آن نیز ممکن است روی ضریب تأثیر اثر داشته باشد. برای مثال مجلاتی که مقالات تجدید نظر چاپ می‌کنند، اغلب ارجاعات بیشتری نسبت به دیگر مجلات دریافت می‌کنند که باعث می‌شود ضرایب تأثیر بالاتر گاهی خیلی بالاتر داشته باشند[۲].

٦– مهم‌ترین انتقادی که به ضریب تأثیر وارد است این است که مفهوم آن به خوبی درک نشده است. زمانی که برای مقایسه‌ی دو مجله از ضریب تأثیر استفاده می‌شود مدلی برای بهتر بودن وجود ندارد که تعریف کند بهتر بودن به چه معنی است. تنها مدل موجود، مدلی است که از خود ضریب تأثیر مشتق می‌شود؛ ضریب تأثیر بزرگ‌تر به معنای مجله بهتر است. در الگوی کلاسیک آماری یک مدل تعریف می‌شود، فرضیات (بدون هیچ تفاوتی) فرموله می‌شوند، و سپس یک آماره ساخته می‌شود که بسته به مقدار آن فرض مورد نظر رد یا پذیرفته می‌شود. استخراج اطلاعات (یا در صورت امکان) مدل از خود داده‌ها یک رهیافت صحیح در تحلیل آماری است، اما در این حالت اطلاعات مشتق شده مشخص نمی‌باشد. چگونه ضریب تأثیر کیفیت را می‌سنجد؟ آیا ضریب تأثیر بهترین آماره برای سنجش کیفیت است؟ دقیقاً چه چیز را اندازه‌گیری می‌کند؟ (بحث بعدی ما در مورد مفهوم ارجاع در ارتباط با همین موضوع است). اطلاع بسیار ناچیزی در رابطه با مدلی برای کیفیت مجله و یا چگونگی ارتباط آن با ضریب تأثیر موجود است.

شش انتقاد وارد به ضریب تأثیر همگی معتبرند، اما این انتقادها تنها به این مفهوم‌اند که ضریب تأثیر کامل نیست، نه اینکه غیر قابل استفاده است. به عنوان

مثال ضریب تأثیر می‌تواند به عنوان یک نقطه شروع در رتبه‌بندی مجلات در گروه‌ها به کار گرفته شود؛ یعنی ابتدا از ضریب تأثیر برای تعریف گروه‌ها استفاده شود و سپس معیار دیگری برای تصحیح رتبه‌بندی و معنی‌دار بودن گروه‌ها به کار گرفته شود. اما استفاده از ضریب تأثیر برای ارزیابی مجلات نیازمند احتیاط است. برای مثال ضریب تأثیر نمی‌تواند برای مقایسه‌ی مجلات بین نظام‌ها مورد استفاده قرار بگیرد و نوع مجلات هنگام استفاده از ضریب تأثیر برای رتبه‌بندی آنها باید به دقت بررسی گردد. همچنین باید توجه زیادی به تغییرات سالیانه، به خصوص برای مجلات کوچک‌تر مبذول داشت و مستحضر بود که اختلاف‌های کوچک ممکن است پدیده‌هایی کاملاً تصادفی باشند. همچنین در نظر داشتن این امر که ضریب تأثیر ممکن است بازتاب دقیقی از دامنه کامل فرایند ارجاع نباشد (هم به خاطر فهرست نشدن مجلات و هم به خاطر کوتاه بودن دوره زمانی) از اهمیت بالایی برخوردار است. آماره‌های دیگر بر اساس دوره زمانی طولانی‌تر و تعداد مجلات بیشتر می‌توانند شاخص‌های بهتری از کیفیت باشند. نهایتاً می‌توان گفت که ارجاعات تنها یک راه برای داوری مجلات‌اند و باید توسط اطلاعات دیگر تکمیل گردند (پیام اصلی این گزارش).

مشابه این احتیاط‌ها در هر نوع رتبه‌بندی بر اساس آماره‌ها مورد نیاز است. رتبه‌بندی بیفکرانه مجلات مطابق ضریب تأثیر در یک سال به خصوص، استفاده نادرست‌ی از آمار است. $Thompson\ Scietific$ به خاطر اعتبار خود این اظهارات را تصدیق می‌کند و به کسانی که ضریب تأثیر را برای این گونه امور مورد استفاده قرار می‌دهند اخطار می‌دهد:

$Thompson\ Scietific$ برای ارزیابی میزان سودمندی مجلات بر روی ضریب تأثیر به تنهایی تکیه نمی‌کند و دیگران نیز نباید چنین کنند. ضریب تأثیر نباید بدون توجه به پدیده‌های بسیاری که نرخ ارجاع را تحت تأثیر قرار می‌دهند از قبیل میانگین تعداد مراجع در متوسط مقالات، مورد استفاده قرار بگیرد. ضریب تأثیر باید به همراه داوری آگاهانه توسط همکاران به‌کار گرفته شود.» $[THOMPSON: IMPACT\ FACTOR]$

اما متأسفانه این توصیه اغلب نادیده گرفته می‌شود.

۳ رتبه‌بندی مقالات

ضریب تأثیر و آماره‌های مشابه مبتنی بر ارجاعات می‌توانند برای رتبه‌بندی مقالات مورد استفاده‌ی نادرست قرار بگیرند. اما سوءاستفاده‌های اساسی‌تر و پرضررتری مانند استفاده از ضریب تأثیر برای مقایسه تک‌تک مقالات، افراد، برنامه‌ها و حتی نظام‌ها وجود دارد. این مشکل رو به گسترش است که در میان بسیاری ملت‌ها و نظام‌ها رواج یافته و با ارزیابی‌های ملی پژوهش اخیر وخیم‌تر نیز شده است. این پدیده، به نوعی، جدید نیست. دانشمندان گاهی برای انجام داوری‌هایی در ارتباط با سابقه نشریات خوانده می‌شوند و نظراتی از قبیل «او مجله خوبی منتشر می‌کند» یا «اکثر مقالات او در مجلات سطح پایین است» شنیده می‌شود. اینها می‌توانند ارزیابی‌های معقولی باشند: کیفیت مجلاتی که یک دانشمند به طور کلی مقالات خود را در آنها

به چاپ می‌رساند یکی از عوامل گوناگونی است که می‌توانند در ارزیابی همه جانبه‌ی پژوهش وی مورد استفاده قرار بگیرند. هر چند ضریب تأثیر تمایل به نسبت دادن ویژگی‌های یک مجله به هر مقاله داخل آن (و هر نویسنده) را افزایش داده است.

Thompson Scietific تلویحاً این عمل را ترویج می‌کند:

«شاید مهم‌ترین و جدیدترین استفاده از ضریب تأثیر در فرایند ارزیابی آکادمیک باشد. ضریب تأثیر می‌تواند برای به دست آوردن یک تخمین ناخالص از اعتبار مجلاتی که افراد مقالات خود را در آنها به چاپ می‌رسانند مورد استفاده قرار بگیرد».

