Journal of Informetrics 5 (2011) 346–359

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

# A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants

Lutz Bornmann[a,*], Rüdiger Mutz[b], Sven E. Hug[b], Hans-Dieter Daniel[b,c]

[a] *Max Planck Society, Hofgartenstr. 8, D-80539 Munich, Germany*
[b] *ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr. 24, CH-8092 Zurich, Switzerland*
[c] *University of Zurich, Evaluation Office, Mühlegasse 21, CH-8001 Zurich, Switzerland*

ARTICLE INFO

*Article history:*
Received 28 September 2010
Received in revised form 7 January 2011
Accepted 17 January 2011

*Keywords:*
*h* index
*h* index variants
Meta-analysis
Multilevel analysis

ABSTRACT

This paper presents the first meta-analysis of studies that computed correlations between the *h* index and variants of the *h* index (such as the *g* index; in total 37 different variants) that have been proposed and discussed in the literature. A high correlation between the *h* index and its variants would indicate that the *h* index variants hardly provide added information to the *h* index. This meta-analysis included 135 correlation coefficients from 32 studies. The studies were based on a total sample size of $N = 9005$; on average, each study had a sample size of $n = 257$. The results of a three-level cross-classified mixed-effects meta-analysis show a high correlation between the *h* index and its variants: Depending on the model, the mean correlation coefficient varies between .8 and .9. This means that there is redundancy between most of the *h* index variants and the *h* index. There is a statistically significant study-to-study variation of the correlation coefficients in the information they yield. The lowest correlation coefficients with the *h* index are found for the *h* index variants MII and *m* index. Hence, these *h* index variants make a non-redundant contribution to the *h* index.



## 1. Introduction

Hirsch's *h* index was proposed as a better alternative to other bibliometric indicators (such as number of publications, average number of citations, and sum of all citations) (Hirsch, 2007). It is based on a scientist's lifetime citedness, which incorporates productivity as well as citation impact: "A scientist has index *h* if *h* of his or her $N_p$ papers have at least *h* citations each and the other $(N_p - h)$ papers have $\leq h$ citations each" (Hirsch, 2005, p. 16569). Nowadays, the *h* index is no longer being used as a measure of scientific achievement only for single researchers (Glänzel, 2006). The index is also being used to measure the scientific output of research groups (van Raan, 2006), scientific facilities (Kinney, 2007), and countries (Csajbók, Berhidi, Vasas, & Schubert, 2007). The indices at the aggregated level (group, scientific facility, country) are calculated analogously to the calculation for single researchers. When single researchers are grouped in research groups, scientific facilities, or countries, in addition to calculation of *h* index values at the higher aggregate level it is also possible to calculate successive *h* index values (Hu, Rousseau, & Chen, 2010; Prathap, 2006; Rousseau, Yang, & Yue, 2010): "The institute has an index $h_2$ if $h_2$ of its *N* researchers have an $h_1$ index of at least $h_2$ each, and the other $(N - h_2)$ researchers have $h_1$ indices lower than $h_2$ each. The succession can then be continued, e.g., for networks of institutions or countries or other higher levels of aggregation" (Schubert, 2007, pp. 201–202).

* Corresponding author. Tel.: +49 089 2108 1265.
*E-mail address:* bornmann@gv.mpg.de (L. Bornmann).

1751-1577/$ – see front matter 
doi:10.1016/j.joi.2011.01.006

Today the *h* index is a widely accepted indicator of scientific performance and "has even become a built-in feature of major bibliographic databases such as Web of Science [Thomson Reuters, Philadelphia, PA, USA] and Scopus [Elsevier, Amsterdam, The Netherlands]" (van Eck & Waltman, 2008, p. 263). The wide acceptance is due mainly to the many advantages of the new index. These include "its simplicity and immediate intuitive meaning" (Franceschini & Maisano, 2010, p. 495), the combination of citation impact and publication numbers in one number (British Academy, 2007), and its robustness against small errors and single peaks (top-cited papers) in a publication list (Liu & Rousseau, 2009). However, in the research on the *h* index (see overviews in Bornmann & Daniel, 2007, 2009; Egghe, 2010; Norris & Oppenheim, 2010; Panaretos & Malesios, 2009; Thompson, Callen, & Nahata, 2009) a number of disadvantages have been pointed out that are either specific to the *h* index or that are shared by the *h* index and other citation-based indicators: It is field-dependent, it may be influenced by self-citations, it does not consider multi-authorship, it is dependent on the scientific age of a scientist, it can never decrease, and it is only weakly sensitive to the highly cited papers.

The disadvantages discussed in connection with the *h* index have led to the development of a now large number of *h* index variants. These are new indices based on the *h* index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009) that are supposed not to have one or more disadvantages. For example, the *g* index proposed by Egghe (2006b) places more weight on the citation performance of a set of papers than the *h* index does. The aim of this study is to determine empirically the extent to which the development of the *h* index variants does in fact result in additional information not provided by the *h* index. Although the proposed variants may be conceptualized differently than the *h* index theoretically or mathematically, in their empirical application they may be highly correlated with the *h* index. "Perfect correlation . . . does not imply that two variables are identical but rather that one of them, *Y*, say, can be written as a linear function of the other, $Y = a + bX$, where *b* is the slope of the regression line and a is the intercept" (Good & Hardin, 2009, p. 198). For Burrell (2009), for instance, each of the *h* index variants "seems to be (approximately) proportional to time and hence each is (approximately) proportional to each of the others" (p. 419).

In this study we present the first meta-analysis of studies that computed correlations between the *h* index and variants of the *h* index that have been proposed and discussed in the literature (i.e., the paper is not about the added value of the *h* index or *h* index variants to other bibliometric indices, e.g., total number of citations). We test whether there is a generally high (or low) correlation between the *h* index and the variants. Additionally, we examine how large the study-to-study variation of the reported correlation coefficients is: Are there *h* index variants that make a non-redundant contribution to the *h* index (that is, that have a low correlation with the *h* index)? Finally, we test the extent to which the study-to-study variation can be explained by covariates. What are the moderators of high or low correlation coefficients?

The term meta-analysis was coined by Glass (1976) to refer to "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3). This quantitative technique has many advantages: Firstly, it allows generalized statements on the strength of the effects (here: correlations), regardless of the specificity of individual studies (Matt & Navarro, 1997). Secondly, it presents findings of the original studies in a manner that are more sophisticated than the usual literature reviews that heavily rely on qualitative summarizing with no respect, for instance, to sample sizes of the primary studies. Thirdly, it is capable of revealing relationships which are obscured in traditional summarizing reviews. Fourthly, it provides a systematic way of getting information from a large number of study findings (Lipsey & Wilson, 2001). Fifthly, it is a widely accepted method to systematically summarize information of primary studies, especially in social sciences and medicine ("evidence based medicine").

The meta-analysis presented here included a total of 135 different bivariate correlation coefficients between the *h* index and an *h* index variant (such as the *g* index; in total 37 different variants). Dependencies in the data, such as (1) several coefficients were reported within one study, and (2) the correlation between the *h* index and a certain *h* index variant was computed in more than one study, made it necessary to perform a multilevel meta-analysis.

## 2. Methods

### *2.1. Literature research*

The literature research was conducted at the beginning of 2010. We performed a systematic search of publications of all document types (journal articles, monographs, collected works, etc.). In a first step, we used tables of contents of issues of information science journals (e.g., *Journal of Informetrics*, *Scientometrics*) and reference lists provided by narrative reviews of research on the *h* index (e.g., Egghe, 2010) to locate several studies that investigated the correlation between the *h* index and its variants. In a second step, to obtain keywords for searching computerized databases; we prepared a bibliogram (White, 2005) for the studies located in the first step. The bibliogram ranks by frequency the words included in the abstracts of the studies located. Words at the top of the ranking list (e.g., *h* index; Hirsch index; and g index) were used for searches in computerized literature databases (e.g., Web of Science; Scopus) and Internet search engines (e.g., Google). In a third step of our literature search, we located all of the citing publications for a series of articles (found in the first and second steps) for which there are a fairly large number of citations in Web of Science.

