

# هشدار اتحادیه بین‌المللی ریاضی

## درباره استفاده نادرست از آمار ارجاعات



از استخدام یا ارتقای کادر علمی و پژوهشگران تا مقایسه و رتبه‌بندی دانشمندان و دانشگاه‌ها و گروه‌ها و مجله‌های پژوهشی، این نوع مقایسه‌ها برای شناسایی بهترین عملکردها و تعیین استاندارد کارپژوهشی – و در نهایت، حفظ و ارتقای کیفیت آن – ضرورت دارد. تباریان، ارزیابی کیفیت پژوهش پدیده تاریخی نیست؛ چیزی که نسبتاً تاریکی دارد، رواج این عقیده است که ارزیابی باید با روش‌های «ساده و عینی» انجام شود، و متنظر از روش‌های «ساده و عینی»، هم استفاده از چند شاخص آماری است که از داده‌های ارجاع استخراج می‌شوند. نظر رایج این است که چنین شاخص‌هایی محکم‌های دقیق‌تر و بین‌طرفانه‌تری برای ارزیابی اند زیرا اعداد ساده را جانشین قضاوت‌های پیچیده شخصی می‌کنند و از عیوب «ذهنی بودن» که نقد توسط همکاران (peer review) دچار آن است متراحتند، ولی این اعتقاد به دلایل زیر مبنای چندان محکم ندارد:

- اگر این شاخص‌های آماری درست درک نشوند و درست به کار نتوانند، اتکا به آنها شیوه دقیقی برای ارزیابی نیست و حتی ممکن است به گمراهمی بینجامد ... تعییر و اعتبار این شاخص‌ها می‌ستی بر دریافت شهودی است [که خود امری ذهنی است].

با سمه تعالی

### در این شماره:

- هشدار اتحادیه بین‌المللی ریاضی درباره ...
- آیا می‌توان آینده علم را در ایران پیش‌بینی کرد؟
- مؤسسه ریاضی کلی: ...
- جایزه ریاضیدانان جوان ۱۳۸۵
- اولین عضو هیأت علمی پژوهشکده نجوم و اخترفیزیک
- جایزه «دانشمند جوان» برای یاسمن فرزان
- خبرها و گزارش‌ها

اتحادیه بین‌المللی ریاضی (IMU) در تاریخ ۱۱ زوئن ۲۰۰۸ گزارشی درباره شاخص‌های می‌ستی بر تعداد ارجاع‌ها در ارزیابی پژوهش‌های علمی منتشر کرده است و جامعه علمی را از انگاری پیش از حد به این گونه معیارها و اعتقاد کور‌گوارانه به دقت، عیبیت، و سادگی آنها برخور داشته است. نویسنده‌گان گزارش با اذعان به سودمندی این شاخص‌ها (لااقل از این نظر که به آسانی قابل محاسبه‌اند و اخلالات خوبی به دست می‌دهند) هشدارهایی درباره کاربرد و تعییر نادرست آنها مطرح کرده‌اند و به هر حال بر این عقیده‌اند که تتابع حاصل از آمار ارجاعات هرچند می‌تواند بخشی از قرایب ارزیابی باشد ولی نباید جانشین همه معیارها و روش‌های دیگر شوند.

این گزارش را اتحادیه بین‌المللی ریاضی با همکاری شورای بین‌المللی ریاضیات مستعمر و کاربردی (ICIAM) و مؤسسه امار ریاضی (IMS) تهییه کرده است. کمیسیون سه‌جانبه‌ای که از طرف این سه نهاد برای پرسی «از ریاضی کنی پژوهش» در ۲۱ مه ۲۰۰۷ تشکیل شده بود کار تهییه گزارش را با کندوکار در انتوه نوشتگان موجود در این زمینه و گزارش‌های ریاضیدانان و دانشمندان دیگر در ۶ زوئن ۲۰۰۸ به پایان رساند. هرچند نتایجی از این گزارش حاوی مطالعه مساله حول حوش ISI است که پاره‌ای در اخبار نظایر آن را خوانده‌اید، نظر به اهمیت موضوع و اهمیت ارگان انتشاره‌هسته آن، ترجمه از ادبی از گزارش را با تطبیص در اینجا می‌آوریم. در این حلاصه‌سازی، مقدمه‌های گزارش را در هم ادحام کرده و ارجاعات را نیز حذف کرده و کمی هم از متن کاسته‌ایم ولی کوشیده‌ایم که مضمون و پیام اصلی گزارش عیناً به خواننده منتقل شود. علاوه‌های دانشمندان به متن اصلی و کامل گزارش می‌توانند آن را در نشانی زیر ببینند:

<http://www.mathunion.org/Publications/Report/CitationStatistics>

### مقدمه

حکومت‌ها و نهادهایی که مبالغه قابل ملاحظه‌ای به پژوهش‌های علمی کنک می‌کنند طبعاً می‌خواهند مطمئن شوند که بول آنها عاقله‌هه مصرف می‌شود؛ می‌خواهند کیفیت پژوهش‌هایی که به هزینه آنها صورت می‌گیرد ارزیابی شود تا باید سرمایه‌گذاری‌های بعدی تصمیمات آگاهانه‌ای بگیرند. همچنین، کیفیت کارپژوهشی یکی از مبانی مهم تصمیم‌گیری در بسیاری از امور آکادمیک است.

باشند. مجموعه ارجاعات، اطلاعاتی درباره مجله‌ها، مقاله‌ها، و داشمندان به دست می‌دهد. ما نمی‌خواهیم این اطلاعات از نظر دور بماند، بلکه می‌خواهیم معنای آنها روش شود.

آلبرت اینشتین گفته است: «هر چیز را تا حد امکان باید ساده کرد ولی نه بیش از این حد!» این نصیحت را باید هنگام ارزیابی پژوهش‌های علمی آویزه‌گوش کرد.