$[THOMPSON : IMPACT\ FACTOR]$

در اینجا چند مثال از از تفاسیری که از توصیه بالا توسط ریاضیدانان سراسر جهان گزارش شده، ارائه شده است:

مثال ۱: دانشگاه من اخیراً یک طبقه‌بندی جدید از مجلات با استفاده از مجلات

Science Citation Index Core معرفی کرده است. مجلات تنها بر اساس ضریب تأثیر به سه گروه تقسیم شده اند. ۳۰ مجله در فهرست برتر موجود است که هیچ مجله ریاضی را شامل نمی‌شود. فهرست دوم شامل ٦٦٧ مجله است که ۲۱ مجله از این فهرست مجله ریاضی است. انتشار مقاله در مجلات فهرست اول حمایت دانشگاه از پژوهش را سه برابر و در مجلات فهرست دوم دو برابر می‌کند. انتشار مقاله در مجلات *Science Citation Index Core* ۱۵ نمره و در هر مجله‌ی پوشش داده شده توسط *Thompson Scientific*، ۱۰ نمره پاداش دارد. ترفیع نیازمند حداقل ثابتی از نمرات است.

مثال ۲: در کشور من هیئت علمی دانشگاه با پست ثابت هر ٦ سال یک بار ارزیابی می‌شوند. ارزیابی‌های موفقیت آمیز مداوم کلید موفقیت آکادمیک است. علاوه بر CV مهم ترین عامل ارزیابی مربوط به رتبه‌بندی پنج مقاله منتشر شده است. در سالهای اخیر در صورتی که مقالات در یک سوم برتر مجلات فهرست *Thompson Scientific* ظاهر شوند ۳ نمره، در دو سوم برتر مجلات ۲ نمره و در یک سوم آخر ۱ نمره به آنها تعلق می‌گیرد این سه فهرست با استفاده از ضریب تأثیر به وجود آمده‌اند.

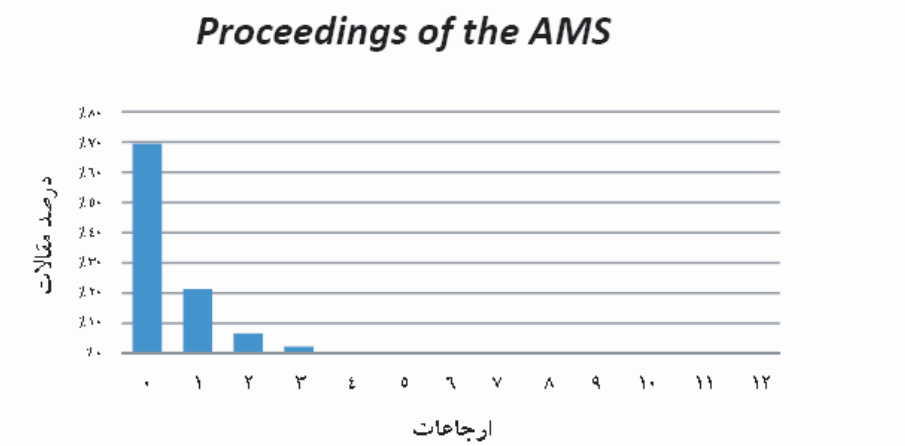
مثال ۳: در دپارتمان ما هر عضو هیئت علمی توسط فرمولی شامل تعداد مقالات معادل یک نویسنده[۱۲] ضرب در ضریب تأثیر مجلاتی که در آنها چاپ شده‌اند، ارزیابی می‌شوند. ترفیع‌ها و استخدام‌ها تا اندازه‌ای بر اساس این فرمول صورت می‌گیرند.

در این مثال‌ها و همچنین بسیاری گزارش‌های دیگر، به صورت مستقیم و غیر مستقیم ضریب تأثیر برای مقایسه مقالات به همراه نویسندگان آنها استفاده می‌شود: ضریب تثیر مجله A بزرگتر از مجله B است، پس مطمئناً مقاله داخل A برتر از مقاله داخل B و نویسنده A برتر از نویسنده B است. در برخی موارد این استدلال به رتبه‌بندی دپارتمان‌ها و حتی نظام‌ها تعمیم می‌یابد. از گذشته می‌دانیم که توزیع تعداد ارجاعات برای مقالات داخل یک مجله به شدت چوله است و تخمینی از قانون

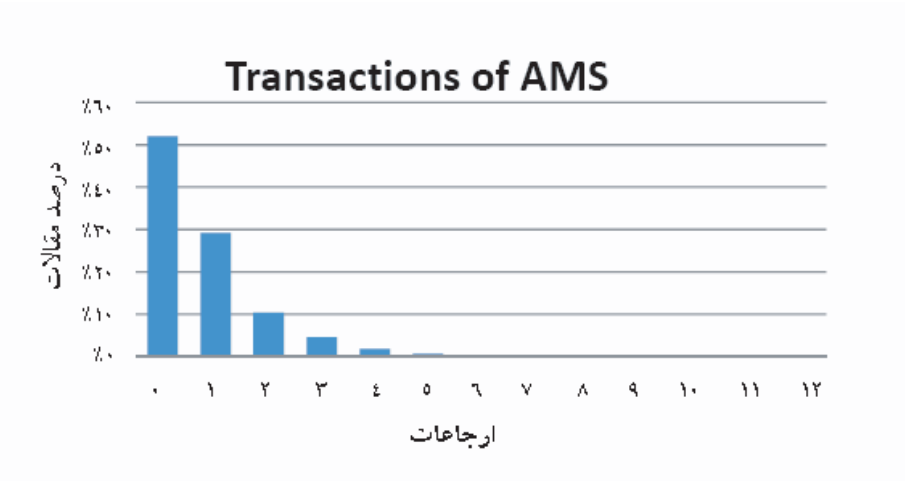
[۱۲] single-author-equivalent

معروف توان[۱۳] است ([۳۰]، [۱۶]). نتایج این امر را با مثالی روشن می‌کنیم.

توزیع مقالات در «مجموعه مقالات انجمن ریاضی امریکا»[۱۴] طی دوره ۲۰۰۴–۲۰۰۲ در زیر دیده می‌شود.



«مجموعه مقالات»[۱۵] کوتاه، معمولاً کوتاه‌تر از ده صفحه از نظر طول را به چاپ می‌رساند و در طول این دوره ۲۳۸۱ مقاله در حدود ۱۵۰۰۰ صفحه، منتشر کرده است. با استفاده از این مجله در اطلاعات ارجاع[۱۶]، در سال ۲۰۰۵ میانگین تعداد ارجاعات هر مقاله یعنی ضریب تأثیر، ۰/۴۳۴ است. مجموعه نسخه های انجمن ریاضی امریکا[۱۷] مقالات طولانی‌تری را چاپ می‌کند که معمولاً از نظر موضوع و محتوا نیز مهم‌ترند.