The search for publications identified 199 studies published between 2005 and 2010 as article ($n = 172$), grey literature ($n = 15$), conference proceeding ($n = 11$), or book section ($n = 1$). Thirty-five out of the 199 publications reported all information required for the meta-analysis presented here: (1) at least one correlation coefficient between the *h* index and an *h* index variant, and (2) the sample size of the study (Antonakis & Lalive, 2008; Arencibia-Jorge & Rousseau, 2009; Bornmann, Marx, &

Schier, 2009; Bornmann, Mutz, & Daniel, 2008; Bornmann, Mutz, Daniel, Wallon, & Ledin, 2009; Cabrerizo, Alonso, Herrera-Viedma, & Herrera, 2010; Costas & Bordons, 2008; de Visscher, 2010; Franceschet, 2009; García-Pérez, 2009; Harzing & van der Wal, 2008; Haslam & Laham, 2010; Hu et al., 2010; Hua, Wan, & Wu, 2010; Jin, Liang, Rousseau, & Egghe, 2007; Kosmulski, 2006; Lee, Kraus, & Couldwell, 2009; Liu & Rousseau, 2007, 2009; Lovegrove & Johnson, 2008; Mingers, 2009; Moussa & Touzani, 2010; Opthof & Wilde, 2009; Rousseau et al., 2010; Ruane & Tol, 2007; Sanderson, 2008; Schreiber, 2008a, 2009a, 2009b; Schubert, Korn, & Telcs, 2009; Sypsa & Hatzakis, 2009; Tol, 2009; Vinkler, 2009; Wohlin, 2009; Wu, 2010). Since three papers by Schreiber (2008a, 2009a, 2009b) and two papers by Liu and Rousseau (2007, 2009) refer to one and the same dataset, the number of studies that could be included in the meta-analysis decreased from 35 to 32. In total, these 32 studies are based on a sample size of 9005 units (e.g., scientists, for whom an *h* index was computed); on average, each study has a sample size of 257 units. Due to the fact that most of these studies were published as peer reviewed publications (89%), their quality is sufficiently guaranteed.

As most of the 32 studies reported more than one correlation (in many studies, a correlation between the *h* index and several variants of the *h* index was computed), a total of 135 correlation coefficients (on average 4 coefficients per study) were gathered. In total, in the 32 studies the *h* index was correlated with 37 different *h* index variants, which are described in the section below. Nearly all studies provided the following quantitative correlation coefficients: Pearson product-moment correlation ($n = 35$) and Spearman's rank correlation ($n = 86$). For a total of 14 coefficients, the authors of the papers did not state what kind of correlation was computed. If different coefficients were reported for the same sample in one single study, Spearman's rank correlations were included in the meta-analyses.

One researcher on our team extracted the information from the publications that was needed for the meta-analysis. The information was then validated by a second researcher on our team, and disagreements were resolved by consensus.

### 2.2. *h index variants included in the meta-analysis*

Table 1 shows in alphabetical order the 37 *h* index variants that were included in the meta-analysis. The *h* index variants are categorized by the advantages proposed by the inventors/authors of the respective *h* index variant themselves. As a categorization scheme the following five advantages of the *h* index variants over the *h* index identified by Egghe (2010) were used:

(1) *The h index variant takes field dependence into account*, that is, the variant tries to compensate for different citation rates across fields. In this meta-analysis, one of the rare *h* index variants of this type is the $h_I$ index (Batista, Campiteli, & Kinouchi, 2006). Batista et al. (2006) assume that co-authorship behavior is characteristic of each research field, and they therefore propose the $h_I$ index, which "indicates the number of papers a researcher would have written along his/her carrier [*sic*] with at least $h_I$ citations if he/she has worked alone" (Batista et al., 2006, p. 188). $h_I$ is the result of the quotient $h^2/N_a^{(T)}$, where $N_a^{(T)}$ is the total number of authors in the considered *h* papers. Batista et al. (2006) claim that by this means it is possible to quantify an individual's scientific research output that is valid across different fields.
(2) *The h index variant takes self-citations into account*. As Schreiber (2007a, 2007b) showed, self-citations can inflate the *h* index substantially. To solve this issue Schreiber proposed the sharpened index $h_s$, "which is defined as the highest number of papers that have received $h_s$ or more independent citations" (Schreiber, 2009b, p. 208). In the present meta-analysis there are further *h* index variants included that correct also for self-citations – for example, Kosmulski's (2006) ch index or Schreiber's (2009b) $h_{ms}$ index.
(3) *The h index variant takes multi-authorship into account*. An example with this feature is the $h_I$ index already mentioned above. Schreiber (2008b) points out that a drawback of the $h_I$ index is the normalization procedure via the mean number of authors, because it can penalize scientists "with some papers with a large number of co-authors" (p. 211) and can reduce "the influence of single-author publications to one's *h*-index" (p. 211). To avoid these biases Schreiber (2008b) proposed the $h_m$ index, which can be obtained by counting citations fractionally according to the inverse of the number of authors. These fractional counts are then added up to reduced numbers of publications, which can be utilized as an "effective" rank to determine the $h_m$ index. There are several other *h* index variants included in the present meta-analysis that deal with the issue of multi-authorship, such as Opthof and Wilde's (2009) *h* index of first authored papers, which incorporates only first-authored papers of a scientist and discards all other publications by the author, or Hu et al.'s (2010) major contribution *h* index, *h* maj, which includes only first-authored papers and papers for which a scientist has been tagged as the corresponding author.
(4) *The h index variant takes career length or the age of publications into account*. In his seminal paper Hirsch (2005) already tackled the issue of the cumulative and non-decreasing nature of the *h* index, which favors older over younger scientists, by introducing the *m* quotient. The *m* quotient is defined as the ratio *h*/y, where *h* represents the *h* index and *y* represents the number of years passed since the first paper of a scientist was published. Most *h* type indices are roughly proportional to the square root of the number of citations, but the *m* quotient is not. Most *h* type indices increase in time (even when the scientist is not active), but the *m* quotient does not. Whereas the *m* quotient corrects for different career lengths, it does not adjust for different ages of publications. In this meta-analysis, one *h* index variant that deals with this latter issue is the contemporary *h* index. Sidiropoulos, Katsaros, and Manolopoulos (2007) define the contemporary *h* index as follows: "A researcher has contemporary *h*-index $h^c$, if $h^c$ of its $N_p$ articles get a score of $S^C(i) \geq h^C$ each, and the rest $(N_P - h^C)$ articles get a score of $S^C(i) \leq h^C$" (p. 258). $S^C(i)$ equals $\gamma^*(Y(\text{now}) - Y(i) + 1)^{-\delta*}|C(i)|$, where $\gamma$ is a constant

**Table 1**
List of *h* index variants included in the meta-analysis, categorized by proposed advantages over the *h* index (see here).

| Full name of *h* index variant | Short name of *h* index variant | Takes into account field dependence | Takes into account self-citations | Takes into account multi-authorship | Takes into account career length or age of the publications | Measures citation intensity in *h* core or gives more weight to highly cited papers | Author(s) |
|---|---|---|---|---|---|---|---|
| *A*-index | *A* index | | | | | X | Jin (2006) |
| *AR*-index | *AR* index | | | | X | X | Jin (2007) |
| ch-index | ch(2) index | | X | | | | Kosmulski (2006) |
| ch(2)-index | ch index | | X | | | X | Kosmulski (2006) |
| Contemporary *h*-index | cont *h* index | | | | X | | Sidiropoulos et al. (2007) |
| Degree *h*-index $h_A$ | degree hA | | | | | | Schubert et al. (2009) |
| Degree *h*-index $h_P$ | degree hP | | | | | | Schubert et al. (2009) |
| *f*-index | *f* index | | | | | X | Tol (2009) |
| *g*-index | *g* index | | | | | X | Egghe (2006a) |
| *h*(2)-index | *h*(2) index | | | | | X | Kosmulski (2006) |
| hg-index | hg index | | | | | X | Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2010) |
| $h_I$-index | hI index | X | | X | | | Batista et al. (2006) |
| *h* index of first authored papers | *h* first author | | | X | | | Opthof and Wilde (2009) |
| $h_m$-index | hm index | | | X | | | Schreiber (2008b) |
| $h_{mol}$-index | hmol index | | | | | | Molinari and Molinari (2008) |
| $h_{ms}$-index | hms index | | X | X | | | Schreiber (2009b) |
| $h_s$-index | hs index | | X | | | | Schreiber (2007b) |
| hw-index | hw index | | | | | X | Egghe and Rousseau (2008) |
| Index of quality and productivity | IQp | X | | | X | X | Antonakis and Lalive (2008) |
| major contribution *h*-index | *h* maj | | | X | | | Hu et al. (2010) |
| Maxprod | Maxprod | | | | | X | Kosmulski (2007) |
| *m*-index | *m* index | | | | | X | Bornmann et al. (2008) |
| Modified impact index | MII | | | | | | Sypsa and Hatzakis (2009) |
| *m*-quotient | *m* quotient | | | | X | | Hirsch (2005) |
| Multidimensional *h* index (second component; $h_2$) | multidim h2 | | | | | | Garcia-Perez (2009) |
| Multidimensional *h* index (third component; $h_3$) | multidim h3 | | | | | | Garcia-Perez (2009) |
| Pi-index | Pi index | | | | | X | Vinkler (2009) |
| Prathaps $h_2$ index | Prathap h2 | | | | | | Prathap (2006) |
| $q^2$ index | q(2) index | | | | | X | Cabrerizo et al. (2010) |
| Raw *h* rate | Raw *h* rate | | | | X | | Burrell (2007) |
| *R*-index | *R* index | | | | | X | Jin et al. (2007) |
| Specific impact index | *s* index | | | | X | | de Visscher (2010) |
| Tapered *h*-index | ht index | | | | | X | Anderson, Hankin, and Killworth (2008) |
| *t*-index | *t* index | | | | | X | Tol (2009) |
| Weighted *h*-index | *h* weight | | | X | | | Lee et al. (2009) |
| Wohlin index | wohlin index | | | | | | Wohlin (2009) |
| Wu index | wu index | | | | | X | Wu (2010) |
| | Total | 2 | 4 | 6 | 6 | 17 | |