استفاده از داده‌های ارجاع برای ارزیابی پژوهش درنهایت به معنای کاربرد این داده‌ها برای رتبه‌بندی مجله‌ها، مقاله‌ها، پژوهشگران، و برنامه‌ها، و رشته‌های است:

## رتبه‌بندی مجله‌ها

این رتبه‌بندی براساس ضریب تأثیر (impact factor) صورت می‌گیرد که در دهه ۱۹۶۰ ابداع شد و مبتنی است بر محاسبه تعداد متوسط ارجاعات به ازای هر مقاله در دوره مشخصی از زمان. به بیان دقیق‌تر، برای یک مجله و یک سال خاص، ضریب تأثیر برابر است با متوسط تعداد ارجاعات داده شده در آن سال به مقالات چاپ شده در آن مجله طی دو سال قبل از آن. مثلاً اگر ضریب تأثیر مجله‌ای در سال ۲۰۰۷ برابر با ۱/۵ باشد معنی آن این است که هر مقاله چاپ شده در آن مجله در سال‌های ۲۰۰۵ و ۲۰۰۶ به طور متوسط ۱/۵ بار در سال ۲۰۰۷ مورد ارجاع قرار گرفته است. اطلاعات مربوط به ارجاع Institute for Scientific Information (ISI) را مؤسسه‌ای که قبلاً Thomson Scientific نام داشت و اکنون Thomson Scientific نامیده می‌شود گرد می‌آورد و هر ساله اطلاعات جدید را به آن می‌افزاید (تعداد مجله‌هایی که در فهرست این مؤسسه قرار دارند بیش از ۹۰۰۰ است ولی تعداد مجله‌های ریاضی موجود در این فهرست، کمتر از نصف مجله‌هایی است که زیر پوشش مستحبه‌کال ریویویز (Mathematical Reviews) و تسنترالبلاط (Zentralblatt)، (دو مجله عمده نقد و معرفی در ریاضیات)، هستند. مؤسسه‌ی تامسون استفاده از ضریب تأثیر را برای مقایسه مجله‌ها (به منظور مثلاً، خرید مهمترین مجلات برای کتابخانه‌ها) تشویق و ترغیب می‌کند.

بسیاری از اهل علم خاطرشنان کرده‌اند که ارزش آکادمیک یک مجله را نباید فقط براساس داده‌های ارجاع سنجید. ضریب تأثیر در واقع متوسط ساده توزیع تعداد استنادهای مقالات یک مجله است و این متوسط فقط اطلاع اندکی درباره آن توزیع می‌دهد و شاخصی نسبتاً خام است. اتکایی صرف به ضریب تأثیر برای داوری درباره یک مجله مانند آن است که برای قضایت درباره سلامت یک شخص فقط وزن او مورد نظر قرار گیرد. به علاوه، ضریب تأثیر از جهات دیگری نیز مورد انتقاد قرار گرفته است که اهم آنها از این قرار است.

(یک) در نظر گرفتن ضریب تأثیر به عنوان متوسط، کار چندان درست نیست. بسیاری از مجلات مطالبی از قبیل سرمقاله و نامه به سردبیر و نظری اینها چاپ می‌کنند که آمار آنها در مخرج کسر ضریب تأثیر به حساب نمی‌آید. ولی، هر چند به ندرت، گاه ارجاعاتی به این گونه مطالب می‌شود

- هرچند اعداد ظاهراً «عینی» هستند ولی این عینی بودن ممکن است مبتنی بر توهم باشد. معنای ارجاع و استناد ممکن است حتی «ذهنی» تراز نقد آثار توسط همکاران باشد. اما چون ذهنی بودن شاخص‌های ارجاع وضوح کمتری دارد، به کاربرندگان آنها بیشتر محتمل است که از محدودیت‌های آنها آگاه نباشند.

- اتکایی صرف به داده‌های ارجاع، در بهترین حالت، شناختی ناقص و اغلب کم‌عمق از جریان پژوهش به دست می‌دهد، شناختی که فقط وقتی معتبر است که با انواع دیگری از قضایت تقویت شود. خلاصه اینکه اعداد ذاتاً برتر از قضایت‌های صحیح نیستند.

پژوهش معمولاً هدف‌های چندگانه‌ای، هم در درازمدت و هم در کوتاه‌مدت، دارد و بنابراین معقول است که ارزش آن با معیارهای چندگانه سنجیده شود. ریاضیدانان می‌دانند که چیزهای سیاری، واقعی و انتزاعی، وجود دارند که نمی‌توان آنها را مرتب کرد، به این معنی که نمی‌توان هر دو تا از آنها را با هم مقایسه کرد. مقایسه غالباً مستلزم تحلیل پیچیده‌تری است که گاه ما را در مورد اینکه کدام یک از دو چیز «بهتر» است به نتیجه قطعی نمی‌رساند. پاسخ درست به پرسش «کدام یک بهتر است؟» گاهی این است: «بستگی دارد!» انتشارات پژوهشی را هم از راه‌های متنوعی، و نه فقط آمار ارجاعات، می‌توان ارزیابی و مقایسه کرد. ملاک‌هایی که معمولاً نشانه ارج و اعتبار کارهای پژوهشی اشخاص هستند، مانند دعوت شدن آنها به کنفرانس‌ها و سخنرانی‌ها، عضویت در هیأت ویراستاران مجله‌ها، دریافت جایزه‌ها و تقديرها و در بعضی کشورها، دریافت کمک‌هزینه پژوهشی، از جمله معیارهایی هستند که می‌توانند در ارزیابی مفید واقع شوند. نقد توسط همکاران مؤلفه مهمی از ارزیابی است. نباید قضایت همکاران را کلان‌نادیده گرفت به این عنوان که احتمال جانبداری یا غرض‌ورزی در آن وجود دارد، همان‌طور که نمی‌توان شاخص‌های ارجاع را کلان‌نادیده گرفت به این عنوان که ممکن است نابه‌جا به کار بروند. خلاصه، راه‌های فراوانی برای ارزیابی پژوهش وجود دارد که آنچه در بالاگفته فقط بخشی از آنهاست، اما جذبۀ فرایند‌های ساده و معیارهای ساده (تجیح) یک تک عدد در ارزیابی، ظاهراً دارد بر عقل سلیم و قضایت صحیح غلبه می‌یابد. این گزارش را چند ریاضیدان درباره استفاده نادرست از شاخص‌ها در ارزیابی پژوهش نوشتند. فرهنگ و عرف خاص ارجاع دهی در ریاضیات، پایین بودن تعداد ارجاعات در این رشته نسبت به سایر رشته‌ها، این رشته را در برای سوء تعبیر و سوء کاربرد شاخص‌ها آسیب‌پذیرتر ساخته و یکی از انگیزه‌های تهیه این گزارش همین است. ولی ما عقیده داریم که همه داشمندان، و کل جامعه علمی، باید در مورد صحیح بودن روش‌ها در ارزیابی پژوهش‌های علمی دغدغه داشته باشند.