[۱۳] power law
[۱۴] Proceedings of the American Mathematical Society
[۱۵] Proceedings
[۱۶] Math Reviews
[۱۷] Transactions of the American Mathematical Society

طی همان دوره زمانی مجموعه نسخه ها[۱۸] ۱۱۶۵ مقاله با بیش از ۲۵۰۰۰ صفحه، با تعداد ارجاعات در دامنه ۰ تا ۱۲ منتشر ساختند. میانگین تعداد ارجاعات هر مقاله ۰/۸۴۶، در حدود دو برابر ضریب تأثیر در مجموعه مقالات بود. حال دو ریاضیدان را در نظر بگیرید که یکی از آنها مقاله‌ای در «مجموعه مقالات» و دیگری در «مجموعه نسخه‌ها» به چاپ می‌رساند. با استفاده از اقدامات مؤسسه‌ای مذکور ریاضیدان دوم با چاپ مقاله‌ای در مجله‌ای با با ضریب تأثیر بزرگتر در واقع دو برابر برتر از ریاضیدان اول مورد قضاوت قرار می‌گیرد. آیا این ارزیابی معتبر است؟ آیا «مقالات نسخه‌ها» دو برابر مقالات در «مجموعه مقالات» برتری دارند؟

زمانی که ادعا می‌کنیم یک مقاله در «مقالات نسخه‌ها» بهتر (از نظر ارجاعات) از یک مقاله در «مجموعه مقالات» است، لازم نیست در مورد میانگین‌ها سؤال کنیم، بلکه باید راجع به احتمال‌ها بپرسیم: احتمال اینکه اشتباه کنیم چقدر است؟ چقدر احتمال دارد که یک مقاله انتخاب شده به صورت تصادفی از «مجموعه مقالات» دارای تعداد ارجاعات حداقل به اندازه ارجاعات یک مقاله‌ی انتخاب شده به صورت تصادفی از «مقالات نسخه‌ها» باشد؟

با محاسبات مقدماتی پاسخ برابر ۶۲ درصد است. یعنی ۶۲درصد اوقات اشتباه می‌کنیم و علی‌رغم این واقعیت که ضریب تأثیر «مجموعه مقالات» نصف ضریب تأثیر «مقالات نسخه‌ها» است، یک مقاله به تصادف انتخاب شده از «مجموعه مقالات» به خوبی (یا بهتر) از یک مقاله‌ی به تصادف انتخاب شده از «مقالات نسخه‌ها» است! لذا اغلب اوقات اشتباه قضاوت می‌کنیم. اکثر افراد این امر را تعجب‌آور می‌یابند، اما این امر نتیجه توزیع شدیداً چوله و پنجره محدود زمانی است که برای محاسبه ضریب تأثیر مورد استفاده قرار می‌گیرد (و دلیل درصد بالای مقالات با ارجاع صفر است)، می‌باشد. این امر ارزش تفکر دقیق آماری را به جای مشاهدات حسی نشان می‌دهد.

این رفتار نوعیِ مجلات است و دلیل خاصی برای انتخاب این دو مجله وجود ندارد (برای مثال *Journal of American Mathematical Society* در طول همان دوره‌ی زمانی دارای ضریب تأثیر ۲/۶۳ یعنی ۶ برابر ضریب تأثیر «مجموعه مقالات» است. با این حال یک مقاله‌ی به تصادف انتخاب شده از «مجموعه مقالات» در ۳۲درصد اوقات از نظر ارجاع حداقل به خوبی مقاله‌ای از *Journal of American Mathematical Society* است).

لذا در حالی‌که درست نیست بگوییم ضریب تأثیر اطلاعی درباره‌ی تک‌تک مقالات داخل یک مجله نمی‌دهد، می توان گفت این اطلاعات به شکل حیرت‌آوری مبهم‌اند و می‌توانند به شدت گمراه کننده باشند.

نتیجه می‌گیریم که نوع محاسبات انجام شده در سه مثال بالا —استفاده از ضریب تأثیر به عنوان شاخصی از تعداد واقعی ارجاعات هر مقاله— دارای اساس عقلایی ناچیزی است. ایراد اظهاراتی که بیش از نیمی از اوقات یا یک

[۱۸] Transactions

سوم اوقات نادرست می‌باشند، مطمئناً راه خوبی برای انجام ارزیابی نیست.

زمانی که در می‌یابیم که قرار دادن ضریب تأثیر برای تعداد ارجاعات مقالات بی‌معنی است، نتیجه می‌شود که استفاده از آن برای ارزیابی نویسندگان این مقالات، برنامه‌هایی که نویسندگان تحت آن عمل می‌کنند و مطمئناً نظام‌هایی که تابع آنها هستند، نیز مفهومی ندارد. ضریب تأثیر و به طور کلی میانگین‌ها خام‌تر از آن هستند که بتوانند مقایسه‌های معقولی از این دست و بدون هیچ اطلاع دیگری انجام دهند.

البته رتبه‌بندی افراد، مشابه رتبه‌بندی مقالاتشان نیست. اما اگر بخواهیم مقالات یک فرد را تنها با استفاده از ارجاعات برای سنجش کیفیت هر مقاله، رتبه‌بندی کنیم، باید از شمارش ارجاعات هر مقاله شروع کنیم. ضریب تأثیر مجله‌ای که مقاله در آن ظاهر شده است، جایگزین قابل اطمینانی نمی‌باشد.

۴ رتبه‌بندی دانشمندان

در حالی‌که ضریب تأثیر بهترین آماره‌ی مبتنی بر ارجاعات شناخته شده است، آماره‌های جدیدتری وجود دارند که به شدت رواج پیدا کرده‌اند. در اینجا نمونه‌ی کوچکی شامل سه مورد از این آماره‌ها برای رتبه‌بندی افراد ارائه می‌کنیم. شاخص h: شاخص h یک دانشمند بزرگ‌ترین n ی است که به ازای آن فرد مورد نظر دارای n مقاله منتشر شده هر یک با حداقل n ارجاع باشد.

این آماره در میان آماره‌هایی که به آنها اشاره شد، محبوب‌ترین است. شاخص h توسط هیرش [۲۱] به‌منظور سنجش بازده علمی یک محقق با تمرکز روی دم بلند انتهایی توزیع ارجاعات فرد، پیشنهاد شده است. هدف جانشین ساختن یک عدد برای تعداد منتشرات و توزیع ارجاعات بوده است.

شاخص m: شاخص m یک دانشمند عبارت است از شاخص h او تقسیم بر تعداد سال‌هایی که از انتشار اولین مقاله‌ی وی گذشته است. این شاخص نیز توسط هیرش در همان مقاله‌ی ذکر شده در بالا پیشنهاد شده است. هدف، جبران کردن برای دانشمندان جوان‌تر که زمان کافی برای انتشار مقاله و به دست آوردن ارجاعات را نداشته‌اند، بوده است.

شاخص g: شاخص g یک دانشمند بزرگ‌ترین n ی است که به ازای آن n مقاله که بیشترین ارجاع را داشته‌اند، در مجموع دارای ارجاع باشند.

این شاخص توسط لئو ایگ[۱۱] پیشنهاد شده است. شاخص h این واقعیت را که برخی از مقالات در n مقاله اول ممکن است به‌صورت غیر معمول تعداد ارجاعات بالایی داشته باشند، در نظر نمی‌گیرد. شاخص g برای جبران این موضوع است.