which equals 4, $Y(i)$ is the publication year of an article $i$ and C($i$) are the articles citing the article $i$. Consequently, the contemporary $h$ index $h^c$ discounts older articles gradually, even if the articles still are gathering new citations.

(5) *The h index variant measures the citation intensity in the h core (all publications in a publication list having at least h citations) or gives more weight to highly cited papers*. One of the core properties of the $h$ index is its robustness; that is, "it is not influenced by a set of infrequently cited papers or by the (even severe) increase of citations to already highly cited papers" (Egghe, 2010, p. 67). Although this property might be conceived as a major advantage of the $h$ index, a host of studies sought to improve the $h$ index by measuring the citation intensity in the $h$ core and/or by giving more weight to highly cited papers. In this meta-analysis, the $g$ index could be regarded as an $h$ index variant that puts emphasis on highly cited papers in an author's publication list, since Egghe (2006a) defines the index as follows: "The highest number $g$ of papers that together received $g^2$ or more citations. From this definition it is already clear that $g \geq h$" (p. 8). In contrast, Jin's (2006) *A* index not only gives more weight to highly cited papers but also measures the citation intensity of all publications in the $h$ core. The *A* index equals $(1/h)\sum_{j=1}^{h} cit_j$, where $h$ stands for the $h$ index and cit stands for citation counts.

As Table 1 shows, a particular $h$ index variant can exhibit several of the above mentioned advantages over the $h$ index simultaneously. For instance, Jin's (2007) *AR* index, defined as $\mathrm{AR} = \sqrt{\sum_{j=1}^{h} \frac{cit_j}{a_j}}$, where $h = h$ index, cit = citation counts, and $a$ = number of years since publishing, not only measures the citation intensity in the $h$ core but also takes the age of the publications into account.

According to our taxonomy depicted in Table 1, 17 of the 37 $h$ index variants attempt to improve the $h$ index by giving more weight to highly cited papers and/or by measuring citation intensity in the $h$ core. Six variants attempt to balance out scientists differing career lengths and/or to take into account the age of a publication. Another six variants tackle the problem of multiple authorship, four variants are concerned with self-citation issues, and two variants seek to account for field normalization. There are a total of eight $h$ index variants in the present meta-analysis that cannot be assigned to any of the five features mentioned above. These unclassified indexes are intended to measure the whole production of a researcher (Wohlin index), the influential weight of a network (degree $h$ index $h_A$ and $h_P$), or higher levels of aggregated $h$ index values (Prathap's $h_2$). They account for different sizes of institutions ($h_{mol}$ index, MII) and seek to increase the discriminatory power in areas with low $h$ index values (multidimensional $h$ index $h_2$ and $h_3$).

In the literature on the $h$ index there can be found some indexes with homonymic names – that is, both Wohlin (2009) and Wu (2010) named their indexes the $w$ index, and both Schreiber (2008b) and Molinari and Molinari (2008) named their indexes the $h_m$ index. To distinguish the one same-named $h$ index variant from the other, we will call Wohlin's $w$ index the wohlin index, Wu's $w$ index the wu index, and Molinari and Molinari's $h_m$ index the $h_{mol}$ index. Schreiber's (2008b) $h_m$ index will not be renamed.

### 2.3. Statistical procedures

In our meta-analysis we are interested in the synthesis of correlations between various $h$ index variants and the $h$ index. This analysis requires three principal decisions:

First, there is broad discussion on whether to use raw correlations in the meta-analysis or Fisher's $Z$-transformed correlations ($Z = \tanh^{-1}(r)$). Overton (1998), for instance, advocated analyzing Fisher's $Z$-transformation, due to its better sample properties such as more stable variance and normality. Others (e.g., Hunter & Schmidt, 2004; Schulze, 2004) recommend the use of raw correlations, due to biased estimates using Fisher's $Z$-transformation. Hafdahl (2008, 2010) and Hafdahl and William (2009) offered a modified Fisher's $Z$-transformation to obtain unbiased estimates of correlations. In this study, we follow Overton (1998) and perform the meta-analysis (random effects model) using Fisher's $Z$-transformed correlations and Hafdahl's (2010) procedure to obtain unbiased mean correlations.

Second, it must be decided whether a fixed effects model or a random effects model should be chosen (Borenstein, Hedges, Higgins, & Rothstein, 2009). Following Hedges (1994) and Hedges and Vevea (1998) the fixed effects model ($M_0$) implies that the correlation estimates in the population are fixed but unknown constants. The correlation between the $h$ index and an $h$ index variant in the population is assumed to be the same for all studies included in the meta-analysis (homogeneous case). Therefore, the correlation estimate of a single study varies only randomly around the true correlation in the population (sampling error). The null hypothesis of homogeneous correlation can be tested using the Q statistic, which follows a $\chi^2$ distribution. As opposed to fixed effects models, random effects or mixed effects models assume that the population correlation themselves vary randomly from study to study (Cooper, Hedges, & Valentine, 2009). Therefore, in a mixture distribution framework the true correlations are sampled from a universe of possible correlations ("super-population"). Whereas fixed-effects models only allow generalizations about the studies that are included in the meta-analysis, in random-effects models the studies are assumed to be a sample of all possible studies that could be done on a given topic about which studies can be generalized.

From a statistical point of view, the main difference between these models is in the calculation of standard errors associated with the combined effect size. Fixed effects models only use within-study variability to estimate the standard errors. In random effect models, however, it is necessary to take into account two sources of error variability: within-study variability and between-study variability, which arises from differences between studies. In the heterogeneous case, the standard errors of the random-effects model are much larger than in the fixed effects model. In our meta-analysis random effects models are favored over fixed effects models in order to allow generalizability of the results over the specific studies included in the meta-analysis.

Third, the structure of the clustered data must be sufficiently captured by the model. In this study, the correlation coefficients are clustered within studies. One single correlation is calculated for a specific *h* index variant within a single study. However, the correlations with one single *h* index variant are not nested within one single study, because several studies report correlations for one and the same *h* index variant. There is an additional level of variability – that is, the correlations are cross-classified by *h* index variants and studies (Jayasinghe, Marsh, & Nigel, 2003). This leads to a three-level cross-classified mixed effects meta-analysis model: with sampling error for each correlation as level 1, with variability within cross-classification as level 2, and with cross-classification of correlations with specific *h* index variants by studies as level 3. Let $t_{i(jk)}$ and $\theta_{i(jk)}$ be the study *k*'s *i*th correlations and correlation parameter of the *j h* index variants, respectively, then the three-level *unstructured* mixed-effects model ($M_1$) can be defined as follows (Goldstein, 1994; Rasbash & Browne, 2008):

$$\begin{aligned} t_{i(jk)} &= \theta_{i(jk)} + \varepsilon_{i(jk)} \\ \theta_{i(jk)} &= \theta_{(jk)} + \tau_{i(jk)} \\ \theta_{(jk)} &= \tau_0 + \tau_{(jk)}, \end{aligned} \tag{1}$$