بعضی اعضای جامعه علمی، در اکنونی منفی به استفاده نادرست از شاخص‌های ارجاع در گذشته، کلان‌شدن شاخص‌ها را کنار می‌گذارند. ولی ما معتقدیم که به این طریق، ابزار مفیدی از دست می‌رود. این شاخص‌ها می‌توانند نقشی در ارزیابی پژوهش داشته باشند به شرط آنکه درست به کار بروند، با احتیاط تعبیر شوند، و فقط بخشی از فرایند ارزیابی

مجلات کوچک، کوچک است. از طرف دیگر، مجله‌ها را اغلب برای یک سال مشخص با هم مقایسه می‌کنند و توجهی به تغییرات زیادتر ضریب تأثیر در مورد مجلات کوچک‌تر ندارند.

(پنج) به مجله‌هایی که به زبانی غیر از انگلیسی انتشار می‌یابند، ارجاعات کمتری داده می‌شود چون بخش بزرگی از جامعه علمی جهانی نمی‌تواند آنها را بخواند (یا نمی‌خواند). همچنین نوع مجله، نه کیفیت به تنها بیان ممکن است اثری در ضریب تأثیر داشته باشد. مثلاً به مجلاتی که مقالات مروری چاپ می‌کنند بسیار بیشتر از مجلاتی که این نوع مقالات را چاپ نمی‌کنند ارجاع داده می‌شود و بنابراین، مجله‌های نوع اول ضریب تأثیری بزرگ‌تر (گاه بسیار بزرگ‌تر) دارند.

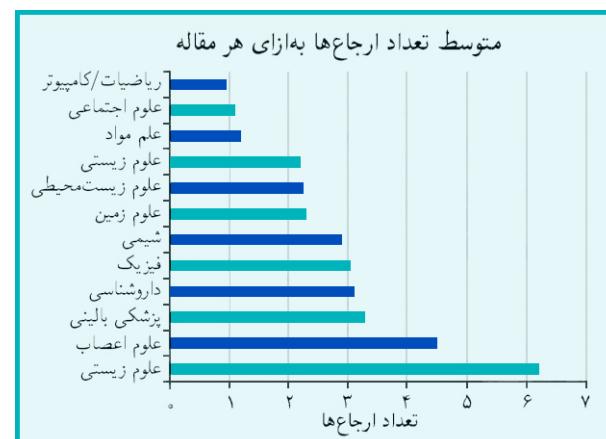
(شش) مهمترین ایرادی که به ضریب تأثیر وارد است، این است که معنای آن کاملاً مشخص نیست. در هنگام استفاده از ضریب تأثیر برای مقایسه دو مجله، هیچ مدل پیشینی وجود ندارد که معنای «بهتر» را تعریف کند. تنها مدل، از خود ضریب تأثیر استنتاج می‌شود، یعنی ضریب بالاتر به معنی بهتر بودن مجله قلمداد می‌شود. حال آنکه در روش کلاسیک آماری، ابتدا مدلی تعریف می‌کنند، فرضی (مبنی بر عدم وجود تفاوت) را بیان می‌کنند، و سپس آماره‌ای [شاخصی آماری] می‌یابند که بر حسب مقادیرش می‌توانند آن فرض را قبول یا رد کنند. استنتاج اطلاعات (و احتمالاً یک مدل) از خود داده‌ها رویکرد مشروعی در تحلیل آماری است ولی در این حالت روشن نیست چه اطلاعاتی استخراج شده است. ضریب تأثیر چگونه کیفیت را می‌سنجد؟ آیا بهترین شاخص برای سنجش کیفیت است؟ دقیقاً چه چیزی را می‌سنجد؟ (بحث بعدی ما درباره «معنای ارجاع» در اینجا موضوعیت دارد). اطلاعات ما درباره مدل کیفیت مجلات و اینکه چگونه این مدل به ضریب تأثیر مربوط می‌شود خیلی کم است.

شش ایراد بالا بر ضریب تأثیر وارد است ولی از آنها فقط این نتیجه را می‌توان گرفت که ضریب تأثیر شاخص خام و نادقیقی است نه اینکه کاملاً هایده است. مثلاً از ضریب تأثیر می‌توان به عنوان نقطه شروعی برای دسته‌بندی مجله‌ها استفاده کرد و سپس بقیه ملاک‌ها را برای دقیق کردن رده‌بندی به کار برد. ولی استفاده از ضریب تأثیر برای ارزیابی مجله‌ها نیازمند دقت است. مثلاً این ضریب را نمی‌توان برای مقایسه مجله‌های رشته‌های مختلف به کار برد و به نوع مجله‌ها در هنگام رده‌بندی آنها باید توجه داشت. همچنین باید تغییرات سالانه را، به خصوص در مورد مجلات کوچک، در نظر گرفت و درک کرد که تفاوت‌های اندک ممکن است کاملاً تصادفی باشند. توجه به این نکته نیز مهم است که در بعضی رشته‌ها ضریب تأثیر نمی‌تواند همه ارجاع‌ها را بازتاب دهد، هم به دلیل اینکه همه مجله‌ها در نظر گرفته نمی‌شوند و هم به دلیل کوتاه بودن دوره در نظر گرفته شده. شاخص‌های دیگری که مبتنی بر دوره‌های زمانی طولانی تر باشند و لحاظ کردن مجلات بیشتر ممکن است کیفیت را بهتر نشان دهند، و بالاخره، در نظر گرفتن ارجاعات فقط یک راه برای داوری درباره مجلات است که باید با اطلاعات دیگری تکمیل شود. (که این پیام اصلی این گزارش

که در صورت کمتر به حساب می‌آید. پس، ضریب تأثیر با متوسط تعداد ارجاع‌ها به ازای هر مقاله کاملاً تطبیق نمی‌کند. در مجلاتی که تعداد زیادی از این گونه مطالب چاپ می‌کنند، این انحراف ممکن است زیاد باشد ولی البته در سیاری از رشته‌ها، از جمله ریاضیات، چنین نیست.

(دو) دوره دوسراله در محاسبه ضریب تأثیر به این منظور در نظر گرفته شده که این ضریب روزآمد باشد. این دوره برای رشته‌هایی مانند علوم زیست‌پژوهشی مناسب است زیرا بیشتر مقاله‌های چاپ شده آنها بیشترین ارجاعات خود را مدت کوتاهی پس از انتشار به دست می‌آورند. ولی در بسیاری از رشته‌های دیگر، از قبیل ریاضیات، بیشتر ارجاع‌ها پس از دو سال صورت می‌گیرند. با بررسی بیش از ۳ میلیون مورد ارجاع اخیر در مجلات ریاضی (پایگاه‌داده‌های ارجاع متریوپوز) دیده می‌شود که تقریباً ۹۰٪ از ارجاعات به یک مجله در خارج از این محدوده دوسراله داده می‌شود. در نتیجه، ضریب تأثیر در واقع برمبنای ۱۰٪ از ارجاعات محاسبه می‌شود و اکثریت ارجاع‌ها نادیده گرفته می‌شوند.

