



## About Highly Cited Researchers

*Highly Cited Researchers 2014*

[Explanation of the Method and Purpose](#)

[Methodology Details](#)

[Frequently Asked Questions](#)

[User's Guide](#)

### EXPLANATION OF THE METHOD AND PURPOSE OF THOMSON REUTERS NEW LIST OF HIGHLY CITED RESEARCHERS 2014

Thomson Reuters has generated a new list of Highly Cited Researchers in the sciences and social sciences to update and complement a previously published list that was presented on the website [ISHighlyCited.com](http://ISHighlyCited.com).

The old list, first issued in 2001, identified more than 7,000 researchers who were the most cited in one or more of 21 broad fields of the sciences and social sciences, fields similar to those used in the Essential Sciences Indicators database. This analysis considered articles and reviews published in Web of Science-indexed journals from 1981 through 1999. Approximately 250 researchers in each field were selected based on total citations to their papers published during this period. An update in 2004 took into account papers published from 1984 to 2003 and cited during the same period, and additional names were added to supplement the original list.

A selection of influential researchers based on total citations gives preference to well-established scientists and social sciences researchers who have produced many publications. It is only logical that the more papers generated, generally the more citations received, especially if the papers have had many years to accumulate citations. Thus, this method of selection favors senior researchers with extensive publication records. It sometimes identifies authors who may, in fact, have relatively few individual papers cited at high frequency. Nonetheless, total citations is a measure of gross influence that often correlates well with community perceptions of research leaders within a field. Such was the nature of the prior lists of highly cited researchers.

Thomson Reuters decided to take a different approach -- and use a different method -- to identify influential researchers, field-by-field, to update the previously published list. First, to focus on more contemporary research achievement, only articles and reviews in science and social sciences journals indexed in the Web of Science Core Collection during the 11-year period 2002-2012 were surveyed. Second, rather than using total citations as a measure of influence or 'impact,' only Highly Cited Papers were considered. Highly Cited Papers are defined as those that rank in the top 1% by citations for field and year indexed in the Web of Science, which is generally but not always year of publication. These data derive from Essential Science Indicators<sup>SM</sup> (ESI). The fields are also those employed in ESI -- 21 broad fields defined by sets of journals and exceptionally, in the case of multidisciplinary journals such as Nature and Science, by a paper-by-paper assignment to a field. This percentile-based selection method removes the citation

Those researchers who, within an ESI-defined field, published Highly Cited Papers were judged to be influential, so the production of multiple top 1% papers was interpreted as a mark of exceptional impact. Relatively younger researchers are more apt to emerge in such an analysis than in one dependent on total citations over many years. To be able to recognize early and mid-career as well as senior researchers was one goal for generating the new list. The determination of how many researchers to include in the list for each field was based on the population of each field, as represented by the number of author names appearing on all Highly Cited Papers in that field, 2002-2012. The ESI fields vary greatly in size, with Clinical Medicine being the largest and Space Science (Astronomy and Astrophysics) the smallest. The square root of the number of author names indicated how many individuals should be selected.

The first criterion for selection was that the researcher needed enough citations to his or her Highly Cited Papers to rank in the top 1% by total citations in the ESI field in which they were considered. Authors of Highly Cited Papers who met the first criterion in a field were ranked by number of such papers, and the threshold for inclusion was determined using the number derived through calculation of the square root of the population. All who published Highly Cited Papers at the threshold level were admitted to the list, even if the final list then exceeded the number given by the square root calculation. In addition, and as concession to the somewhat arbitrary cut-off, any researcher with one fewer Highly Cited Paper than the threshold number was also admitted to the list if total citations to his or her Highly Cited Papers were sufficient to rank that individual in the top 50% by total citations of those at the threshold level or higher. The justification for this adjustment at the margin is, it seemed to work well in identifying influential researchers, in the judgment of Thomson Reuters citation analysts.

Of course, there are many highly accomplished and influential researchers who are not recognized by the method described above and whose names do not appear in the new list. This outcome would hold no matter what specific method was chosen for selection. Each measure or set of indicators, whether total citations, h-index, relative citation impact, mean percentile score, etc., accentuates different types of performance and achievement. Here we arrive at what many expect from such lists but what is really unobtainable: that there is some optimal or ultimate method of measuring performance. The only reasonable approach to interpreting a list of top researchers such as ours is to fully understand the method behind the data and results, and why the method was used. With that knowledge, in the end, the results may be judged by users as relevant or irrelevant to their needs or interests.

---

## DETAILS OF METHOD USED TO SELECT HIGHLY CITED RESEARCHERS 2014 IN 21 FIELDS OF THE SCIENCES AND SOCIAL SCIENCES

The data used in the analysis and selection of the new highly cited researchers came from Essential Science Indicators, 2002-2012, which then included 113,092 Highly Cited Papers. Each of these papers ranked in the top 1% by total citations according to their ESI field assignment and year of publication (as stated above, more correctly, year processed for the Web of Science). For more information on the identification of Highly Cited Papers in ESI, see the ESI help file at [http://esi.webofknowledge.com/help/h\\_dathic.htm](http://esi.webofknowledge.com/help/h_dathic.htm)

ESI surveys the SCI-E and SSCI components of the Web of Science, meaning journal articles in the sciences and social sciences. The analysis is further limited to items indexed as articles or reviews only, and does not include letters to the editor, correction notices, and other marginalia.

