

Perspective

# Productivity above replacement: a novel approach to assessing academic productivity in orthopedics and spine surgery

Andrew J. Schoenfeld, MD, MSc<sup>a,c,\*</sup>, Roman J. Schoenfeld<sup>b,c</sup>

<sup>a</sup> Department of Orthopaedic Surgery, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA

<sup>b</sup> Medfield Public Schools, 24 Pound Street, Medfield, MA 02052, USA

<sup>c</sup> Investigation performed at Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Received 9 April 2019; revised 11 April 2019; accepted 11 April 2019

---

**Keywords:** Academic productivity; h-index; Performance measures; Sabermetrics; Publication productivity; Performance above replacement (PAR)

---

In an era of increased financial constraints, academic programs are continuously challenged to improve the efficiency by which they deliver high-quality clinical care while also maintaining an educational mission and contributions to research [1]. As a result, department chairs, financial administrators, and other medical center leadership are pressured to increasingly scrutinize the contributions of their faculty to the components of the academic mission [1]. Although clinical productivity and participation in educational endeavors are fairly intuitive, effectively measuring academic productivity in an equitable way remains difficult [1,2]. This derives in part from the fact that no standardized measure exists to benchmark faculty productivity, or compare performance across members of a division, department or different institutions [1–6].

Historically, within the field of orthopedic surgery, deference has been given to the sheer volume of papers published by a faculty member [3–5]. Over the last decade, other measures that incorporate both prolificacy and impact (such as the Hirsch Index [h-index] [7]) have gained popularity [2,3,6]. Both estimates, however, have substantial limitations. For example, an unadjusted assessment of papers published in the literature does not account for recent

productivity and the commonly held maxim that two published first, or senior, author papers represents adequate research output may be outdated caused by advances in computing and the large-scale availability of clinical registries and claims data in recent years [2–5]. The h-index is biased in favor of established researchers (eg, scholars who have been publishing high volumes of literature for longer periods of time) and continues to increase whether a scholar is actively contributing to the literature [2,6]. Women and under-represented minorities in medicine may also be adversely affected by assessments restricted to the h-index alone. Given the increased number of outlets for publishing and availability of published manuscripts through electronic platforms, h-indices may also be rising at an unprecedented rate, such that what was previously viewed as an impressive h-index must be adjusted in light of inflation. There is also no standardized benchmark as to what constitutes a sufficient number of published manuscripts or a suitably high h-index [1–7]. Furthermore, many of these determinations are influenced by the researcher's field and the type of studies typically performed (eg, basic science research vs. prospective clinical trials vs. retrospective study) [7].

Similar challenges in terms of standardized assessments of productivity and contribution to a group effort are present in other fields, including professional sports. In that arena, sabermetrics have been developed to communicate the magnitude of an individual's standardized performance—especially in professional baseball [8]. One such measure is the *Wins Above Replacement*, which is reported to be the number of additional wins a team would anticipate if substituting a given player for a generic replacement [8]. In this paper, we describe a

---

FDA device/drug status: Not applicable.

Author disclosure: *AJS*: Royalties: Wolters Kluwer (B), Springer (A); Board of directors: JBJS (D); Grants: NIH-NIAMS (G), OREF (F), DoD (H), CMS-OMH (F); *RJS*: Nothing to disclose.

\* Corresponding author. Department of Orthopaedic Surgery, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA. Tel: 330-329-2594, fax: 617-278-6919.

E-mail address: [ajschoen@neomed.edu](mailto:ajschoen@neomed.edu) (A.J. Schoenfeld).

corollary for academic productivity, termed *Productivity Above Replacement* (PAR). This measure takes into account recent paper production, as a measure of current contribution to the literature, as well as the h-index which accounts for prominence in the academic community, prior publication record, and impact of the research portfolio as a whole.

Here, we describe the rationale behind the PAR measure, its application and present an example of its performance in a group of orthopedic faculty from seven academic medical centers. This is the first publication we are aware of to attempt to standardize the evaluation of academic productivity as a component of orthopedic practice, or apply sabermetrics to the study of academic performance.