(سه) ضریب تأثیر ممکن است از رشته‌ای به رشتۀ دیگر تفاوت قابل ملاحظه‌ای داشته باشد. بخشی از تفاوت از نکته (دو) ناشی می‌شود: اگر در رشته‌هایی بسیاری از ارجاع‌ها پس از دوره دوسراله صورت گیرند، ضریب تأثیر مجله‌ها در آن رشته‌ها به مراتب پایین‌تر خواهد بود. بخشی دیگر از اختلاف هم از تفاوت فرهنگ ارجاع دادن در رشته‌های مختلف منشأ می‌گیرد یعنی دانشمندان در رشته‌های گوناگون به دلایل متفاوت و میزان‌های متفاوت ارجاع می‌دهند (به این نکته بعداً در بحث «معنای ارجاع» باز خواهیم گشت زیرا مفهوم ارجاع فوق العاده مهم است).



نمودار تعداد ارجاعات برای چند رشته مختلف که نشان‌دهنده تفاوت بارز رشته‌ها در ارجاع‌دهی است

(چهار) ضریب تأثیر ممکن است از سالی به سال دیگر بسیار تغییر کند، و این تغییر برای مجلات کوچک‌تر، زیادتر است. مثلاً برای مجله‌هایی که کمتر از ۵۰ مقاله چاپ می‌کنند، تغییر متوسط ضریب تأثیر از ۲۰۰۰ تا ۲۰۵۰ بقریباً ۵٪ بوده است. این امر کاملاً قابل انتظار است زیرا اندازه نمونه برای

کمتر از ده صفحه، چاپ می‌کند. در این دوره، این مجله ۲۳۸۱ مقاله (در حدود ۱۵۰۰۰ صفحه) چاپ کرده است. با مراجعه به اطلاعات مربوط به مجله‌های سال ۲۰۰۵ در پایگاه داده‌های ارجاع متریویز می‌بینیم متوسط تعداد ارجاعات به ازای هر مقاله (یعنی ضریب تأثیر) برای این مجله (۴۳٪) است. مجله ترازنزکشنز متعلق به همین انجمن (*Transactions of the AMS*) مقالاتی بلندتر و معمولاً اساسی‌تر، هم از لحاظ دامنه موضوع و هم از لحاظ محتوا، چاپ می‌کند. طی همین دوره زمانی، مجله دوم ۱۱۶۵ مقاله (در بیش از ۲۵۰۰۰ صفحه) منتشر داده است که شمار ارجاع‌های آنها از ۰ تا ۱۲ است. تعداد متوسط ارجاع‌ها به ازای هر مقاله، ۸۴۶٪، یعنی تقریباً دو برابر مقالات پراسیدینگر است.



حال دو ریاضیدان را در نظر بگیرید که یکی از آنها مقاله‌ای در پراسیدینگز و دیگری مقاله‌ای در ترازنزکشنز چاپ کرده است. بنا به روال مرسوم که در بالا ذکر شد، ریاضیدان دومی بهتر از اولی به حساب می‌آید چون مقاله‌ای در مجله‌ای با ضریب تأثیر بالاتر منتشر داده است. آیا این ارزیابی معین است؟ آیا هر مقاله مجله دوم دو برابر هر مقاله اول ارزش دارد؟ وقتی ادعا می‌کنیم مقاله‌ای در مجله دوم (براساس ضریب تأثیر مجله) بهتر از مقاله‌ای در مجله اول است، باید ببینیم احتمال درستی ادعای ما چقدر است. چقدر احتمال دارد مقاله خاصی از پراسیدینگر که به تصادف انتخاب شده، دست‌کم همان قدر ارجاع داشته باشد که مقاله‌ای از ترازنزکشنز که آن هم به تصادف انتخاب شده است؟ محاسبه این احتمال ساده است و پاسخ آن ۶۲٪ است. یعنی در ۶۲ درصد موارد، ادعای ما غلط است و مقاله‌ای تصادفی از مجله اول به خوبی مقاله‌ای تصادفی در مجله دوم (یا بهتر از آن) است — هرچند ضریب تأثیر مجله اول فقط نصف ضریب تأثیر مجله دوم است. خیلی‌ها ممکن است از این موضوع تعجب کنند ولی این امر نتیجه توزیع بسیار چاوله و کوتاه بودن دوره زمانی است که برای محاسبه ضریب تأثیر به کار می‌رود. این وضع به هیچ وجه

است. خود مؤسسه تامسون هم با این نظر موافق است و در بیانیه‌ای می‌گوید: «ضریب تأثیر نباید بدون توجه به بسیاری عوامل دیگر که ممکن است بر میزان ارجاعات تأثیر بگذارند به کار رود ... این ضریب باید همراه با نقد آگاهانه توسعه همکاران به کار برد شود ...»

## رتبه‌بندی مقاله‌ها

استفاده غلط از ضریب تأثیر منحصر به رتبه‌بندی مجله‌ها نیست بلکه بدتر از آن، کاربرد نابهجهای این ضریب در مقایسه مقاله‌ها، دانشمندان، برنامه‌های پژوهشی، و حتی رشته‌های است که به مسئله رشیدابندی در مقیاس کشورها و رشته‌ها و ارزیابی‌های ملی تحقیقات تبدیل شده است.