where $\tau_0 = E(\theta_{i(jk)})$ is the mean correlation (fixed effect), $\tau_{i(jk)}$ is the random effect of a correlation *i* within the combination of *h* index variant *j* and study *k* with $E(\tau_{i(jk)}) = 0$, and the variance component $\sigma_{\tau i(jk)}{}^2$, and $\tau_{(jk)}$ is the random effect of the combination of *h* index variant *j* and study *k* with $E(\tau_{(jk)}) = 0$, and the variance component of $\sigma_{\tau(jk)}{}^2$. The sampling error $\varepsilon_{ij}$ for a Fisher's *Z*-transformed correlation is normally distributed with $E(\varepsilon_{i(jk)}) = 0$ and $\sigma^2\varepsilon_{ijk} = 1/(n_{ijk} - 3)$, which is known for each correlation *i* of study *k* and *h* index variant *j* with sample size $n_{ijk}$. In the three-level *cross-classified* mixed effects model ($M_2$), the combined effect $\tau_{(jk)}$ is further split into two random effects: the random effects of the studies ($\tau_k$) with $E(\tau_k) = 0$, and variance component $\sigma_{\tau k}{}^2$, and the random effects of the *h* index variants ($\tau_j$) with $E(\tau_j) = 0$, and variance component $\sigma_{\tau j}{}^2$, as follows (Rasbash & Browne, 2008):

$$\begin{aligned} t_{i(jk)} &= \theta_{i(jk)} + \varepsilon_{i(jk)} \\ \theta_{i(jk)} &= \theta_{(jk)} + \tau_{i(jk)} \\ \theta_{(jk)} &= \tau_0 + \tau_j + \tau_k, \end{aligned} \tag{2}$$

The level-3 variation results from the sum of the variation of random effects of studies and *h* index variants. Additionally, the model can be extended by introducing explanatory covariates to explain the variation of random effects.

All in all, the following four models are estimated in this paper:

1. $M_0$: fixed effects model.
2. $M_1$: unstructured mixed effects model with three levels.
3. $M_2$: cross-classified mixed effects model with three levels.
4. $M_3$: cross-classified mixed effects meta-regression model.

In the meta-regression model $M_3$ the covariates described in the following Section 2.4 are included to explain random-effects variance. In case of correlations, the meta-regression is a moderator analysis: Statistically significant regression parameters indicate conditions where the correlations between *h* index and *h* index variants are significantly larger or lower than the average correlation.

All analyses were performed using SAS PROC MIXED in SAS, version 9.2. (Houwelingen van, Arends, & Stijnen, 2002; Little, Milliken, Stroup, Wolfinger, & Schabenberger, 2007).

### *2.4. Proposed covariates*

Two groups of covariates were included in the meta-regression analyses to explain the study-to-study variation of the reported correlation coefficients (see Table 2). One group – with the covariates "Correlation," "Sample units," "Literature database," "Discipline," "Country," "Journal Impact Factor," and "Sample size for correlation" – describes the dissimilarity of the studies that were included in the meta-analysis. The other group – with the covariates "Citation impact related *h* index variant," "*h* index variant considers scientific age," and "*h* index variant considers field, self-citations and/or multi-authorship" – describes the dissimilarity of the *h* index variants that were examined in the studies.

#### *2.4.1. Covariates describing the studies*

As higher correlation coefficients are to be expected when using the one method rather than the other, Table 2 lists the method used for the calculation of the correlation in a study (Pearson correlation or Spearman's rank correlation) as a covariate. Further, it can be assumed that the correlation coefficients vary in dependency on the sample (covariate: "Sample

**Table 2**
Description of the covariates.

| Covariate | | *N* | Percent | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|---|
| Correlation | Pearson | 35 | 25.9 | | | | |
| | Rank | 86 | 63.7 | | | | |
| | Not specified | 14 | 10.4 | | | | |
| Sample units | Person | 97 | 71.9 | | | | |
| | Other (journal, subfield etc.) | 38 | 28.1 | | | | |
| Literature database | Web of Science | 98 | 72.6 | | | | |
| | Other | 37 | 27.4 | | | | |
| Discipline | Social sciences, business, law | 22 | 16.3 | | | | |
| | Science | 62 | 45.9 | | | | |
| | Health and welfare | 45 | 33.3 | | | | |
| | Other fields | 6 | 4.4 | | | | |
| Country | European country | 42 | 31.1 | | | | |
| | Country not in Europe | 44 | 32.6 | | | | |
| | Not specified | 49 | 36.3 | | | | |
| Citation impact related *h* index variant | Yes | 72 | 53.3 | | | | |
| | No | 63 | 46.7 | | | | |
| *h* index variant considers scientific age | Yes | 13 | 9.6 | | | | |
| | No | 122 | 90.4 | | | | |
| *h* index variant considers field, self-citations and/or multi-authorship | Yes | 44 | 32.6 | | | | |
| | No | 91 | 67.4 | | | | |
| Journal Impact Factor[a] | | 123 | | 1.7 | .6 | .7 | 4.1 |
| Sample size for correlation[b] | | 135 | | 148.6 | 543.7 | 4 | 6000 |

*Notes*: The categorical covariates are effect-coded in the meta-regression analysis; the continuous covariates ("Journal Impact Factor" and "Sample size for correlation") are grand-mean centered.

[a] If there were missing values, a value of 0 was imputed.

[b] The variable is divided by 1000.

units"): For a study that calculated index values for persons (scientists), different coefficients are to be expected than for studies that calculated index values for journals or scientific subfields. In addition to these two covariates, as Table 2 shows, there are the further sample describing covariates "Literature data base" (this variable captures whether index values in a study were calculated using Web of Science data or using data from another literature data base, such as Scopus), "Discipline" (this covariate contains the information about what subfield a study was about, such as health and welfare), and "Country" (as the studies in part referred to scientists in certain countries, it was recorded whether the study concerned a European country or a country outside Europe).

With the covariates "Sample size for correlation" and "Journal Impact Factor" we examined whether there was a publication bias in the studies that were included in the meta-analysis. Publication bias refers to the case where certain results of a study (such as statistically significant results) increase the probability that the study will be published (or vice versa). With the covariate "Sample size for correlation" it is examined whether low correlation coefficients are published when there is a large sample (Hox, 2010) or high coefficients also when there is a small sample. This would result in a high correlation between sample size and coefficients across studies. Journal Impact Factors (JIFs), the second covariate here, are published by Thomson Reuters and are a measure of the "average" and fast response of the scientific community to a paper in a journal (Bornmann, Leydesdorff, & Marx, 2007). With this covariate it can be examined whether studies that are published in a journal with a high JIF (journals having greater visibility in the community) report lower correlation coefficients than studies from a journal with a lower JIF. The supposition is that journals with a high JIF are in favor of low correlations. For those studies for which no JIF could be found ($n = 8$), a value of 0 was used in the analysis. Thomson Reuters lists a JIF only for those journals that have a JIF that differs from 0.

### 2.4.2. Covariates describing the h index variants

As set out in the Introduction section above, the prime goal of developing *h* index variants was and is to compensate for certain weaknesses of the *h* index. In Section 2.2 we introduced the indices that were used in the studies included in the meta-analysis. The indices could be assigned to a total of five groups; each group contains indices that were developed with a similar aim. For example, in the group "*h* index variant considers scientific age" we included the indices that (in contrast to the *h* index) attempt to consider the scientific age of a researcher when calculating the index values. As Table 2 shows, three covariates were included in the meta-regression that referred to the grouping of the *h* index variants. They were: (1) "Citation impact related *h* index variant," (2) "*h* index variant considers scientific age", and (3) "*h* index variant considers field,

**Table 3**
Results for two mixed effects models without covariates (Fisher's *Z*-transformed correlations).

| | | $M_1$: Three-level unstructured | | $M_2$: Three-level cross-classified | |
|---|---|---|---|---|---|
| Term | Parameter | Estimate | S.E. | Estimate | S.E. |
| *Fixed effects* | | | | | |
| Intercept | $\beta_0$ | 1.42 | .08 | 1.37 | .11 |
| *Random effects* | | | | | |
| Level 3 "*h* index variant" | ${\sigma_{\tau j}}^2$ | | | .19* | .09 |
| Level 3 "Study × *h* index variant"/"Study" | ${\sigma_{\tau(jk)}}^2/{\sigma_{\tau k}}^2$ | .38* | .08 | .14* | .06 |
| Level 2 "Within level 3" | ${\sigma_{\tau i(jk)}}^2$ | .10* | .03 | .12* | .03 |
| Level 1 "Single result" | ${\sigma_{SE}}^2$ | .007 | | .007 | |
| $-2 \log L$ | | 248.7 | | 222.8 | |
| BIC | | 261.8 | | 222.8 | |

*Notes*: ${\sigma_{SE}}^2$ is identical to the standard error of the intercept of the fixed effects model and is fixed. S.E. = standard error, BIC = Schwarz–Bayesian information criterion, $-2 \log L$ = deviation.
* $p < .05$.

self-citations and/or multi-authorship" – a combination of the three groups "*h* index variant considers field dependence," "*h* index variant considers self-citations," and "*h* index variant considers multi-authorship," as only few indices could be assigned to each of these groups.