## Development of a proof of concept

### *Description and calculation of the PAR*

The PAR takes into account an individual's recent publication record as well as their h-index and how these measures compare to peers in the field. A weighted sum of the difference between recent publishing productivity, h-index and the population medians are used to determine the PAR via the following formula:

$$\text{PAR} = [(3 - \text{year publications}_{\text{indiv}} - 3 - \text{year publications}_{\text{med}})(0.7) + [(h - \text{index}_{\text{indiv}} - h - \text{index}_{\text{med}})(0.3)]$$

Indiv and med represent the value for the individual and the population median, respectively. Individual publications are limited to first or senior author contributions to the literature in the preceding three calendar years. This restriction is in line with the stance of granting organizations and university promotion committees that generally consider these two author positions to be those which can claim *ownership* of the scholarly product [2]. Population medians are used in place of means to limit the potential for extreme outliers (eg, a small number of highly prolific and influential faculty) to skew the average estimate. The  $\Delta$ 3-year publication is weighted by 70% as the measure is intended to emphasize current productivity over established reputation. The 30% weight of the  $\Delta$ h-index accounts for the importance of prior achievements, standing in the field, and expertise.

The resultant PAR is thus a composite measure that encapsulates the aspects of both recent research productivity and the h-index. The score is standardized such that a PAR of 0 represents performance at the group median. For a score of zero, theoretical substitution with a generic replacement would result in no change in research productivity. Negative scores indicate that the individual is performing below the group median and positive scores indicate superior performance. There is no predetermined upper or lower limit for the PAR, as values are reflective of the average performance in the population compared against an individual's own academic productivity.

Each 1.0 point change in PAR is reflective of a standardized change of one in both the number of 3-year publications

and the h-index. As an example of how the weighting enables a balanced assessment, an industrious junior clinician scholar with 10 publications above the median in the last three years, but an h-index only three points above the median, would have a PAR of 7.9. This PAR would approximate that of a more established faculty member with an h-index 19 points above the median, who published three papers over the population median in the same time frame (PAR=7.8).

### *Evaluation of the PAR in a cohort of academic orthopedic faculty*

We obtained data on 3-year publications and h-index for listed faculty in the divisions of joint arthroplasty and spine surgery at seven academic medical centers (University of Southern California, Mayo Clinic, Washington University in St. Louis, Harvard Medical School [faculty practicing at Brigham and Women's Hospital and Massachusetts General Hospital only], Rush Medical College, Thomas Jefferson University (Rothman Institute) and the University of Pennsylvania). These institutions were selected in part based on a recent publication listing them as among the top performing academic centers in the field [9], as we wanted to ensure robust variation in research productivity and h-indices across junior and senior faculty to demonstrate the performance of the PAR metric.

Faculty members at each institution were obtained from listings with the North American Spine Society Fellowship Directory [10] for spine faculty and the American Academy of Orthopaedic Surgeons Post Graduate Fellowships Directory [11] for joint arthroplasty, as well as through a search of program websites. Part-time faculty or clinical adjunct faculty were not included. H-index and the number of first or senior author publications in calendar years 2016 to 2018 were obtained through Scopus for each faculty member identified [12].

STATA v15.0 (STATA Corp., College Station, TX, USA) was used to determine the population medians and calculate the PAR. The performance of the measure was evaluated by assessing the distribution, including mean and median PAR, range and 95% CI, across the entire population, as well as within each institution. Normality was determined using skew and kurtosis. The PAR performance was assessed across all faculty members without respect to institution, but results are summarized at the institutional level to preserve privacy. Data collection was completed on March 4, 2019. This effort did not require institutional review board approval as it relies on publicly available information.

## Demonstration of the proof of concept

PAR scores were determined for 92 faculty members across seven institutions (Table 1). Forty-nine (53%) faculty members were members of their department's spine division. The median number of 3-year publications was nine (interquartile range [IQR] 3, 18.5)], with a mean of 12.2 (SD 12.4) and a range of 0 to 78. The median h-index

Table 1

Average 3-year publications and h-index of the clinician-researchers used to demonstrate the performance of the Productivity Above Replacement (PAR) metric, by institution

Institution	Number of clinician-researchers	Publications (range)	h-index (range)
University of Southern California	11	8.55 (0–24)	27.09 (2–64)
Mayo Clinic	16	13.06 (2–30)	31.44 (4–66)
Washington University in St. Louis	13	11.46 (3–24)	33.46 (11–91)
Harvard Medical School	19	11.68 (0–35)	21.89 (2–47)
Rush Medical College	12	10.00 (0–42)	41.17 (8–83)
Thomas Jefferson University	13	21.15 (2–78)	38.00 (13–81)
University of Pennsylvania	8	7.00 (0–21)	15.25 (0–24)

was 24.5 (IQR 14.5, 44), with a mean of 30.0 (SD 21.2) and a range of 0 to 91. At the institutional level, mean 3-year publications ranged from 7.0 to 21.15 and mean h-index from 15.25 to 38.0.