شاخص‌های بسیار دیگری شامل انواع گوناگونی از شاخص‌هایی که عمر مقاله و تعداد نویسندگان را به‌حساب می‌آورند، وجود دارند([۳]، [۴]،[۳۱]).

هیرش در مقاله‌ی خود در تعریف شاخص h می‌نویسد که او این شاخص را به عنوان یک شاخص که به سادگی قابل محاسبه است و برآوردی از اهمیت، معنی‌داری و تأثیر کلی همکاری پژوهشی انباشته‌ی یک دانشمند ارائه می‌کند، پیشنهاد کرده است. او اضافه می‌کند که «زمانی

که معیار ارزیابی، موفقیت علمی است، این شاخص می‌تواند یک معیار سودمند برای مقایسه‌ی افراد مختلفی که برای منابع یکسان رغابت می‌کنند باشد.»

هیچ یک از این ادعاها توسط گواه متقاعد کننده‌ای حمایت نمی‌شوند. هیرش برای حمایت ادعای خود که شاخص h اهمیت و مفهوم پژوهش انباشته یک دانشمند را می‌سنجد، شاخص h را برای مجموعه‌ای از برندگان جایزه نوبل و به‌صورت جداگانه برای اعضای آکادمی ملی تحلیل می‌کند. او نشان می‌دهد که افراد این گروه به طور کلی دارای شاخص‌های h بالایی هستند. می‌توان نتیجه گرفت که در صورتی که دانشمندی برنده‌ی جایزه نوبل باشد، احتمالاً دارای شاخص h بالا است. اما بدون اطلاعات بیشتر در مورد امکان اینکه شخصی با شاخص h بالا برنده جایزه نوبل یا عضو آکادمی ملی شود اطلاع ناچیزی داریم. و این همان نوع اطلاعاتی است که برای تصدیق اعتبار شاخص h مورد نیاز است.

هیرش در مقاله‌ی خود همچنین ادعا می‌کند که با استفاده از شاخص h می‌توان دو دانشمند را مورد مقایسه قرار داد:

«من استدلال می‌کنم که دو نفر با h مشابه روی هم رفته از نظر تأثیر علمی قابل مقایسه‌اند، حتی اگر تعداد کل مقالات یا تعداد کل ارجاعاتشان خیلی متفاوت باشد. بالعکس، بین دو نفر با تعداد مشابه مقالات و ارجاعات و مقادیر متفاوت h، شخصی که مقدار h بزرگتری داشته باشد، با احتمال بیشتری دانشمند فاضل‌تری است».

به‌نظر می‌رسد که این اظهارات توسط درک متعارف رد می‌شوند (دو دانشمند را در نظر بگیرید، هر یک با ۱۰ مقاله دارای ۱۰ ارجاع، اما یکی از آنها دارای ۹۰ مقاله‌ی اضافی با ۹ ارجاع باشد؛ یا فرض کنید یکی دارای دقیقا ۱۰ مقاله با ۱۰ ارجاع و دیگری دارای ۱۰ مقاله با ۱۰۰ ارجاع باشد. آیا می‌توان این دو را معادل فرض کرد؟) هیرش با این ادعا که«h بر بقیه‌ی معیارهای تک عددی که معمولاً برای ارزیابی خروجی علمی یک محقق مورد استفاده قرار می‌گیرند، ارجح است...» [۲۱] در مورد خواص آن اغراق می‌کند، اما او نه واژه‌ی ارجح را تعریف می‌کند و نه توضیح می‌دهد که چرا باید به دنبال معیار تک عددی بگردیم.

در حالی‌که انتقادهایی به این رهیافت وارد شده اما تحلیل جدی ناچیزی انجام گرفته است. اغلب تحلیل‌ها عبارت‌اند از نشان دادن اعتبار همگرا، یعنی این‌که شاخص h با دیگر متریک‌های انتشار/ ارجاع، مانند تعداد مقالات انتشاریافته یا تعداد کل ارجاعات، به‌خوبی پیوسته است. از آن‌جا که کلیه این متغیرها توابعی از پدیده‌ای واحد(انتشار) می‌باشند، این همبستگی اهمیت زیادی ندارد. در یک مقاله‌ی برجسته در مورد شاخص h [۲۳] نویسندگان تحلیل دقیق‌تری انجام داده‌اند و نشان می‌دهند که شاخص h (در واقع شاخص m) به«خوبیِ» در نظر گرفتن میانگین تعداد ارجاعات هر مقاله به تنهایی نیست. هر چند، حتی در این مقاله، نویسندگان تعریف کافی از واژه «خوب» ارائه نمی‌دهند. زمانی که الگوی کلاسیک آماری به‌کار گرفته می‌شود[۲۳]، اثبات می‌شود که شاخص h کم اعتبارتر از معیارهای دیگر است.

اَشکال متنوعی از شاخص h برای مقایسه کیفیت پژوهشگران نه تنها داخل یک نظام بلکه حتی بین نظام‌ها، توصیه شده است[۴]، [۲۷]. برخی ادعا می‌کنند که شاخص h می‌تواند برای مقایسه‌ی مؤسسات و

دپارتمان‌ها نیز مورد استفاده قرار بگیرد[۲۲]. این کوشش‌ها برای در برگیری یک سابقه‌ی پیچیده‌ی ارجاعات در یک عدد، اغلب اوقات کامل نیستند. در واقع مزیت اولیه‌ی این شاخص‌های جدید بر هیستوگرام‌های ساده‌ی تعداد ارجاعات این است که این شاخص‌ها تقریبا کلیه‌ی جزئیات سابقه‌ی ارجاعات را دور می‌ریزند و این امر امکان رتبه‌بندی برای هر دو دانشمند را فراهم می‌سازد. هرچند حتی مثال‌های ساده نشان می‌دهند که اطلاعات دور ریخته شده برای درک سابقه‌ی پژوهش، مورد نیازند. مطمئناً به هنگام ارزیابی پژوهش، هدف نهایی باید درک آن باشد، نه صرفاً اطمینان حاصل کردن از اینکه هر دو فردی قابل مقایسه‌اند.

در برخی موارد هیئت‌های ارزیابی ملی شاخص h یا یکی از اَشکال آن را به عنوان قسمتی از اطلاعات خود جمع آوری می‌کنند. متأسفانه یک عدد برای رتبه‌بندی هر دانشمند، نظریه‌ی گمراه کننده‌ای است که ممکن است چنان در میان عموم گسترش یابد که استفاده‌ی درست از استدلال آماری را برای شرایط ساده‌تر بد تعبیر کند.

۵ مفهوم ارجاع

افرادی که آمار ارجاعات را به عنوان معیار غالب برای کیفیت پژوهش ترویج می‌کنند پاسخی برای سؤال اساسی «مفهوم ارجاع چیست؟» ندارند. این افراد مقدار زیادی داده در مورد تعداد ارجاعات جمع آوری می‌کنند، برای استخراج آماره‌هااین داده‌ها را پردازش می‌کنند، و آنگاه اظهار می‌دارند که فرایندارزیابی حاصل «عینی» است.