در مورد مقاله‌ها به جای آنکه تعداد واقعی ارجاعات به مقاله مبنای مقایسه قرار گیرد، غالباً ضریب تأثیر مجله‌ای را که مقاله در آن چاپ شده است مبنای قرار می‌دهند. اعتقاد بر این است که بالا بودن ضریب تأثیر مجله به معنی بالا بودن تعداد ارجاع‌های هر مقاله است. حال آنکه اغلب چنین نیست! این امر، استفاده غلطی از آمار است که هرگاه و هرجا مطرح می‌شود باید به چالش طلبیده شود. البته، کیفیت مجله‌ای که یک دانشمند معمولاً مقالات خود را در آنها چاپ می‌کند یکی از معیارهای متدالوی است که می‌تواند در ارزیابی کلی کار پژوهشی او به کار رود، ولی مسئله این است که ضریب تأثیر، تمايل گسترشده‌ای پدید آورده است که کیفیت یک مجله به هر یک از مقالات آن (و حتی هر یک از نویسندهان آن) منسوب شود. مثلاً در دانشگاهی، ارتقای عضو هیأت علمی براساس فرمولی انجام می‌شود که عبارت از تعداد مقالات تکنفره از ضرب در ضریب تأثیر مجله‌هایی است که مقاله‌ها در آنها چاپ شده‌اند. استخدام اعضای جدید هم تا حدودی براساس این فرمول است. در دانشگاهی دیگر، مهمترین معیار ارزیابی مبتنی است بر رتبه‌بندی مقاله‌های شخص بر این اساس که هر مقاله ۳ امتیاز می‌گیرد اگر در مجله‌ای که جزو ثبت اول مجلات ISI است چاپ شده باشد، ۲ امتیاز می‌گیرد اگر در مجلات ثالث دوم باشد و ۱ امتیاز می‌گیرد اگر در مجله‌ای که جزو مجلات انتهایی فهرست ISI است چاپ شده باشد. در این نمونه‌ها و در نمونه‌های دیگری که به نویسندهان این مقاله شده باشد. در این نمونه‌ها و در نمونه‌های دیگری که به نویسندهان آنها به کار می‌رود: اگر ضریب تأثیر مجله A بیشتر از مجله B باشد، می‌سلم فرض می‌شود که هر مقاله چاپ شده در A بیشتر از هر مقاله چاپ شده در B است و نویسنده آن هم برتر و مهمتر است. این استدلال در مواردی به رتبه‌بندی دانشگاهها و حتی رشته‌ها تعیین داده می‌شود. از مدت‌ها قبل معلوم شده است که توزیع تعداد ارجاعات به مقاله‌های چاپ شده در یک مجله به شدت چاوله و تقریباً حاکمی از قانون توانی است. این امر پیامدهای دارد که آنها را ذکر یک مثال روشن می‌کنیم. این توزیع را در مورد مقالات چاپ شده در مجله پراسیدینگز متعلق به انجمن ریاضی آمریکا (*Proceedings of the AMS*) طی دوره ۲۰۰۴–۲۰۰۵ در زیر می‌بینید. پراسیدینگز معمولاً مقالات کوتاهی در

شاخص برای تعدیل حالتی در نظر گرفته شده است که  $n$  مقاله  
بالایی پژوهشگر تعداد خارق العاده‌ای ارجاع داشته است.

هرش در همان مقاله‌ای که شاخص  $h$  را معرفی کرد آن را «شاخصی که به آسانی قابل محاسبه است و براوردی از اهمیت و تأثیر مجموعه آثار پژوهشی دانشمند به دست می‌دهد» نامید. ولی شواهد قانع‌کننده‌ای در تأیید این ادعا وجود ندارد. هرش برای تقویت ادعای خود شاخص  $h$  را برای عده‌ای از برنده‌گان جایزه نوبل (و به طور جداگانه برای عده‌ای از اعضای آکادمی ملی علوم آمریکا) تحلیل کرد و نشان داد که اعضا این گروه‌ها معمولاً شاخص‌های  $h$  بالایی دارند. می‌توان نتیجه گرفت که اگر دانشمندی برنده جایزه نوبل باشد شاخص  $h$  او احتمالاً بالاست. ولی اگر شاخص  $h$  شخصی بالا باشد، احتساب اینکه برنده نوبل یا عضو آکادمی شود معلوم نیست مگر آنکه اطلاعات دیگری در دست باشند.

وی در همان مقاله نوشت: «من از این نظر دفاع می‌کنم که دو نفر با شاخص  $h$  یکسان از لحاظ تأثیر علمی کلی خود مشابه‌اند حتی اگر تعداد کل مقاله‌ها یا تعداد کل ارجاع‌های آنها خیلی متفاوت باشد. به عکس، بین دو نفر (با سابقه کار علمی یکسان) که تعداد کل مقاله‌هایشان یا تعداد کل ارجاع‌هایشان یکی باشد اما مقدار  $h$  بسیار متفاوتی داشته باشند، آنکه مقدار  $h$  بالاتری دارد بیشتر محتمل است که شخص برجسته‌تری باشد.» چنین ادعا‌هایی را عقل سليم تأیید نمی‌کند. (دو پژوهشگر را در نظر بگیرید که هر یک  $10$  مقاله با  $10$  ارجاع دارد و یکی،  $90$  مقاله دیگر هم، هر یک با  $9$  ارجاع، دارد؛ یا فرض کنید یکی دقیقاً  $10$  مقاله با  $10$  ارجاع و دیگری  $10$  مقاله، هر یک با  $10$  ارجاع، دارد. آیا کارنامه پژوهشی این دو نفر با هم معادل است؟)

هرش ادعا می‌کند که «شاخص  $h$  بر سایر معیارهای تک عددی که معمولاً برای ارزیابی کار علمی پژوهشگران به کار می‌رود مرجع است» ولی او نه معنی «مرجع» را تعریف می‌کند و نه توضیح می‌دهد که چرا باید در جستجوی «معیار تک عددی» بود.

هرچند انتقاداتی به این شاخص وارد شده است، تحلیل جدی از آن به عمل نیامده است. بیشتر تحلیل‌ها معطوف به نشان دادن همبستگی شاخص  $h$  با سایر شاخص‌های انتشار/ارجاع، از قبیل تعداد مقاله‌های انتشاری‌افته یا تعداد کل ارجاعات بوده است. ولی این همبستگی مهم نیست زیرا همه این متغیرها تابع یک پدیده بنیادی — انتشارات — هستند. گونه‌های دیگری از شاخص  $h$  نیز برای مقایسه دانشمندان ابداع شده است که هدف از طراحی همه آنها استفاده از معیارهای ساده تک عددی به جای اطلاعات تفصیلی درباره ارجاعات است. با این معیارها، مقایسه هر دو پژوهشگر امکان‌پذیر می‌شود ولی سیاری از اطلاعات لازم برای شناخت دقیق کارنامه پژوهشی افراد کنار گذاشته می‌شود، حال آنکه هدف ارزیابی پژوهش باید شناخت دقیق باشد نه فقط حصول اطمینان از اینکه هر دو پژوهشگر را می‌توان با هم مقایسه کرد. در مواردی دیده می‌شود که گروه‌های ارزیابی در سطح ملی، به گردآوری داده‌های شاخص  $h$  به عنوان بخشی از

منحصر به این دو مجله نیست. مثلاً ضریب تأثیر جورنال انجمان ریاضی آمریکا (*Journal of the AMS*) در همین دوره  $2,632$  یعنی شش برابر پراسیدنی‌نگر بوده است. با این حال مقاله‌ای از پراسیدنی‌نگر که به تصادف انتخاب شود به احتمال  $32\%$  به خوبی مقاله‌ای از جورنال است.