## 3. Results

### 3.1. General correlation between h index and h index variants

In Table 3 the results of model $M_1$ and $M_2$ are shown, the results for $M_0$ are reported only in the text. As mentioned above the results refer to Fisher-*Z*-transformed correlations, not to raw correlations. The null hypothesis that all bivariate correlations between *h* index and *h* index variants are the same except for random fluctuations due to sampling errors (${\sigma_{SE}}^2 = .007$) must be rejected (fixed-effects model, $M_0$). A statistically significant *Q*-statistic ($Q = 16{,}459.6^*$ $p < .05$) and a statistically significant variance component (${\sigma_{\tau(jk)}}^2 = .38^*$ $p < .05$) in model $M_1$ show that the correlations are heterogeneous and are sampled from a mixture population distribution. Therefore, the fixed-effects model does not fit the data. It cannot be assumed that there is the same high correlation between the *h* index variants and the *h* index across all *h* index variants and all studies (Field, 2001): The correlations are very different. Even correlations within a study for the same *h* index variant (calculated for example based on the data from different data bases) in the unstructured model $M_1$ vary statistically significantly with a variance component of ${\sigma_{\tau i(jk)}}^2 = .10^*$.

The statistically significant decrease of the deviation ($-2 \log L$) from 248.7 to 228.8 ($\chi_{M1-M2}$ (1) = 25.9* $p < .05$) and the decrease of the Schwarz–Bayesian information criterion (BIC) from 261.8 to 222.8 in Table 3 show that the cross-classified model ($M_2$) should be favored over the unstructured model ($M_1$). $M_2$ depicts the data structure better than $M_1$. Using $M_2$ we can answer the question as to whether the difference between the individual correlations can be attributed to differences between the *h* index variants (${\sigma_{\tau j}}^2$), to differences between the studies (${\sigma_{\tau k}}^2$), or to differences between correlations that were calculated within a study for one and the same *h* index variant (${\sigma_{\tau i(jk)}}^2$). All three variance components are statistically significant (${\sigma_{\tau j}}^2 = .19^*$, ${\sigma_{\tau k}}^2 = .14^*$, ${\sigma_{\tau i(jk)}}^2 = .12^*$). 42.2% of the sum of the three variance components is attributable to different correlations of the individual *h* index variants with the *h* index, 31.1% results from the heterogeneity of the studies, and 26.7% goes back to different correlations reported by a study for one *h* index variant (e.g., by using different samples). As almost half of the total random variance results from the different correlations of the *h* index variants with the *h* index, there appear to be – in addition to the *h* index variants where there is redundancy between them and the *h* index (that is, they correlate highly with the *h* index) – *h* index variants that make a non-redundant contribution to the *h* index (that is, the correlation between them and the *h* index is low).

The intercepts of $M_1$ und $M_2$ as expectancy values and the 95% confidence intervals (Fisher's *Z*-scale) are transformed to correlations using the procedure suggested by Hafdahl (2010) to determine the mean correlation of the *h* index variants with the *h* index across all studies. As the results in Table 4 show, the expectancy or mean value amounts to .826 at the 95% confidence interval [.784, .861] for the unstructured model and .791 [.753, .831] for the cross-classified model, respectively. Thus, the *h* index und *h* index variants generally correlate very highly. There are, however, slight differences between the

**Table 4**
Mean correlations between *h* index and *h* index variants (mixed effects models).

| Model | Description | *N* | Mean | CL95% | CU95% |
|---|---|---|---|---|---|
| $M_1$ | Unstructured | 135 | .826 | .784 | .861 |
| $M_2$ | Cross-classified | 135 | .791 | .743 | .831 |

*Note*: The correlations for the random effects model are calculated using Hafdahl's (2010) procedure rather than the usual Fisher's *Z*-transformation ($z \rightarrow r$).

two models in the expectancy values due to the fact that both models are conditional models: The mean values are the estimated expectancy values under the condition that the random effects of the Fisher's *Z*-transformed correlations are zero (and the covariates in $M_3$ are zero, respectively).

### 3.2. Explanation of the study-to-study variation by covariates

Table 5 shows the results of different meta-regressions ($M_{31} - M_{3final}$) that were calculated with the different covariates (see Section 2.4). First, an analysis was performed for each individual covariate ($M_{31} - M_{38}$), and then a multiple meta-regression ($M_{3final}$) was calculated in model $M_{3fin}$ with the covariates that were statistically significant in the individual analyses (see here Marsh, Bornmann, Mutz, Daniel, & O'Mara, 2009). Due to the fact that the categorical variables included are effect-coded and continuous variables are grand-mean centered, statistically significant parameters indicate conditions where the correlations between *h* index and *h* index variant are significantly higher or lower than the mean correlation.

As the results of the individual analyses ($M_{31} - M_{38}$) in Table 5 show, the covariates "Correlation" ($\beta_1$, $\beta_2$), "Sample units" ($\beta_3$), and "Country" ($\beta_5$, $\beta_6$) are statistically significant and therefore important moderators. Among these covariates the type of correlation ("Correlation") serves as the most important moderator variable: Using only the −2 log Likelihood deviation – which is here identical with the BIC – as a comparison standard, model $M_{31}$ is better than all other models except the final model. Hence, independently of the *h* index variant with which the *h* index was correlated, Spearman's rank correlation coefficients are statistically significantly higher than Pearson product-moment correlation coefficients. So after applying the usual Fischer's *Z*-transformation, the average rank correlation is .92 and the average Pearson correlation is .81. When the covariate "Correlation" is included in the model, the variance component of studies in $M_2$ ($\sigma_{\tau i(j)}{}^2$) decreases from .14 to .06 in model $M_{31}$, that is (.14 − .06)/.14 = 57.1% of the between-study variance is explained.

For the two other statistically significant covariates the following can be said with regard to the individual analyses in Table 5 ($M_{32} - M_{38}$): If a correlation refers to persons instead of other analysis units (such as journals) ($\beta_3$ = −.24, covariate "Sample units) or if the study refers to a country in Europe and not a country on another continent ($\beta_5$ = −.29, covariate "Country"), then the correlations of the *h* index variant with the *h* index are significantly lower than the mean correlation. As described in Section 2.4 above, using the covariates "Sample size for correlation" and "Journal Impact Factor" it was examined whether there is a publication bias in the studies included in the meta-analysis. The effects of the variables on the correlations are not statistically significant. Thus, there are no indications of a publication bias such as, for example, that high correlations are published more by journals with a high JIF.

In the final model $M_{3fin}$ all important covariates were included, that is, all covariates that were statistically significant in the individual analyses. As the results in Table 5 show, all moderator variables are statistically significant also here. In total, when the covariates are included in the final model ($M_{3fin}$), 100*(.14 − .009)/.14 = 93.5% of the between-study variance is explained. As the study variance component $\sigma_{\tau(jk)}{}^2$ is no longer statistically significant, in the meta-regression the differences between the studies are almost completely eliminated. As many cells of the cross-classification of *h* index variants by studies are empty (the individual studies deal with very different sets of *h* index variants), this can result in the reduction of the variance in one factor (here: study) leading to an increase of the variance in another factor (here: *h* index variant). As compared to $M_2$, in model $M_{3fin}$ the variance in the factor "*h* index variant" increases from .19 to .25.

### 3.3. Normalized correlation coefficients between h index and h index variants

Based on model $M_{3fin}$, empirical Bayes estimates and confidence intervals are calculated for the correlation of the *h* index with each of the total of 37 *h* index variants. These estimates take into account the different sample sizes and sampling errors of the studies, respectively, and are adjusted for the included covariates. The Fisher's *Z* Bayes estimates (and confidence intervals) are transformed to correlations by means of the usual Fisher's *Z*-transformation, to make possible better interpretation of the results. They are thus normalized correlation coefficients that can be compared with each other directly and can provide information on the "true" strength of the correlation between an *h* index variant and the *h* index. The stronger a correlation and Fischer's *z* value, respectively, deviate from the mean value, the more the value is shrunken to the mean to minimize the impact of unreliable outliers. For calculation of the confidence intervals we utilized the adjustment proposed by Goldstein and Healy (1995), which allows differences between correlation estimates to be interpreted as statistically significant at $\alpha$ = .05, provided that their confidence intervals do not overlap.