در حالیکه این ارزیابی نتیجه تفسیر آماره‌هاست و تفسیر بر مفهوم ارجاع تکیه دارد که امری کاملاً شخصی است.

در ادبیات مروج این رهیافت، یافتن عبارات روشن در مورد مفهوم ارجاع به شکل حیرت‌آوری دشوار است.

«مفهوم نهفته در فهرست‌بندی ارجاعات اصولاً ساده است. با در نظر داشتن اینکه ارزش اطلاعات توسط افرادی که آن را مورد استفاده قرار می‌دهند تعیین می‌گردد، چه راهی بهتر از سنجش کیفیت کار با اندازه‌گیری اثری که به طور کلی روی جامعه می‌گذارد وجود دارد. وسیع‌ترین جامعه‌ی ممکن داخل جامعه‌ی دانش‌پژوهان یعنی هر کسی که منابع اصلی را مورد استفاده قرار داده یا تولید می‌کند، اثر این ایده و بنیانگذار آن را بر بدنه دانش ما تعیین می‌کند». $[THOMPSON : HISTORY]$

«با این‌که تعیین کیفیت شخص دانشمندان به‌صورت کمی کار دشواری است، نظر عموم این است که انتشار بیشتر بهتر از انتشار کمتر است، و لذا تعداد ارجاعات یک مقاله (متناسب با عادات ارجاع در یک رشته) معیار سودمندی از کیفیت است»[۲۳].

«فراوانی ارجاعات ارزش یک مجله و استفاده‌ی حاصل از آن را منعکس می‌کند...»[۱۵].

«زمانی که یک فیزیکدان یا یک محقق زیست پزشکی به مقاله‌ی مجله‌ای ارجاع می‌دهد، این امر نشان می‌دهد که مجله‌ی مرجع او را به نحوی تحت تأثیر قرار داده است»[۱۶].

«ارجاع نوعی قدردانی به خاطر دین علمی[۱۹] است». $[THOMPSON :$

[۱۹] intellectual debt

$FIFTY\ YEARS$]

واژه‌های مربوطه عبارتند از «کیفیت»[۲۰]، «ارزش»[۲۱]، «اثر»[۲۲] و «دِین علمی». واژه‌ی «تأثیر»[۲۳] به صورت یک کلمه‌ی عمومی برای انتساب مفهوم به ارجاعات در آمده است؛ واژه‌ای که اولین بار در یک مقاله‌ی کوتاه در[۱۴] برای ترویج ایده‌ی تولید فهرستی از ارجاعات، به کار برده شد. در این مقاله چنین نوشته شده است:

«به این ترتیب، در مورد یک مقاله‌ی مهم، یک مقدار کمی برای فهرست ارجاعات وجود دارد که می‌تواند نویسنده را در سنجش اثر مقاله —یعنی ضریب تأثیر آن— یاری دهد»[۱۴].

نسبتاً واضح است که در اینجا نیز مانند هر جای دیگر مقصود از به کارگیری عبارت ضریب تأثیر اظهار این امر است که مقاله‌ی ارجاع دهنده بر اساس کار مقاله‌ی مرجع بنا شده است. یعنی ارجاع مکانیزمی است که پژوهش از طریق آن گسترش می‌یابد.

ادبیاتی غنی در مورد مفهوم ارجاع اظهار می‌دارد که معنای ارجاع خیلی پیچیده‌تر از آن است که این اظهارات مبهم ما را وادار به باور آن می‌دارند. برای مثال، مارتین و ایروین [۲۵] در مقاله خود در مورد ارزیابی پژوهش در ۱۹۸۳ می‌نویسند:

«استفاده از ارجاعات به عنوان سنجشی از کیفیت در زمینه‌ی کلیه مشکلات، به نوعی نادیده گرفتن دلایلی است که نویسندگان به‌خاطر آنها بخش‌های به خصوصی از کار را مرجع قرار می‌دهند. تحلیل‌های ساده‌ی ارجاعات، مدل بسیار گویایی از ارائه‌ی مراجع را به عنوان پیش فرض در نظر می‌گیرند که در آن ارجاع عمدتاً سپاسگزاری علمی از کارهای با کیفیت یا اهمیت بالا در نظر گرفته می‌شود و ارجاع دهندگان دارای شانس مساوی برای ارجاع دادن به یک مقاله‌ی به خصوص می‌باشند...»[۲۵]. کزنس [۱۰] در مقاله ۱۹۸۸ خود در مورد مفهوم ارجاع اظهار می‌دارد که ارجاع حاصل دو سیستم در زمینه‌ی هدایت انتشارات علمی است، یکی سیستم «پاداش»[۲۴] و دیگری «لفظی»[۲۵]. نوع اول مفهومی است که اغلب اوقات به ارجاع منسوب می‌گردد؛ قدردانی به این دلیل که مقاله‌ی ارجاع دهنده به مقاله‌ی مرجع «دِین علمی» دارد. اما نوع دوم دارای مفهومی کاملاً متفاوت است؛ ارجاع به یک مقاله‌ی قبلی برای توضیح نتیجه‌ای، شاید خلاف نتیجه‌ی نویسنده‌ای که به او ارجاع داده شده است. چنین ارجاعاتی، تنها راهی برای ایراد یک مکالمه‌ی علمی‌اند و مصداق دِین علمی نمی‌باشند. البته در برخی موارد، ارجاع می‌تواند هر دو مفهوم را به همراه داشته باشد.

کزنس اظهار می‌کند که اکثر ارجاعات لفظی می‌باشند. این موضوع به تجربه توسط اغلب ریاضیدانان ورزیده تیید گردیده است (برای مثال در پایگاه داده‌ی ارجاعات $Math\ Reviews$ نزدیک ۳۰ درصداز بیش از ۳ میلیون ارجاع، ارجاع به کتاب است و نه به مقالات پژوهشی مجلات). چرا این مسئله اهمیت دارد؟ زیرا بر خلاف

[۲۰] quality
[۲۱] value
[۲۲] influence
[۲۳] impact
[۲۴] reward
[۲۵] rhetorical

ارجاع از نوع پاداش که تمایل به ارجاع به خود مقالات دارد، انتخاب اینکه به کدام مقاله ارجاع لفظی داده شود بستگی به عوامل زیادی از قبیل اعتبار نویسنده‌ی مرجع، رابطه نویسنده‌ی ارجاع دهنده و نویسنده‌ی مرجع، میزان دسترسی به مجله (آیا به مجلات با دسترسی آزاد با احتمال بیشتری ارجاع داده می‌شود؟)، تناسب ارجاع به چند نتیجه از یک مقاله و ... بستگی دارد. تعداد کمی از این عوامل مستقیماً با کیفیت مقاله مرجع مرتبطند.