In ESI, all papers, including Highly Cited Papers, are assigned to one of 22 broad fields (the 22nd is Multidisciplinary, on which see below). Each journal in ESI is assigned to only one field and papers appearing in that title are similarly assigned. In the case of multidisciplinary journals such as Science, Nature, Proceedings of the National Academy of Sciences of the USA, and others, however, a special analysis is undertaken. Each article in such publications is individually reviewed, including an examination of the journals cited in its references as well as the journals from which citations to it derive. The 'weight' of these cited and citing relationships helps to assign papers to a specific ESI field. For more information about this reclassification process, see our article at <http://archive.sciencewatch.com/about/met/classpapmultijour/>. In fact, this procedure includes not only those journals that are widely recognized as multidisciplinary, but others, too, that within Clinical Medicine, publish papers across a wide number of specialty areas (these include: Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Society, The Lancet, New England Journal of Medicine, Journal of Clinical Investigation, Cell, Journal of Experimental Medicine, and Nature Medicine). Some few papers (> 5% in the case of Science, Nature, and PNAS) are unable to be assigned to a specific field and remain in the Multidisciplinary category in ESI. As such, Multidisciplinary in ESI should be recognized as a grouping of 'leftovers' and, therefore, not very useful for performance analysis.

A ranking of author names in each ESI category by number of Highly Cited Papers produced during 2002-2012 was the first step in identification and selection of our new list of highly cited researchers. We used algorithmic analysis to help distinguish between individuals with the same name or name form (surname and initials). In instances where any ambiguity remained, manual inspection was needed. This entailed searching for papers by author surname and one or multiple initials, ordering them chronologically, visually inspecting each (noting journal of publication, research topic or theme, institutional addresses, co-authorships, and other attributes), and deciding which ones could be attributed to a specific individual. As noted in the FAQ section, we examined original papers, if necessary, as well the websites of researchers themselves and their curricula vitae. This was often required if a researcher changed institutional affiliations several times during the period surveyed.

Once the data on Highly Cited Papers within an ESI field were verified and assigned to specific individuals, the authors in the field were ranked by number of Highly Cited Papers. To determine how many researchers to select for inclusion in the new list, we considered the size of each ESI field in terms of number of authors (as a proxy for population) represented on the Highly Cited Papers for the field. The ESI fields are of very different sizes, the result of the definition used for the field which includes the number of journals assigned to that field. Clinical Medicine, for example, makes up some 19% of the content of ESI while Economics and Business, Microbiology, and Space Science (Astronomy and Astrophysics) account for 1.7%, 1.4%, and 1.1%, respectively. For each ESI field, author names (before use of the PDE algorithm and therefore not disambiguated) were counted, and then the square route of that number was calculated. That number was used to decide approximately how many researchers to include in each ESI field. From the list of authors in a field ranked by number of Highly Cited Papers, the number of papers at the rank represented by the square root score determined the threshold number of Highly Cited Papers required for inclusion. If an author had one fewer highly cited paper than this threshold, but whose citations to their Highly Cited Papers were sufficient to rank them in the top 50% by citations among those with Highly Cited

Papers at or above the threshold, these individuals were also selected. Finally, citations to an individual's Highly Cited Papers had to meet or exceed the threshold for total citations used in the 2002-2012 version of ESI for including a researcher in the top 1% (highly cited list) for an ESI field.

The methodology described above was applied to all ESI fields with the exception of Physics. The relative large number of Highly Cited Papers in Physics dealing with high-energy experiments typically carried hundreds of author names. Using the whole counting method produced a list of high-energy physicists only and excluded those working in other subfields. For example, the number of Highly Cited Papers required for inclusion in Physics, using the standard methodology, turned out to be a remarkable 63. So, as an expedient, it was decided to eliminate from consideration any paper with more than 30 institutional addresses. This removed 436 out of 10,373 Highly Cited Papers in physics and the problem of overweighting to high-energy physics. An analysis without these papers produced a list in which the threshold for number of Highly Cited Papers was 14. It also produced a ranking in which the 2010 Nobel Prize winner in Physics Andre Geim of the University of Manchester appeared first, with 40 Highly Cited Papers. Fields of physics other than high-energy alone now appear as represented by the scientists selected.

The final new list contains some 3,200 highly cited researchers in 21 fields of the sciences and social sciences.

---

**4. I have been named a Highly Cited author in Engineering but my field and departmental affiliation is actually Mathematics. Would you change my designation to Mathematics?**

We understand that you identify yourself as a mathematician, but we found your greatest impact, according to our analysis, to be in Engineering as it is defined in ESI. There is no universally agreed field classification scheme, and the use of journals to define fields is approximate at best. The practical advantage of our method is that we can fairly compare individuals against one another in the same consistently defined sphere.