پس هرچند درست نیست که بگوییم ضریب تأثیر هیچ اطلاعی در مورد تک‌تک مقاله‌های یک مجله به دست نمی‌دهد، این اطلاع سیار مبهم است و ممکن است گمراه کننده باشد. خلاصه اینکه، استفاده از ضریب تأثیر به عنوان جانشین تعداد واقعی ارجاعات هر مقاله، مبنای معقولی ندارد و صدور ادعاهایی که در نیمی (یا در ثلث) موارد غلط است، راه خوبی برای ارزیابی نیست.

اگر پنیزیریم که در نظر گرفتن ضریب تأثیر مجله به جای تعداد ارجاعات یک مقاله بی معنایست، نتیجه باید گرفت که مبنای قرار دادن این ضریب برای ارزیابی نویسنده‌گان مقاله‌ها و برنامه‌های پژوهشی که در آنها مشارکت دارند و نیز رشته‌هایی که در آنها کار می‌کنند معنایی ندارد. ضریب تأثیر و این گونه شاخص‌های متوسط به طور کلی برای مقایسه‌های معنادار مناسب نیستند مگر اینکه اطلاعات دیگری هم در دست باشد.

البته، رتبه‌بندی افراد با رتبه‌بندی مقاله‌های آنها یکی نیست ولی اگر بخواهید ارزش مقالات یک پژوهشگر را فقط با استفاده از ارجاعات به یک مقاله خاص او رتبه‌بندی کنید باید با شمارش ارجاعات آن مقاله شروع کنید و ضریب تأثیر مقاله‌ای که مقاله در آن چاپ شده جانشین قابل اعتمادی برای این کار نیست.

## رتبه‌بندی پژوهشگران

هرچند ضریب تأثیر تاکنون معروفترین شاخص مبتنی بر ارجاعات برای رتبه‌دادن به پژوهشگران بوده است، شاخص‌های جدیدتری نیز مطرح شده‌اند که عده‌ای به شدت از آنها طرفداری می‌کنند. سه تا از این شاخص‌ها عبارت‌اند از:

شاخص  $h$  برای یک پژوهشگر بزرگترین عدد  $n$  ای است که آن پژوهشگر  $n$  مقاله هر یک با دست کم  $m$  ارجاع داشته باشد. این شاخص را هرش (J.E. Hirsch) به منظور «سنجه کارنامه علمی پژوهشگر» براساس «دَم» بالای توزیع ارجاعات او ابداع کرده است.

شاخص  $m$  برای یک پژوهشگر، شاخص  $h$  او تقسیم بر تعداد سال‌هایی است که از انتشار اولین مقاله او گذشته است. هدف از طرح این شاخص، رعایت انصاف در مورد دانشمندان جوان و تازهکار است که فرصت نداشته‌اند مقالات زیادی مستشتر کنند یا ارجاعات زیادی به دست آورند.

شاخص  $g$  برای یک پژوهشگر، بزرگترین عدد  $n$  ای است که  $n$  مقاله پارچای او مجموعاً، دست کم  $n^2$  ارجاع داشته باشند. این

است، مثلاً در مقاله‌ای گفته شده است: «همه مشکلات استفاده از تعداد ارجاعات به عنوان شاخص کیفیت، ناشی از ناآگاهی ما از دلایل ارجاع دادن نویسنده‌گان به بعضی از کارها و ارجاع ندادن به کارهای مشابه است ... پیش‌فرض تحلیل ساده ارجاع، مدلی بسیار عقلایی از یادکرد آثار دیگران است که نشان‌دهنده قدرشناسی از کارهای علمی مهم قبلي است و اینکه همه ارجاع‌دهنده‌های بالقوه به یک میزان احتمال دارد که به کارهای خاص ارجاع بدھند. در مقاله دیگری این نظر مطرح می‌شود که ارجاع بر دونوع است، پکی به منظور قدرشناسی و استناد واقعی به نوشته دیگران انجام می‌شود و دیگری کم محتوا و لفاظانه است. معنی ارجاع نوع اول این است که مؤلف ارجاع‌دهنده، «دین علمی» خود را به مؤلف ارجاع‌شونده ادا می‌کند ولی ارجاع نوع دوم معنای کاملاً متفاوتی دارد و فقط اشاره به مقاله‌ای است که نتیجه‌ای را توضیح داده است و آن نتیجه ممکن است اصلاً متعلق به نویسنده آن مقاله نباشد. این گونه ارجاع‌های کم محتوا حداکثر می‌تواند در خدمت نوعی گفتگوی علمی باشد و نه ادای دین علمی. البته در مواردی، ارجاع ممکن است هر دو جنبه را داشته باشد. نویسنده‌گان این گزارش بیشتر ارجاعات را از نوع دوم می‌دانند و تجربه بیشتر ریاضیدانان دست‌اندرکار این نظر را تأیید می‌کند (مثلاً در پایگاه‌داده‌های متریپوز بیش از ۳۰٪ از ۳ میلیون ارجاع به کتاب‌های است و نه به مقالات تحقیقی در مجله‌ها). چرا این موضوع مهم است؟ زیرا برخلاف استنادهای قدرشناسانه و جدی که عمدها به مقاله‌های مهم تحقیقی انجام می‌شوند، ارجاع کم محتوى و غیرجدی به مقاله‌ای خاص، به عوامل زیادی بستگی دارد — از جمله، اعتبار مؤلف ارجاع‌شونده، رابطه مؤلفان ارجاع‌دهنده و ارجاع‌شونده، دسترس پذیری مجله (مثلاً ممکن است مجله‌هایی که دسترسی به آنها ریگان است شناسه بیشتری داشته باشند که مورد ارجاع قرار گیرند)، و عوامل دیگری که ارتباط مستقیمی با «کیفیت» مقاله ارجاع‌شونده ندارند. حتی استنادهای جدی هم ممکن است از انگیزه‌های مختلفی ناشی شده باشند. مؤلف ممکن است به قصد دادن هشداری به خوانده و یا بیان نظری منطقی درباره نقص کار دیگری، به آن ارجاع بدهد. دستاوردهای شایان توجه ممکن است به فاصله کوتاه جزو دستاوردهای بعدی قرار گیرد و ارجاعات به آن دستاوردهای بعدی داده شود. خلاصه اینکه، معنای ارجاع چندان ساده نیست و شاخص‌های مبتنی بر ارجاع هم، چندان که طرفدارانش ادعا می‌کنند، «عینی» نیست. بعضی ممکن است استدلال کنند که مفهوم ارجاع مهم نیست زیرا شاخص‌های مبتنی بر ارجاع همبستگی زیادی با سایر ملاک‌های کیفیت پژوهش (مانند نقد از طرف همکاران) دارند. مثلاً در گزارشی استدلال می‌شود که به علت این همبستگی، شاخص‌های مبتنی بر ارجاع می‌توانند (و باید) جانشین سایر شیوه‌های ارزیابی شوند: «شواهد نشان می‌دهند که با روش‌های مبتنی بر استناد می‌توان شاخص‌هایی از کیفیت پژوهش به دست آورد که با برداشت پژوهشگران سازگار باشد.» نتیجه ظاهراً این است که این نوع شاخص‌ها، صرف نظر از اینکه معنای دقیق آنها چیست، باید جانشین سایر روش‌های ارزیابی شوند زیرا اغلب با آنها همخوانی دارند. صرف نظر از اینکه در این استدلال، دور و تسلیل وجود دارد، بی‌بردن به خطای آن دشوار نیست.