حتی زمانی که ارجاعات از نوع «پاداش» هستند، می‌توانند بازتاب تنوعی از انگیزه‌ها شامل انتشار[۲۶]، اعتبار منفی[۲۷]، اطلاعات قابل استفاده[۲۸]، قوه اقناع[۲۹]، اعتبار مثبت[۳۰]، آگاهی خواننده[۳۱]، و توافق اجتماعی[۳۲] باشند [۸]. در اکثر موارد، ارجاعات توسط بیش از یکی از این عوامل انگیخته می‌شوند. برخی از نتایج قابل توجه می‌توانند گرفتار اثر «بطلان»[۳۳] شوند، به این ترتیب که به سرعت در کار دیگری که تبدیل به اساس ارجاعات بعدی می‌شود ترکیب می‌شوند. برخی ارجاعات دیگر، پاداش پژوهش برجسته نیستند بلکه اخطاری در رابطه با نتایج و پندارهای نادرست‌اند. گزارش حاضر مثال‌های متعددی از این ارجاعات «اخطار دهنده» ارائه می‌کند.

جامعه‌شناسی ارجاع مبحث پیچیده‌ای است و ورای اهداف این گزارش است. هر چند حتی این بحث مختصر نشان می‌دهد که مفهوم ارجاع به هیچ عنوان ساده نیست و آماره‌های مبتنی بر ارجاعات چنان که طرفداران آنها اظهار می‌دارند «عینی» نیستند.

برخی افراد ممکن است بحث کنند که مفهوم ارجاعات اهمیتی ندارد زیرا آماره‌های مبتنی بر ارجاعات به شدت با دیگر معیارهای کیفیت پژوهش (مانند داوری توسط همکاران) همبسته‌اند. به‌عنوان مثال،[۱۲] که قبلاً به آن اشاره شد، اشاره می‌کند که آماره‌های مبتنی بر ارجاعات به خاطر این همبستگی می‌توانند (و باید) جانشین اَشکال دیگر ارزیابی گردند:

«گواه موجود نشان می‌دهد که روش‌های بیبلومتریک می‌توانند شاخص‌هایی از کیفیت پژوهش بسازند که با ادراک محقق متجانس‌اند»[۱۲]. به‌نظر می‌رسد نتیجه این باشد که آماره‌های مبتنی بر ارجاعات، بدون توجه به مفهوم دقیقشان جانشین دیگر روش‌های ارزیابی می‌شوند. گذشته از چرخه‌وار بودن این بحث، سفسطه آمیز بودن این استدلال به‌راحتی مشهود است.

۶ استفاده‌ی هوشمندانه از آمار

تکیه‌ی بیش از حد بر متریک‌های عینی (آماره‌ها) برای ارزیابی پژوهش پدیده‌ای است که نه جدید است و نه منحصر به فرد. این موضوع به فصوح در کتاب معروف «آمار و دروغ‌ها»[۳۴] نوشته‌ی جامعه‌شناس جول بست[۳۵]

[۲۶] currency
[۲۷] negative credit
[۲۸] operational information
[۲۹] persuasiveness
[۳۰] positive credit
[۳۱] reader alert
[۳۲] social consensus
[۳۳] obliteration
[۳۴] Damned lies and statistics
[۳۵] Joel Best

در سال ۲۰۰۱ شرح داده شده است:

«برخی فرهنگ‌ها وجود دارند که مردم در آنها به قدرت جادویی برخی اجسام اعتقاد دارند؛ انسان‌شناسان این اجسام را طلسم می‌نامند. در جامعه‌ی ما، آماره‌ها نوعی طلسم‌اند. ما تمایل داریم آماره‌ها را طوری لحاظ کنیم گویی جادویی هستند و بیش از عدد محض می‌باشند. ما با آنها به صورت نمایش قدرتمندی از حقیقت برخورد می‌کنیم و چنان عمل می‌کنیم گویی آنها پیچیدگی و ابهام حقیقت را در واقعیت‌های ساده خلاصه می‌کنند. آماره‌ها را برای تبدیل مسائل پیچیده اجتماعی به برآوردهای قابل درک ، درصدها و نرخ‌ها مورد استفاده قرار می‌دهیم. آماره‌ها نگرانی‌های ما را هدایت می‌کنند، و نشان می‌دهند که در چه موردی و تا چه حد باید نگران باشیم. در مفهومی، مشکل اجتماعی به صورت یک آماره در می‌آید و چون آماره‌ها را درست و مسلم می‌پنداریم، آنها به نوعی کنترل جادویی و طلسم گونه بر چگونگی نگاه ما به مسائل اجتماعی را در دست می‌گیرند. ما آماره‌ها را در قالب واقعیت‌های قابل کشف می‌پنداریم، نه اعدادی که خود آنها را ساخته‌ایم»[٦].

این باور سحرآمیز در مورد جادوی آماره‌های ارجاع می‌تواند در مدارک مؤسسه‌ای و ملی روش‌های ارزیابی پژوهش یافت شود. حتی می‌توان آن را در کار افرادی که شاخص h و دیگر انواع آن را ترویج می‌کنند مشاهده کرد.

این گرایش همچنین در تلاش‌های اخیر برای بهبود ضریب تأثیر با استفاده از الگوریتم‌های پیچیده‌تر ریاضی، شامل الگوریتم‌های رتبه صفحه، برای تحلیل ارجاعات، مشهود است. ([۵]،[۳۲]) طرفداران آنها ادعاهایی درباره سودمندی آنها دارند که با تحلیل‌ها توجیه نمی‌شوند و ارزیابی آنها دشوار است. چون این رهیافت‌ها بر مبنای محاسبات پیچیده‌تری می‌باشند، اکثر افراد تشخیص فرضیات (معمولاً پنهان) در پس آنها را مشکل می‌یابند. اغلب مایلیم اعداد و رتبه‌بندی‌ها را با هیبت فرض کنیم و به صورت حقایق، نه آنچه خود خلق کرده‌ایم، به آنها می نگریم.

پژوهش اولین فعالیت سرمایه‌گذاری شده به شکل عمومی نیست که تحت بررسی درآمده است و طی دهه‌های گذشته افراد سعی بر اجرای ارزیابی‌های کمی کارایی هر چیزی، از آموزش (مدارس) گرفته تا بهداشت (بیمارستان‌ها و جراحان)، داشته‌اند. در برخی موارد، آماردانان برای ارائه‌ی توصیه‌هایی در مورد متریک‌های معقول و استفاده‌ی درست از آماره‌ها به افرادی که عمل ارزیابی را انجام می‌دهند، دخالت کرده‌اند. اگر شخصی برای استفاده از داروها با پزشکان مشورت می‌کند، مطمئناً باید به هنگام استفاده از آماره‌ها نیز با آماردانان مشورت کرده و توصیه‌های ایشان را لحاظ کند. دو مثال خوب برای این موضوع را می‌توان در ([۷] و [۱۹]) یافت. در حالیکه هر یک از این دو مثال با ارزیابی کارایی چیزی به جز پژوهش سر و کار دارند (اولی در ارتباط با کارایی بخش عمومی و دومی در رابطه با بهداشت/آموزش)، بینشی در ارتباط با استفاده‌ی معقول از آماره‌ها را در ارزیابی پژوهش فراهم می‌کنند. مقاله‌ی نوشته شده توسط گلداستین و اسپیگل‌هالتر به خصوص در ارتباط با استفاده از جداول اتحادیه‌ها (رتبه‌بندی‌ها) بر اساس اعداد ساده (برای مثال موفقیت دانش آموزان یا نتایج

پزشکی) می‌باشد، لذا به ارزیابی پژوهش توسط رتبه‌بندی مجلات، مقالات یا نویسندگان با استفاده از آماره‌های ارجاع مربوط است. این دو نویسنده در مقاله‌ی خود یک چارچوب کاری سه بخشی برای ارزیابی کارایی تهیه کرده‌اند:

۱.٦ داده‌ها

«هیچ اندازه‌ای از کار آماری نمی‌تواند بر نارسایی اولیه‌ی داده‌های جمع‌آوری شده چه در تناسب و چه در صحت، فائق آید»[۱۹].