**5. I have a very common name, and some other people with the same name form (surname and initials) actually work in the same field that I do. How did you make sure not to confuse me and my papers with others and their papers in your analysis?**

Ensuring correct attribution of papers to authors involved a manual inspection of each Highly Cited Paper. The Highly Cited Papers in an ESI field for a specific name and its variants (such as, for example, surname plus one initial or two) were ordered in chronological sequence, the subject of the papers examined as well as the journals in which they were published, the institutional addresses reviewed, and the co-authorships inspected. Often this was sufficient to resolve questions of authorship for a unique individual. Original papers were sometimes consulted to obtain a full name not present in the Web of Science bibliographic record (papers indexed before 2006). Reference was made to websites of researchers themselves and their curricula vitae if questions remained, which sometimes arose when a researcher changed institutional affiliations several times during the period surveyed. We would like to think our efforts to resolve authorship questions resulted in 100% clean data, but with any such effort, and more than 3,000 researchers, we likely fell short in some few specific instances and will make adjustments where required.

**6. You say you selected top researchers according to specific fields as defined in ESI. But what about researchers who have Highly Cited Papers across several fields, such as Molecular Biology and Genetics, Clinical Medicine, and Psychiatry/Psychology? How did you account for such cross-disciplinary impact?**

We analyzed each ESI field separately. If we had attempted to identify researchers across fields we would have faced the problem of determining a cross-field baseline for inclusion in an all-field list. In other words, the percentage of Highly Cited Papers required for inclusion differs across fields. In some fields, the number of Highly Cited Papers required for selection was low compared to others – as much as five to one – and each individual would present a different mix of fields in which they published their Highly Cited Papers and in different quantities, thereby creating an unmanageable number of combinations by which to calculate a baseline for inclusion.

**7. Did you apportion credit for Highly Cited Papers according to the number of authors on a paper? You know, in some fields and especially with some highly cited reports (high-energy physics, for example), papers reflect the work of large teams.**

You raise the issue of whole vs. fractional counting, which has been likened to measuring participation versus contribution, respectively. The question is a good one but the solution to apportioning credit is not straightforward, and it highlights deep questions about the contemporary meaning of authorship. First, we should recognize that scientists themselves do not explicitly apportion credit on their papers, although there has been discussion about making individual contributions more transparent on papers. The norm – and appearance – is that each author contributed significantly to the publication. How much is left opaque. Let us postulate that fractional credit is better than full credit for each author, i.e., whole counting. In this case a paper with five authors would yield .2 of a paper for each. But, in the absence of specific information on contribution, is not fractional counting just as arbitrary a choice as whole counting? There is more. What does one then do with the citation counts to the paper with respect to each author? Should the total citation count, too, be divided proportionally among the authors? This is an even more perplexing question, calling into question whether contribution may be partial but influence whole. There seems to be a qualitative difference between credit for contribution and credit for impact. Finally, what would appear to be fairer in terms of assigning credit using fractional counting may in fact bias an analysis against papers resulting from teamwork, since some fields, such as mathematics, typically exhibit a very low average number of authors in comparison to, say, biomedical fields. Fractionating credit may undercount contributions from team members and overcount contributions in fields that typically show many single-author publications.

Despite discounting the superiority of fractional counting of publications over whole counting above, we must here comment on an exceptional procedure used in the treatment of Highly Cited Papers in Physics and the selection of Highly Cited Researchers in Physics, 2002-2012. The rise of papers with hundreds and even thousands of authors is a fairly recent phenomenon, one that reveals important changes in the scale of collaboration required in certain realms (see: [http://archive.sciencewatch.com/newsletter/2012/201207/multiauthor\\_papers/](http://archive.sciencewatch.com/newsletter/2012/201207/multiauthor_papers/)). High-energy physics is one such area, and others that may be cited are large-scale clinical trials, genome mapping studies, and astronomy projects and related space-based deployments of new instrumentation. But the phenomenon is particularly acute in high-energy physics and evident on many Highly Cited Papers of recent vintage, such as those of the CMS and ATLAS teams at CERN. A preliminary analysis of Physics, using the methodology described above, resulted in a list that included high-energy physicists only, most of whom achieved their ranking through vanishingly small fractional credits. This outcome, and its effect of suppressing all else in physics, required a different approach. In the case of Physics – and Physics only – we removed from our analysis any paper with 30 or more institutional addresses ( $n = 436$  out of 10,373), and this eliminated the problem of dealing with high-energy physics papers exhibiting 'a cast of thousands.' The use of 30 as a threshold for institutional addresses was determined heuristically through examination of all Highly Cited Papers in physics, their content, and the desire to include reports deriving from extensive multi-institute collaborations but not those from huge teams, as found in many high-energy physics experiments. The result was a more balanced view of high-impact researchers and topics in Physics. High-energy physicists, especially theoreticians, nonetheless do appear on the new list. This 'Gordian-knot' solution will not please some, but it solved the problem we faced in a practical way.

**8. I believe I have a method that produces a result more consistent with the scientific community's perception of top researchers in a field. Would you take into account my feedback?**