Fig. 1 shows the mean correlation (intercept $M_{3fin}$, "Mean") and the correlations of the *h* index with the *h* index variants with the adjusted confidence intervals. The mean correlation is .90. The *h* index variants were ranked from minimal correlation (at left) to maximal correlation (at right) with the *h* index. It is clearly visible that a large part of the *h* index variants correlates highly with the *h* index; however, for some *h* index variants the correlation coefficients are relatively low. Whereas the correlations between *f* index and *t* index and the *h* index are very high, the estimated correlations for the MII and *m* index are relatively low. These latter two differ statistically significantly from the first two in the correlation with the *h* index. Based on these results, we can assume that MII and *m* index make a non-redundant contribution to the *h* index. In contrast, it can be assumed that there is redundancy between the *f* index and the *t* index and the *h* index.

Overall, the confidence interval of the individual *h* index variants becomes clearly smaller from left to right in Fig. 1. This is not only due to the different standard errors of the random effects but also due to the nonlinear Fisher's *Z*-transformation. This means that an identical confidence interval on the Fisher's *Z* scale can produce large confidence intervals for correlations after

**Table 5**
Results for the cross-classified three-level mixed effects models $M_3$ with covariates (Fisher's *Z*-transformed correlations).

| Term | Parameter | $M_{31}$ est. | $M_{32}$ est. | $M_{33}$ est. | $M_{34}$ est. | $M_{35}$ est. | $M_{36}$ est. | $M_{37}$ est. | $M_{38}$ est. | $M_{3final}$ est. |
|---|---|---|---|---|---|---|---|---|---|---|
| *Fixed effects* | | | | | | | | | | |
| Intercept | $\beta_0$ | 1.38* (.11) | 1.47* (.12) | 1.43* (.11) | 1.39* (.11) | 1.40* (.12) | 1.37* (.16) | 1.37* (.11) | 1.35* (.11) | 1.49* (.11) |
| Rank correlation | $\beta_1$ | .23* (.10) | | | | | | | | .22* (.08) |
| Pearson correlation | $\beta_2$ | −.27* (.10) | | | | | | | | −.26* (.08) |
| Sample units (person) | $\beta_3$ | | −.24* (.07) | | | | | | | −.17* (.06) |
| Literature database | $\beta_4$ | | | −.11 (.06) | | | | | | |
| European country | $\beta_5$ | | | | −.29* (.09) | | | | | −.20* (.07) |
| Country not in Europe | $\beta_6$ | | | | .24* (.12) | | | | | .23* (.09) |
| Social sciences, business, law | $\beta_7$ | | | | | .09 (.14) | | | | |
| Science | $\beta_8$ | | | | | −.05 (.13) | | | | |
| Health and welfare | $\beta_9$ | | | | | −.19 (.17) | | | | |
| Citation impact related *h* index variant | $\beta_{10}$ | | | | | | −.08 (.14) | | | |
| *h* index variant considers scientific age | $\beta_{11}$ | | | | | | −.07 (.13) | | | |
| Citation impact × scientific age | $\beta_{12}$ | | | | | | −.24 (.15) | | | |
| *h* index variant considers field, self-citations and/or multi-authorship | $\beta_{13}$ | | | | | | .13 (.14) | | | |
| Journal Impact Factor | $\beta_{14}$ | | | | | | | .10 (.17) | | |
| Sample size for correlation/1000 | $\beta_{15}$ | | | | | | | | .19 (.11) | |
| *Random effects* | | | | | | | | | | |
| Level 3 "*h* index variant" | $\sigma_{\tau j}{}^2$ | .24* (.10) | .22* (.10) | .18* (.09) | .19* (.09) | .18* (.09) | .18* (.09) | .19* (.09) | .17* (.08) | .25* (.09) |
| Level 3 "Study" | $\sigma_{\tau k}{}^2$ | .06 (.04) | .12* (.06) | .11* (.06) | .10* (.05) | .11* (.06) | .13* (.06) | .14* (.06) | .14* (.06) | .009 (.03) |
| Level 2 "Within level 3" | $\sigma_{\tau i(jk)}{}^2$ | .12* (.03) | .10* (.02) | .12* (.03) | .11* (.03) | .13* (.03) | .12* (.03) | .12* (.03) | .12* (.03) | .11* (.03) |
| Level 1 "Single result" | $\sigma_{SE}{}^2$ | .007 | .007 | .007 | .007 | .007 | .007 | .007 | .007 | .007 |
| $-2 \log L$ | | 216.3 | 212.5 | 220.0 | 213.4 | 221.3 | 220.0 | 222.5 | 220.0 | 197.9 |

*Notes*: $\sigma_{SE}{}^2$ is identical to the standard error of the intercept of the fixed effects model and is fixed; est. = estimated parameter, S.E. = standard error.
$-2 \log L$ = deviation.
* $p < .05$.

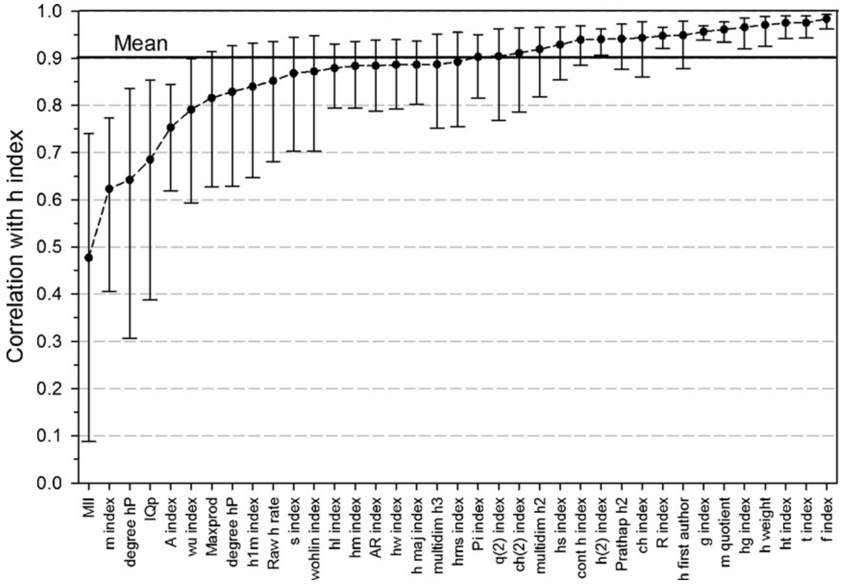


**Fig. 1.** Bayes estimates of the correlations between the *h* index and 37 *h* index variants with Goldstein-adjusted 95% confidence intervals (*Z*-values) for model $M_{3final}$, ranked from lowest to highest correlation (Goldstein & Healy, 1995). Two correlations between *h* index and *h* index variant with non-overlapping confidence intervals differ statistically significantly ($p < .05$).

Fisher's *Z*-transformation of small *Z*-values (small correlations) and small confidence intervals after Fisher's *Z*-transformation of high *Z* values (large correlations). As the standard errors and the accompanying confidence intervals were estimated correctly and these hardly differ between coefficients with high and low correlations, this is not a matter of a statistical artifact but a real result of our analyses.

## 4. Discussion

This study reports the results of the first meta-analysis that has been conducted on the *h* index and its variants. Through the analysis, the aim was to find out whether the development of the *h* index variants has made a non-redundant contribution to the *h* index. As the results show, with an overall mean value between .8 and .9, there is a high correlation between the *h* index and the *h* index variants. According to Navon (2009) "high correlations indicate that despite the differences in how the metrics are calculated, there is too much redundancy in the information they yield" (p. 2). Even if the "actual" correlations might be higher than that claimed in the original papers introducing the new *h* type indices due to systematic sampling selection effects, a mean correlation coefficient of between .8 and .9 is still high and might not justify the development of more and more *h* index variants.

The generally high correlation found between the *h* index and its variants does not imply that the relationship is valid for each individual case. According to the results of Vinkler (2010) journals with similar *h* index values show significantly different Pi index values and journals with similar Pi index values reveal different *h* index values. According to Vinkler (2009) the Pi index equals one hundredth of the number of citations obtained to the top square root of the total number of journal papers ranked by the decreasing number of citations. Furthermore, not all *h* index variants have a high correlation with the *h* index. Our results indicate that some *h* index variants have been developed that have a relatively low correlation with the *h* index, and it can be assumed that they can make a non-redundant contribution to the *h* index. These variants are mainly the MII (Sypsa & Hatzakis, 2009) and the *m* index (Bornmann et al., 2008). For evaluative purposes, then, the *h* index could be combined with these *h* index variants to better depict research performance bibliometrically (Garcia-Perez, 2009).