این عبارت دیدگاه مهمی در ارزیابی کارایی آماره‌های مبتنی بر ارجاع است. برای مثال ضریب تأثیر بر اساس زیر مجموعه‌ای از داده‌ها محاسبه می‌شود که تنها شامل مجلاتی است که توسط $Thompson\ Scietific$ انتخاب شده‌اند. (اشاره می‌کنیم که ضریب تأثیر خود نقش مهمی در معیار انتخاب دارد). برخی افراد صحت این داده‌ها را زیر سؤال برده‌اند[۲۹]. برخی دیگر اشاره می‌کنند که ممکن است مجموعه داده‌های کامل‌تری وجود داشته باشد[۲٦]. بسیاری گروه‌ها ایده‌ی استفاده از Google Scholars را برای تکمیل آماره‌های مبتنی بر ارجاع مانند شاخص h تحمیل کرده‌اند، اما داده‌های موجود در $Google\ Scholars$ اغلب دقیق نیست زیرا اطلاعاتی مانند نام نویسندگان به صورت اتوماتیک از صفحات وب استخراج می‌شود. بدست آوردن آمار ارجاعات برای دانشمندان گاهی دشوار است، زیرا نویسندگان به‌صورت یکتا مشخص نمی‌شوند و در برخی سیستم‌ها و کشورهای بخصوص این امر می‌تواند مانعی بزرگ برای گردآوری داده‌های صحیح از ارجاعات باشد.

۲.٦ تحلیل آماری و ارائه

«باید توجه بخصوصی به تعیین یک مدل آماری مناسب، اهمیت قاطع عدم حتمیت در ارائه کلی نتایج، و تکنیک‌های تعدیل نتایج، عوامل ابهام آمیز و نهایتاً میزان تکیه بر رتبه‌بندی‌های مستقیم مبذول داریم»[۱۹].

همان‌طور که قبلاً اشاره کردیم در اکثر مواردی که آماره‌های ارجاع برای رتبه‌بندی مقالات، افراد و برنامه‌ها به کار می‌روند، هیچ مدل خاصی از پیش تعیین نمی‌شود. در عوض خود داده‌ها یک مدل پیشنهاد می‌کنند که اغلب مبهم است. به نظر می‌رسد که یک فرایند چرخه‌وار اشیا را به این دلیل که رتبه‌ی بالاتری دارند، در مرتبه‌ی بالاتر قرار می‌دهد. مکرراً توجه کمی به عدم حتمیت در هریک از این رتبه‌بندی‌ها می‌شود و تحلیل ناچیزی از چگونگی تأثیرگذاری این عدم حتمیت (برای مثال تغییرات سالیانه ضریب تأثیر) روی رتبه‌بندی‌ها انجام می‌شود. نهایتاً عوامل ابهام آمیز (برای مثال یک نظام خاص، نوع مقالاتی که یک مجله چاپ می‌کند، اینکه آیا یک دانشمند خاص یک آزمایش کننده است یا یک نظریه‌پرداز) مکرراً در چنین رتبه‌بندی‌هایی به خصوص در ارزیابی کارایی‌های ملی نادیده گرفته می‌شود.

۳.٦ تفسیر و تأثیر

«مقایسه‌هایی که در این مقاله مورد بحث قرار گرفته‌اند مورد علاقه‌ی عموم می‌باشند، و این به وضوح موضوعی است که در آن توجه دقیق به محدودیت‌ها هم حیاتی

است و هم احتمال نادیده گرفته شدن آن وجود دارد. اینکه آیا نتایج تصحیح شده به طریقی معیارهای معتبری از کیفیت مؤسسه‌ای هستند موضوع مهمی است، اما تحلیل‌گران همچنان باید آگاه به اثرات احتمالی نتایج بر شکل تغییرات رفتاری آینده توسط مؤسسات و افرادی که در جستجوی بهبود رتبه‌بندی‌های بعدی هستند، باشند»[۱۹]. ارزیابی و پژوهش نیز مورد علاقه عموم است. برای یک دانشمند یک ارزیابی می‌تواند اثرات عمیق بلند مدت روی حرفه‌ی وی داشته باشد؛ برای یک دپارتمان می‌تواند چشم‌اندازهای آن را برای موفقیت در آینده‌ی دور تغییر دهد؛ برای نظام‌ها مجموعه‌ای از ارزیابی‌ها می‌تواند تفاوتی مانند پیشرفت و پسرفت ایجاد کند. برای چنین امر مهمی مطمئناً باید در درک اعتبار و محدودیت‌های ابزار مورد استفاده در انجام آن بکوشیم. ارجاعات تا چه اندازه‌ای کیفیت پژوهش را می‌سنجند؟ به نظر می‌رسد که تعداد ارجاعات با کیفیت همبسته باشد و یک درک شهودی می‌گوید که مقالات با کیفیت بالا تعداد ارجاعات بیشتری دارند. اما همانگونه که در بالا توضیح داده شد، برخی مقالات به خصوص، در برخی نظام‌ها، به‌خاطر دلایلی به جز کیفیت بالا مورد ارجاع قرار می گیرند و لزوماً نمی‌توان نتیجه گرفت که مقالات با تعداد ارجاعات بالا دارای کیفیت بالا هستند. تفسیر دقیق رتبه‌بندی‌های بر مبنای آمار ارجاعات نیازمند درک بهتر است. علاوه بر این، اگر آمار ارجاعات نقش اساسی در ارزیابی پژوهش بازی می‌کند، واضح است که نویسندگان، ویراستاران، و حتی منتشرکنندگان راه‌هایی برای دست کاری سیستم به نفع خود پیدا می‌کنند[۲۴]. معنای بلند مدت این موضوع روشن نبوده و مورد مطالعه قرار نگرفته است.