Among spine faculty, median number of 3-year publications was 9 (interquartile range [IQR 3, 14]), with a mean of 11.8 (SD 13.7) and a range of 0 to 78. The median h-index was 25 (IQR 15, 37), with a mean of 28.8 (SD 21.1) and a range of 2 to 91.

In the population as a whole, the resulting PARs ranged from  $-13.65$  to  $63.75$ , with a mean PAR of  $4.2$  (SD  $12.6$ ) and a median of  $3.5$  (IQR  $-5.7, 12$ ). The scores approximated a normal distribution (Figure) with a skew of  $1.4$  and a kurtosis of  $7.1$ . When limited to spine faculty, PARs ranged from  $-12.45$  to  $63.75$ , with a mean PAR of  $3.7$  (SD  $13.0$ ) and a median of  $2.6$  (IQR  $-5.5, 11.65$ ).

The average PAR varied substantially across institutions, however, all had overlapping 95% CIs and publication ranges (Table 2). All institutional 95% CIs and publication ranges included the population point estimates for the mean and median PAR.

### Assessment of the PAR and its utility

Here we have presented how to calculate the PAR and demonstrated the score's performance in a cohort of

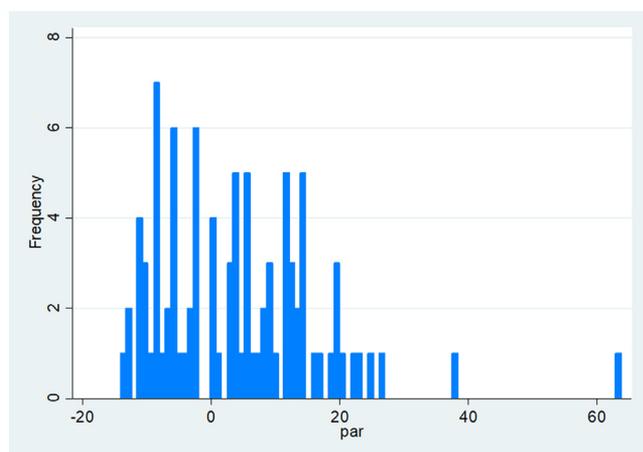


Figure. Distribution of the PAR among 92 faculty members at the seven academic institutions considered in this analysis.

academic orthopedic faculty from different medical centers. The PAR is relatively intuitive and represents an effective way to standardize comparisons of academic productivity across faculty members within a department, institution, or healthcare organization. Furthermore, in the example presented here, the score maintained a relatively normal distribution with a mean and median close to each other and the null. The score also did not fundamentally disadvantage, or favor, any particular institution with the range and 95% CI of scores at the various academic centers all overlapping. We believe that the PAR is likely to perform equally well in any healthcare setting where the expectations, motivating factors, and emphasis on research productivity are similar to the institutions [9] included in our study example.

The PAR incorporates aspects of recent publishing activity and h-index and, thus is reflective of both current research productivity as well as the impact of an individual's research portfolio, past productivity, and standing in the research community. We believe that the PAR is uniquely advantaged over historical measures of academic productivity, such as raw number of publications or h-index as a stand-alone measure. An individual's total publication count may be more reflective of past activity than recent contributions [1–6]. Unadjusted consideration of publications, without deference to direct contribution to the work (as is usually conveyed by position in the author order), may also artificially inflate the sense of an individual's presence in the literature [2]. Faculty members who are early in their career, by definition, will not be able to have a high h-index which only rises through increased publications in the literature that also garner sufficient number of citations [6,7]. Often, the number of citations a particular paper may attain is beyond the control of the first or senior author. By incorporating a weighted component of both recent productivity and the h-index, the PAR balances many of these concerns, although limiting consideration to first or senior authors ensures that the work is attributed to the individuals making the most robust contributions to the research effort. Faculty members are also not anticipated, or expected, to maintain the same levels of productivity at all times over the course of a career. The PAR accounts for this fact, but in a way that still provides equity to junior faculty or less established researchers.