The characteristics of the *h* index variants that correlate less strongly with the *h* index could yield information about how to best complement the *h* index. For example, if we look at the *m* index (or also the *A* index) – it measures the citation intensity in the *h* core – we can assume that indicators that focus on the impact of the publications with the highest citation counts within a publication list could yield promising improvements to the original *h* index. This is in line with Bornmann et al. (2008; see also Bornmann, Mutz, et al., 2009), who showed by factor analysis that there are two independent types of *h* index variants, namely, those that describe the number of papers in the most productive core (e.g., *h* index or *g* index) and those that depict the impact of the papers in that core (e.g., *A* index or *m* index). Similarly, Antonakis and Lalive (2008) concluded that "what *h* seems to tap is quantity more than anything else" (p. 968). The MII (Sypsa & Hatzakis, 2009) – another *h* index variant that correlates relatively low with the *h* index – is a sophisticated indicator that does not make use of the *h* index algorithm and aims to facilitate the comparison of research units (e.g., institutions) of disparate sizes like Molinari and

Molinari's (2008) $h_{mol}$ but in a more sophisticated manner. Thus, it could be argued that a sophisticated version or genuine alteration of the *h* index algorithm could provide a promising complement of the *h* index.

A limitation of this study is that the method of meta-analysis is recommended for a quantitative review of experimental studies in which subjects were randomly assigned to a treatment or control group (Skrondal & Rabe-Hesketh, 2004). However, evaluative bibliometrics is not a 'medication' that can be tested in a randomized clinical trial (Marusic, 2005). For the meta-analysis we were therefore restricted to non-experimental studies that investigated the *h* index and its variants in a "natural setting." The restriction to the "natural setting" makes it very difficult to establish unambiguously whether the individual *h* index variants in fact provide information additional to the *h* index or not. This limitation of this meta-analysis is unavoidable. However, to check the stability of the results, it would be desirable a few years from now to conduct to a new meta-analysis on the *h* index and its variants using studies published up to then.

## References [1]


Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). *h*-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289. doi:10.1016/j.joi.2009.04.001

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2010). hg-Index: A new index to characterize the scientific output of researchers based on the *h*- and *g*-indices. *Scientometrics*, *82*(2), 391–400.

Anderson, T., Hankin, R., & Killworth, P. (2008). Beyond the Durfee square: Enhancing the *h*-index to score total publication output. *Scientometrics*, *76*(3), 577–588.

*Antonakis, J., & Lalive, R. (2008). Quantifying scholarly impact: *IQp* versus the Hirsch *h*. *Journal of the American Society for Information Science and Technology*, *59*(6), 956–969. doi:10.1002/Asi.20802

*Arencibia-Jorge, R., & Rousseau, R. (2009). Influence of individual researchers' visibility on institutional impact: An example of Prathap's approach to successive *h*-indices. *Scientometrics*, *79*(3), 507–516.

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the *h* index? *Journal of the American Society for Information Science and Technology*, *58*(9), 1381–1385.

Bornmann, L., & Daniel, H.-D. (2009). The state of *h* index research. Is the *h* index the ideal way to measure research performance? *EMBO Reports*, *10*(1), 2–6.

Bornmann, L., Leydesdorff, L., & Marx, W. (2007). Citation environment of *Angewandte Chemie*. *CHIMIA*, *61*(3), 104–109.

*Bornmann, L., Marx, W., & Schier, H. (2009). Hirsch-type index values for organic chemistry journals: A comparison of new metrics with the Journal Impact Factor. *European Journal of Organic Chemistry*, *2009*(10), 1471–1476.

*Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the *h* index? A comparison of nine different variants of the *h* index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, *59*(5), 830–837.

*Bornmann, L., Mutz, R., Daniel, H.-D., Wallon, G., & Ledin, A. (2009). Are there really two types of *h* index variants? A validation study by using molecular life sciences data. *Research Evaluation*, *18*(3), 185–190.

British Academy. (2007). *Peer review: The challenges for the humanities and social sciences*. London, UK: The British Academy.

Burrell, Q. (2007). Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. *Scientometrics*, *73*(1), 19–28.

Burrell, Q. L. (2009). Some comments on "A Proposal for a Dynamic *h*-Type Index" by Rousseau and Ye. *Journal of the American Society for Information Science and Technology*, *60*(2), 418–419.

*Cabrerizo, F. J., Alonso, S., Herrera-Viedma, E., & Herrera, F. (2010). q2-Index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. *Journal of Informetrics*, *4*(1), 23–28. doi:10.1016/j.joi.2009.06.005

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russel Sage Foundation.

*Costas, R., & Bordons, M. (2008). Is *g*-index better than *h*-index? An exploratory study at the individual level. *Scientometrics*, *77*(2), 267–288.

Csajbók, E., Berhidi, A., Vasas, L., & Schubert, A. (2007). Hirsch-index for countries based on Essential Science Indicators data. *Scientometrics*, *73*(1), 91–117.

*de Visscher, A. (2010). An index to measure a scientist's specific impact. *Journal of the American Society for Information Science and Technology*, *61*(2), 319–328.

Egghe, L. (2006a). An improvement of the *h*-index: The *g*-index. *ISSI Newsletter*, *2*(1), 8–9.

Egghe, L. (2006b). Theory and practise of the *g*-index. *Scientometrics*, *69*(1), 131–152.

Egghe, L. (2010). The Hirsch-index and related impact measures. *Annual Review of Information Science and Technology*, *44*, 65–114.

Egghe, L., & Rousseau, R. (2008). An *h*-index weighted by citation impact. *Information Processing & Management*, *44*(2), 770–780.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, *6*(2), 161–180.

*Franceschet, M. (2009). A cluster analysis of scholar and journal bibliometric indicators. *Journal of the American Society for Information Science and Technology*, *60*(10), 1950–1964.

Franceschini, F., & Maisano, D. A. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research*, *203*(2), 494–504. doi:10.1016/j.ejor.2009.08.001

Garcia-Perez, M. A. (2009). The Hirsch *h* index in a non-mainstream area: Methodology of the behavioral sciences in spain. *Spanish Journal of Psychology*, *12*(2), 833–849.

*García-Pérez, M. A. (2009). A multidimensional extension to Hirsch's *h*-index. *Scientometrics*, *81*(3), 779–785. doi:10.1007/s11192-009-2290-1

Glänzel, W. (2006). On the opportunities and limitations of the *H*-index. *Science Focus*, *1*(1), 10–11.

Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher*, *5*, 3–8.

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research*, *22*(3), 364–375.

Goldstein, H., & Healy, M. J. R. (1995). The graphical presentation of collection of means. *Journal of the Royal Statistical Society A*, *158*(Part 1), 175–177.

Good, P. I., & Hardin, J. W. (2009). *Common errors in statistics (and how to avoid them)* (3rd ed.). Hoboken, NJ, USA: Wiley-Interscience.

Hafdahl, A. R. (2008). Combining heterogeneous correlation matrices: Simulation analysis of fixed-effects methods. *Journal of Educational and Behavioral Sciences*, *33*(4), 507–533.

Hafdahl, A. R. (2010). Random-effects meta-analysis of correlations: Monte Carlo evaluation of mean estimators. *British Journal of Mathematical and Statistical Psychology*, *63*, 227–254.

Hafdahl, A. R., & William, M. A. (2009). Meta-analysis of correlations revisited: Attempted replication and extension of Field's (2001) simulation studies. *Psychological Methods*, *14*(1), 24–42.

*Harzing, A. W., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, *8*, 61–73.


[1] Studies included in the meta-analysis are marked with an asterisk.

*Haslam, N., & Laham, S. M. (2010). Quality, quantity, and impact in academic publication. *European Journal of Social Psychology*, *40*(2), 216–220. doi:10.1002/ejsp.727

Hedges, L. V. (1994). Fixed effects models. In H. M. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Russell Sage Foundation.

Hedges, L. V., & Vevea, J. L. (1998). Fixed and random effects models in meta-analysis. *Psychological Methods*, *3*, 486–504.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Hirsch, J. E. (2007). Does the *h* index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19193–19198. doi:10.1073/pnas.0707962104

Houwelingen van, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine*, *21*, 589–624.

Hox, J. J. (2010). *Multilevel analysis techniques and applications* (2nd ed.). New York: Routledge Taylor & Francis Group.