داده‌های خود می‌پردازند. متأسفانه در دست داشتن معیار تک عددی برای رتبه دادن به دانشمندان بسیار سوشه کننده است زیرا چنین معیاری را افراد بسیار بیشتری درک می‌کنند، افرادی که نحوه کاربرد درست استدلال آماری را در وضعیت‌های به مراتب ساده‌تر هم درک نمی‌کنند.

## معنای ارجاع

کسانی که از شاخص‌های ارجاع به عنوان برترین معیارهای کیفیت پژوهش طرفداری می‌کنند پاسخی به این پرسش اساسی نمی‌دهند: معنای ارجاع چیست؟ آنها مقدار زیادی داده درباره تعداد ارجاعات گردآوری می‌کنند، داده‌ها را پژوهش می‌کنند تا آماره‌هایی [شاخص‌هایی آماری] از آنها استخراج کنند، و سپس مدعی می‌شوند که ارزیابی حاصل، «عینی» است. ولی تعبیر این آماره‌های است که به ارزیابی می‌انجامد، و این تعبیر به معنای ارجاع وابسته است، که کاملاً «ذهنی» است.

در مطلبی که طرفداران این آماره‌ها نوشتند، مشکل بتوان توضیح روشنی درباره معنای ارجاع یافت. از جمله، گفته‌اند: «مفهوم نهفته در پس نمایه‌سازی ارجاعات خیلی ساده است. اگر قبول کنیم که ارزش دستاوردها را کسانی تعیین می‌کنند که آن را به کار می‌برند، برای سنجش کیفیت کار پژوهشی چه راهی بهتر از اینکه تأثیر آن را بر کل جامعه علمی موردنظر سنجیم؟ ... وسیع‌ترین جامعه ممکن از میان کل جامعه علمی (یعنی هر کسی که مطلب مورد نظر را به کار می‌برد یا از آن به عنوان منبع نقل می‌کند) تعیین‌کننده تأثیر ایده موردنظر و صاحب آن ایده بر مجموعه معرفت ماست» یا آنکه گفته‌اند: «هرچند کمی سازی کیفیت کار دانشمندان دشوار است، نظر کلی این است که زیاد چاپ کردن بهتر از کم چاپ کردن است و تعداد استنادها به یک مقاله (با توجه به عرف رایج در رشته موردنظر) شاخص مفیدی از کیفیت آن مقاله است؛ و یا «بسامد ارجاعات نشان‌دهنده ارزش یک مجله و میزان استفاده از آن است» و یا «وقتی یک پژوهش یا محقق زیست‌پژوهشی به مقاله ارجاع می‌دهد معنی آن این است که آن مجله او را به نحوی تحت تأثیر قرار داده است» و یا «[ارجاع دادن نوعی ادای دین علمی است]. اصطلاحات مطرح اساسی در این اظهارات، «کیفیت»، «ارزش»، و «تأثیر» است و در میان آنها، واژه «تأثیر» واژه‌ای عام است که برای انتساب معنا به ارجاع به کار می‌رود. این کلمه را اول بار یوجین گارفیلد در سال ۱۹۵۵ در مقاله کوتاهی به کار برد که برای تبلیغ ایده شاخص ارجاع نوشت: «بنابراین، در مورد مقالات بسیار مهم، شاخص ارجاع ارزش کمی دارد زیرا به تاریخدان کمک می‌کند که میزان نفوذ یک مقاله یعنی «ضریب تأثیر» آن را سنجد.»

واضح است که منظور از «ضریب تأثیر» این است که مقاله ارجاع‌دهنده براساس مقاله ارجاع شده «پنا شده است» — یعنی ارجاع دادن به میثا سازوکاری که نحوه پیشرفت علم را نشان می‌دهد. از مطالب زیادی که درباره معنای واقعی ارجاع نوشته شده است چنین بر می‌آید که دلایل ارجاع دادن پیچیده‌تر و متنوع‌تر از این گونه اظهارات مبهم

## کاربرد عاقلانه شاخص‌ها

تحلیل و ارائه نتایج آماری. چنانکه قبل‌گفتیم، در بیشتر مواردی که شاخص‌های ارجاع برای رتبه‌بندی مقاله‌ها، پژوهشگران، و برنامه‌ها به کار می‌روند، هیچ مدل مشخصی پیش‌پیش در دست نیست بلکه مدل، که اغلب مبهم است، از خود داده‌ها برمی‌آید. به نظر می‌رسد یک فرایند دوری وجود دارد که چیزهای موردنظر رتبه بالا می‌گیرند چون (در پایگاه داده‌ها) بالارتبه‌اند. غالباً توجه چندانی به عدم قطعیت این رتبه‌بندی‌ها نمی‌شود و تحلیل چندانی از اثر این عدم قطعیت (مثلًا، تغییرات سالانه ضریب تأثیر) بر رتبه‌بندی به عمل نمی‌آید. و بالاخره، عوامل پیچیده (مانند رشته خاص، نوع مقالاتی که یک مجله انتشار می‌دهد، اینکه یک پژوهشگر خاص، تجربی کار است یا نظری کار) اغلب در این گونه رتبه‌بندی‌ها — به خصوص هنگامی که در ارزیابی‌های ملی انجام می‌شوند، — نادیده گرفته می‌شود.