مقاله‌ی نوشته شده توسط گلداستین و اسپیگل هالتر امروزه بسیار ارزشمند است زیرا روشن می‌سازد که تکیه‌ی بیش از حد بر آماره‌های ساده—انگار در ارزیابی پژوهش برنامه‌ای منحصر به فرد نیست. در گذشته دولت‌ها ، مؤسسات و افراد با مشکلات در دیگر زمینه‌ها منازعه داشته‌اند و طرقی برای درک بهتر ابزارهای آماری و الصاق آنها با دیگر وسائل ارزیابی یافته‌اند. گلداستین و اسپیگل هالتر مقاله‌ی خود را با عبارت امیدوار کننده‌ای به پایان می‌برند:

«نهایتاً، با اینکه ما به‌طور کلی از بسیاری تلاش‌های رایج برای فراهم کردن داوری‌هایی در مورد مؤسسات انتقاد کرده‌ایم، اما قصد نداریم این گمان را به وجود آوریم که ما باور داریم کلیه‌ی چنین مقایسه‌هایی معیوب‌اند. چنین به نظر می‌رسد که مقایسه‌ی مؤسسات و تلاش برای درک اینکه چرا مؤسسات متفاوت‌اند فعالیت خیلی مهمی است و بهتر است با روحیه‌ی همکاری و نه منازعه صورت بگیرد. این شاید تنها روش مطمئن برای حصول اطلاعات مبتنی بر عینیت است که می‌تواند منجر به درک شود و نهایتاً اصلاح امور را نتیجه دهد. مشکل اصلی در روش‌های ساده که ما آنها را مورد انتقاد قرار دادیم این است که آنها تمرکز توجه و منابع را از این هدف با ارزش از بین می‌برند»[۱۹]. شاید یافتن اظهار بهتری برای نشان دادن اهدافی که همه‌ی افراد مشمول در ارزیابی پژوهش باید در آن سهم داشته باشند، دشوار باشد.

مراجع

[1] Adler, R. (2007), The impact of impact factors, *IMS Bulletin*, 36(5), 4.

[2] Amin, M. and Mabe, M. (2000), Impact factor: use and abuse, *Perspectives in Publishing*, 1, 1-6.

[3] Batista, P. D., Campiteli, M.G., Kinouchi, O. and Martinez, A.S. (2005), Universal behavior of a research productivity index, *ar Xiv: physics*, 1, 1-5.

[4] Batista, P. D., Campiteli, M.G. and Kinouchi, O. (2006), Is it possible to compare researchers with different scientific interests?, *Scientometrics*, 68(1), 179-189.

[5] Bergstrom, C. (2007), Eigenfactor: measuring the value and prestige of scholarly journals, *College and Research Libraries News*, 68(5).

[6] Best, J. (2001), *Damned Lies and Statistics: Untangling the Numbers from the Media Politicians and Acivities*, University of California Press, Berkeley.

[7] Bird, S., et al. (2005), Performance indicators: good bad and ugly; Report of working party on performance monitoring in the public services, *J.R.Statist. Soc. A*, 168, Part 1, 1-27.

[8] Brooks, T. (1986), Evidence of complex citer motivations, *Journal of the American Society for Information Science*, 37(1), 34-36.

[9] Cary, A. L., Cowling, M. G. and Taylor, P. G. (2007), Assessing research in the mathematical sciences, *Gazette of the Australian Math Society*, 34(2), 84-89.

[10] Cozzens, S. E. (1989), What do citations count? The rhetoric-first model, *Scientometrics*, 15(5-6), 437-447.

[11] Egghe, L. (2006), Theory and practice of the g-index, *Scientometrics*, 69(1), 131-152.

[12] Evidence Report (2007), *The Use of Bibliometrics to Measure Research Quality in the UK Higher Education System*,(A report produced for the Research Policy Committee of Universities UK by Evidence Ltd a company specializing in research performance analysis and interpretation. Evidense Ltd. Has "strategic alliance" with Thompson Scientific.)

[13] Ewing, J. (2006), Measuring Journals, *Notices of the AMS*, 53(9),1049-1053.

[14] Garfield, E. (1955), Citation indexes for science: A new dimension in documentation through association of ideas, *Science*, 122(3159), 108-11.

[15] ——- (1972), Citation analysis as a tool in journal evaluation, *Science*, 178(4060), 471-479.

[16] ———(1987) Why are the impacts of the leading medical journals so similar and yet so different?, *Current Comments 2*, p. 3.

[17] ———(1998) Long-term vs. short-term journal impact (part II), *The Scientist*, 12(14), 12-3.

[18] ———(2005), Agony and the ecstasy- the history and meaning of the journal impact factor, Presented at the *International Congress on Peer Review and Bibliomedical Publication*, Chicago.

[19] Goldstein, H. and Spiegelhalter, D. J. (1996), League tables and their limitations: Statistical issues in comparisons of institutional performance, *Journal of Royal Statistcal Society A*, 159(3), 385-443.

[20] Hall, P. (2007), Measuring research performance in the mathematical sciences in Australian universities, *The Australian Mathematical Society Gazzete*, 34(1), 26-30.

[21] Hirsch J. E. (2006), An index to quantify an individual's scientific research output, *Proc. Natl. Acad. Sci. USA*, 102(46), 16569-16573.

[22] Kinney, A. L. (2007). National scientific facilities and their science impact on nonbiomedical research, *Proc. Natl. Acad. Sci. USA*, 104(46), 17943-17947.

[23] Lehman, Sune; Jackson, Andrew D.; Lautrup, B. E. (2006), Measures for measures, *Nature*, 444(21) , 1003-1004.

[24] Macdonald, S. and Kam, J. (2007), Aardvark et al.: quality journals and gamesmanship in management studies, *Journal of Information Science*, 33, 702-717.

[25] Martin, B. R. and Irvine, J. (1983), Assessing basic research, *Research Policy*, 12, 61-90.

[26] Meho, L. and Yang, K. (2007), Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar, *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.

[27] Molinari, J. F. and Molinari, A. (2008), A new methodology for ranking scientific institutions. To appear in *Scientometrics*.

[28] Monastersky, R. (2005), The number that's devouring science, *Chronicle Higher Ed*, 52(8).

[29] Rossner, M., Van Epps, H. and Hill, E. (2007), Show me the data, *Journal of Cell Biology*, 179(6), 1091-1092.

[30] Seglen, P. O. (1997), Why the impact factor for journals should not be used for evaluating research, *BMJ*, 314-497 .

[31] Sidiropoulos, A., Katsaros, D. and Manolopoulos, Y. (2006), Genelalized h-index for disclosing latent facts in citation networks, *ar Xiv: cs*, 1.

[32] Stringer, M. J., Sales-Pardo M. and Nunes Amaral L. A. (2008), Effectiveness of journal ranking schemes as a tool for locating information,*Plos one* 3(2): e1683.

[33] THOMPSON: JOURNAL CITATION REPORTS, (2007), (Thompson Scientific website)

[34] THOMPSON: SELECTION, (2007), (Thompson Scientific Website)

[35] THOMPSON: IMPACT FACTOR, (2007), (Thompson Scientific website)

[36] THOMPSON: HISTORY, (2007), (Thompson Scientific website)

[37] THOMPSON: FIFTY YEARS, (2007), (Thompson Scientific website)