*Hu, X. J., Rousseau, R., & Chen, J. (2010). In those fields where multiple authorship is the rule, the *h*-index should be supplemented by role-based *h*-indices. *Journal of Information Science*, *36*(1), 73–85. doi:10.1177/0165551509348133

*Hua, P.-h., Wan, J.-k., & Wu, J.-h. (2010). A perfect Hirsch-type index? Experiences using the tapered *h*-index ($h_T$). *Chinese Journal of Scientific and Technical Periodicals*, *21*(1), 33–37.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA, USA: Sage Publications.

Jayasinghe, U. W., Marsh, H. W., & Nigel, B. (2003). A multilevel cross-classified modelling approach to peer review of grant proposals: The effects of assessor and researcher attributes on assessor ratings. *Journal of the Royal Statistical Society Series A – Statistics in Society*, *166*(Part 3), 279–300.

Jin, B. (2006). *h*-Index: An evaluation indicator proposed by scientist. *Science Focus*, *1*(1), 8–9.

Jin, B. (2007). The *AR*-index: Complementing the *h*-index. *ISSI Newsletter*, *3*(1), 6.

*Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The *R*- and *AR*-indices: Complementing the *h*-index. *Chinese Science Bulletin*, *52*(6), 855–863.

Kinney, A. L. (2007). *National scientific facilities and their science impact on non-biomedical research*. Retrieved July 16, 2007, from http://arxiv.org/abs/0705.1797

*Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original *h*-index. *ISSI Newsletter*, *2*(3), 4–6.

Kosmulski, M. (2007). MAXPROD – A new index for assessment of the scientific output of an individual, and a comparison. *Cybermetrics*, 11.

*Lee, J., Kraus, K. L., & Couldwell, W. T. (2009). Use of the *h* index in neurosurgery clinical article. *Journal of Neurosurgery*, *111*(2), 387–392. doi:10.3171/2008.10.jns08978

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* London: Sage Publications.

Little, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2007). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute Inc.

*Liu, Y., & Rousseau, R. (2007). Hirsch-type indices and library management: The case of Tongji University Library. In D. Torres-Salinas, & H. F. Moed (Eds.), *Proceedings of the 11th conference of the International Society for Scientometrics and Informetrics (Vol. 1)* (pp. 514–522). Madrid, Spain: Spanish Research Council (CSIC).

*Liu, Y., & Rousseau, R. (2009). Properties of Hirsch-type indices: The case of library classification categories. *Scientometrics*, *79*(2), 235–248.

*Lovegrove, B. G., & Johnson, S. D. (2008). Assessment of research performance in biology: How well do peer review and bibliometry correlate? *Bioscience*, *58*(2), 160–164. doi:10.1641/B580210

Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*, *79*(3), 1290–1326.

Marusic, A. (2005). Peers on peer review: 5th international congress on peer review and biomedical publication. *European Science Editing*, *31*(4), 112.

Matt, G. E., & Navarro, A. M. (1997). What meta-analyses have and have not taught us about psychotherapy effects: A review and future directions. *Clinical Psychology Review*, *17*(1), 1–32.

*Mingers, J. (2009). Measuring the research contribution of management academics using the Hirsch-index. *Journal of the Operational Research Society*, *60*(9), 1143–1153, 10.1057/jors.2008.94.

Molinari, J. F., & Molinari, A. (2008). A new methodology for ranking scientific institutions. *Scientometrics*, *75*(1), 163–174. doi:10.1007/s11192-007-1853-2

*Moussa, S., & Touzani, M. (2010). Ranking marketing journals using the Google Scholar-based hg-index. *Journal of Informetrics*, *4*(1), 107–117. doi:10.1016/j.joi.2009.10.001

Navon, D. (2009). The *h–i* index: A proposed new metric of individual scientific output. *Cybermetrics*, *13*(1), paper 3.

Norris, M., & Oppenheim, C. (2010). The *h*-index: A broad review of a new bibliometric indicator. *Journal of Documentation*, *66*(5), 681–705. doi:10.1108/00220411011066790

*Opthof, T., & Wilde, A. A. M. (2009). The Hirsch-index: A simple, new tool for the assessment of scientific output of individual scientists. The case of Dutch professors in clinical cardiology. *Netherlands Heart Journal*, *17*(4), 145–154.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, *3*, 354–379.

Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics*, *81*(3), 635–670. doi:10.1007/s11192-008-2174-9

Prathap, G. (2006). Hirsch-type indices for ranking institutions' scientific research output. *Current Science*, *91*(11), 1439.

Rasbash, J., & Browne, W. J. (2008). Non-hierarchical multilevel models. In J. de Leeuw, & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 301–334). New York: Springer.

*Rousseau, R., Yang, L., & Yue, T. (2010). A discussion of Prathap's $h_2$-index for institutional evaluation with an application in the field of HIV infection and therapy. *Journal of Informetrics*, *4*(2), 175–184. doi:10.1016/j.joi.2009.11.007

*Ruane, F., & Tol, R. S. J. (2007). Centres of research excellence in economics in the Republic of Ireland. *The Economic and Social Review*, *38*(3), 289–322.

*Sanderson, M. (2008). Revisiting *h* measured on UK LIS and IR academics. *Journal of the American Society for Information Science and Technology*, *59*(7), 1184–1190.

Schreiber, M. (2007a). *A case study of the Hirsch index for 26 non-prominent physicists*. Retrieved August 20, 2007, from http://arxiv.org/abs/0708.0120

Schreiber, M. (2007b). Self-citation corrections for the Hirsch index. *Europhysics Letters*, *78*(3), 30002.

*Schreiber, M. (2008a). An empirical investigation of the *g*-index for 26 physicists in comparison with the *h*-Index, the *A*-index, and the *R*-index. *Journal of the American Society for Information Science and Technology*, *59*(9), 1513–1522. doi:10.1002/Asi.20856

*Schreiber, M. (2008b). A modification of the *h*-index: The $h_m$-index accounts for multi-authored manuscripts. *Journal of Informetrics*, *2*(3), 211–216.

*Schreiber, M. (2009). A case study of the modified Hirsch index $h_m$ accounting for multiple coauthors. *Journal of the American Society for Information Science and Technology*, *60*(6), 1274–1282.

Schreiber, M. (2009). The influence of self-citation corrections and the fractionalised counting of multi-authored manuscripts on the Hirsch index. *Annalen der Physik*, *18*(9), 607–621. doi:10.1002/andp.200910360

Schubert, A. (2007). Successive *h*-indices. *Scientometrics*, *70*(1), 201–205.

*Schubert, A., Korn, A., & Telcs, A. (2009). Hirsch-type indices for characterizing networks. *Scientometrics*, *78*(2), 375–382.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA, USA: Hogrefe and Huber.

Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized Hirsch *h*-index for disclosing latent facts in citation networks. *Scientometrics*, *72*(2), 253–280.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models*. London, UK: Chapman & Hall/CRC.

*Sypsa, V., & Hatzakis, A. (2009). Assessing the impact of biomedical research in academic institutions of disparate sizes. *BMC Medical Research Methodology*, *9* doi:10.1186/1471-2288-9-33
Thompson, D. F., Callen, E. C., & Nahata, M. C. (2009). New indices in scholarship assessment. *American Journal of Pharmaceutical Education*, *73*(6), 111.
*Tol, R. (2009). The *h*-index and its alternatives: An application to the 100 most prolific economists. *Scientometrics*, *80*(2), 317–324.
van Eck, N. J., & Waltman, L. (2008). Generalizing the *h*- and *g*-indices. *Journal of Informetrics*, *2*(4), 263–271.
van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, *67*(3), 491–502.
*Vinkler, P. (2009). The pi-index: A new indicator for assessing scientific impact. *Journal of Information Science*, *35*(5), 602–612. doi:10.1177/0165551509103601
Vinkler, P. (2010). The pv-index: A new indicator to characterize the impact of journals. *Scientometrics*, *82*(3), 461–475. doi:10.1007/s11192-010-0182-z
White, H. D. (2005). On extending informetrics: An opinion paper. In P. Ingwersen, & B. Larsen (Eds.), *Proceedings of the 10th international conference of the International Society for Scientometrics and Informetrics (Vol. 2)* (pp. 442–449). Stockholm, Sweden: Karolinska University Press.
*Wohlin, C. (2009). A new index for the citation curve of researchers. *Scientometrics*, *81*(2), 521–533. doi:10.1007/s11192-008-2155-z
*Wu, Q. (2010). The *w*-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology*, *61*(3), 609–614.