تعییر و تأثیر، تأثیر ارزیابی پژوهش در جامعه علمی آشکار است. برای خود دانشمند، این ارزیابی ممکن است تأثیر عمیق و درازمدتی بر زندگی حرفه‌ای اش داشته باشد؛ برای یک دانشکده یا پژوهشکده، ممکن است چشم‌انداز موفقیت در آینده را دگرگون کند، برای یک رشته علمی، مجموعه‌ای از این ارزیابی‌ها ممکن است باعث رونق و شکوفایی یا آفت و نزول شود. در کاری با این اهمیت، باید هم اعتبار و هم محدودیت‌های ابزارهای مورد استفاده را شناخت. تعداد ارجاع‌ها تا چه حدی کیفیت پژوهش را می‌سنجد؟ به نظر می‌رسد این تعداد با کیفیت همبستگی داشته باشد و بنا به دریافت شهودی، مقاله‌های با کیفیت، پراجاع‌اند. ولی همان‌طور که در بالا توضیح داده شد، برخی مقالات، به خصوص در بعضی از رشته‌ها، به دلایلی به جز کیفیت از ارجاعات زیاد برخوردار می‌شوند، و لزوماً چنین نیست که مقاله‌های پراجاع خوش‌کیفیت باشند. تعییر دقیق رتبه‌بندی‌های مبتنی بر ارجاعات باید روشن شود. به علاوه، اگر شاخص‌های ارجاع نقش اساسی در ارزیابی پژوهش داشته باشند، روشن است که نویسنده‌گان، ویراستاران، و حتی ناشران مقاله‌ها راه‌هایی برای دستکاری در نظام به نفع خودشان خواهند یافتد.

به عنوان حسن ختام، جملات پایانی مقاله‌ای را ذکر می‌کنیم که درباره مسائل آماری در مقایسه عملکرد انسیتوها نوشته شده است و گویای آرزو و هدف همه‌کسانی است که دستی در ارزیابی پژوهش دارند: «و بالاخره، هرچند ما از بسیاری از شیوه‌های داوری درباره انسیتوها انتقاد کرده‌ایم، نمی‌خواهیم چنین برداشت شود که به عقیده‌ما همه‌این روش‌های مقایسه‌لزوماً معیوب است. به نظر ما، مقایسه انسیتوها و تلاش برای فهم دلایل تفاوت آنها فوق العاده مهم است و این کار در صورتی به بهترین وجه انجام می‌شود که روحیه همکاری حاکم باشد نه تقابل و رویارویی. این شاید تنها راه مطمئن برای کسب اطلاعات عینی باشد که ممکن است در نهایت به پیشرفت و اصلاح بینجامد. مشکل واقعی در روش‌های ساده‌انگارانه‌ای که ما از آنها انتقاد کردیم این است که هم توجه و هم منابع و امکانات ما را از این هدف بالارزش‌تر دور می‌سازند.»

اتکای بیش از حد به شاخص‌های عینی منحصر به ارزیابی پژوهش نیست و تازگی هم ندارد. جامعه‌شناسی به نام جوئل بست (Joel Best) نوشته است: «در بعضی از فرهنگ‌ها، مردم اعتقاد به قدرت سحرآمیز بعضی چیزها دارند. در جامعه‌ما، آمار چنین چیزی است. سنجه‌های آماری را موجوداتی ماورای اعداد در نظر می‌گیریم که گویی عصارة واقعیات پیچیده و گیج‌کننده را به صورت داده‌ای ساده عرصه می‌کنند. از آمار برای تبدیل مسائل اجتماعی بغرنج به براوردها، درصدها، و نیزهایی که فهم آنها آسان‌تر است استفاده می‌کنیم؛ و اینها هستند که تعیین می‌کنند نسبت به چیزی، و چقدر، باید دغدغه داشته باشیم. مسئله اجتماعی، به مفهومی، به شاخص‌های آماری تقسیل می‌بایند و دیدگاه ما به مسائل اجتماعی را زیرسلطه خود می‌گیرند. آمار و ارقام رانه به عنوان اعدادی که خلقشان کرده‌ایم بلکه به عنوان واقعیاتی که کشفشان کرده‌ایم، در نظر می‌گیریم.»

این گونه اعتقاد به سحرآمیز بودن شاخص‌های آماری در هنگام ارزیابی پژوهش — چه در سطح ملی و چه در سطح مؤسسات — مستجملی است، و در اقدامات جدیدی هم که برای بهبود ضریب تأثیر با استفاده از الگوهای ریاضی پیچیده، از قبیل الگوریتم‌های رتبه‌صفحه‌ای، به عمل می‌آید مشهود است. طرفداران این تدابیر جدید ادعاهایی درباره کارایی شاخص‌های جدید مطرح می‌کنند که توجیه و ارزیابی آنها دشوار است. چون این شاخص‌ها مبتنی بر محاسبات پیچیده‌اند، تشخیص فرض‌هایی (اغلب پنهان) که در پس آنها نهفته است برای بیشتر افراد آسان نیست. ما با اعداد و رتبه‌بندی‌ها با خوف و حمیت رفتار می‌کنیم زیرا آنها را واقعیات مسلم و نه مخلوقات خودمان می‌پنداrim.

پژوهش تنها حوزه از حوزه‌های فعالیت عمومی نیست که مورد ارزیابی قرار می‌گیرد. در چند دهه گذشته، سعی کرده‌اند که عملکردها را تقریباً در همه حوزه‌ها مورد ارزیابی کشی قرار دهند، از نظام‌های آموزشی تا بهداشت و غیره. در مواردی که آماردانان در انتخاب سنجه‌های معقول و استفاده صحیح از آمار دخالت داشته‌اند، نتایج نیکویی به بار آمده است. همان‌طور که برای درمان بیماری به پژوهش مراجعه می‌کنیم، هنگام استفاده از آمار هم باید به سراغ آماردانان رفت (و به نصایح آنها گوش داد).

داده‌ها، مناسب بودن و انسجام و کفایت داده‌ها از ملزمات هر بررسی آماری است و هیچ ترفند آماری نمی‌تواند بر نارسایی‌های بنیادی داده‌ها فائق آید. این نکته در ارزیابی‌هایی مبتنی بر تعداد ارجاعات نیز مهم است حال آنکه مثلاً ضریب تأثیر براساس زیرمجموعه‌ای از داده‌ها محاسبه می‌شود زیرا فقط مجله‌هایی در نظر گرفته می‌شوند که مؤسسه‌تامسون آنها را انتخاب کرده است (و ضریب تأثیر نیز خود یکی از ملاک‌های مهم در این انتخاب است).