Table 2  
Average Productivity Above Replacement (PAR) metrics with 95% confidence interval (CI), by institution

Institution	PAR	95% CI	Range
University of Southern California	2.58	−5.34, 10.49	−12.45, 22.35
Mayo Clinic	4.93	0.05, 9.80	−8.35, 22.95
Washington University in St. Louis	4.41	−1.42, 10.25	−6.15, 20.25
Harvard Medical School	1.10	−4.25, 6.44	−13.05, 19.85
Rush Medical College	5.7	−3.11, 14.51	−11.25, 26.15
Thomas Jefferson University	12.56	1.17, 23.95	−8.35, 63.75
University of Pennsylvania	−4.2	−10.03, 1.68	−13.65, 7.65

PAR values are predicated on the overall performance of the community used to develop the median scores. In professional sports, productivity measures are widely available and updated almost daily for all players [8]. More effort is required to determine productivity scores for orthopedic faculty [2–6]. The example presented here, by intention, used faculty members from high performing institutions where academic productivity was known to be robust [9] and variation in publication numbers and h-indices were anticipated to be present across scholars. The median values for 3-year publications and h-index may be higher for the population considered here when compared with other contexts. For example, Bastian et al. reported that the average h-index for professors of orthopedic surgery across the country was 15.1 and 17.8 for department chairs [6]. Ence et al. estimated that the median h-index for academic orthopedic faculty as a whole was 5 (IQR 1, 12) [4]. In a study limited to spine fellowship faculty in academic and private practice environments, Schoenfeld et al. found that the median 3-year publication figure was two (range 0–54) and the mean h-index was 13.2 (SD 0–63) [2]. The time that has transpired between the completion of these prior works [2–7] and the current study prevents meaningful comparison of the h-index as this score continues to evolve in real time as additional publications accrue and citations are added [7]. Interestingly, the median figure of nine publications over a 3-year period encountered here approximates that of Internal Medicine faculty at institutions where research productivity was integral to the compensation plan [1].

We envision that the PAR has a number of immediate applications in academic practice. At a minimum, it can be used as a standardized benchmark of research productivity. It can also be used to establish goals for academic progress, or performance incentives intended to increase academic productivity [1]. In the appropriate context (eg, other academic environments where productivity, incentives, and performance are assumed to be comparable to the institutions included in our case example [9]), the population medians presented here may be used as a referent. The beauty of the PAR, however, is that the formula is not dependent on a particular base estimate for 3-year publications or the h-index. Thus, those responsible for calculation of the PAR are able to choose the source population which will serve as the comparator for individuals, based on the derivation of figures for a *generic replacement*. This could

be as restrictive as members of a single orthopedic department, or extended to all faculty in a particular region, members of a specialty society, orthopedic sub-specialty, or the field as a whole as long as figures are available and the means of calculation remain transparent. The figures for the *generic replacement* also do not have to be set at the median; 25% performance percentile, 75% percentile or the mean can serve as viable substitutes depending on the desired emphasis on academic achievement. In the case example we chose to include all first or senior author publications listed in Scopus. Further modifications that may be appropriate could restrict consideration to original research contributions, as opposed to review articles, or publications that appear in journals with impact factors above a predetermined value. One important point to consider is that the 3-year publications<sub>med</sub> and h-index<sub>med</sub> should be updated semi-annually or annually, as these figures are unlikely to remain constant with time and h-index inflation invariably occurs as the cumulative healthcare literature continues to expand [2,6,7].

Although the PAR is composed of the two most common measures used to evaluate academic productivity in orthopedics today [2–6,9], we realize other aspects of academic productivity are not considered in this estimate. These include grants received and courses directed at a medical school, among others. However, recent research would indicate that these aspects of the academic portfolio are not regularly encountered among the plurality of orthopedic faculty in the United States [13]. Although we demonstrate the score's performance in the context of academic orthopedic departments, nothing about the measure would restrict its use to surgical fields, or orthopedics in particular. We report our findings in the case example at the institutional level to demonstrate the score's performance across healthcare organizations and also to protect the privacy of individual faculty members. The score is not intended to provide meaningful comparisons between institutions as a whole. It should not be seen as a ranking measure and it is inappropriate to use the PAR as such.

In conclusion, we have described and demonstrated the performance of the PAR metric, a novel measure which can be used to evaluate standardized academic performance among faculty. The PAR is relatively easy to calculate, intuitive to understand, and performed well in our case

example within and across institutions known to have productive academic orthopedic departments. We believe that the PAR can be used as a standardized benchmark of academic productivity, a means to set goals for academic achievement, or a way to incentivize increased academic performance. This descriptive paper may serve as a general literature citation when the PAR is employed as a means to evaluate academic productivity.

## References

- [1] Leverence R, Nuttall R, Palmer R, Segal M, Wood A, Yancey F, et al. Using organizational philosophy to create a self-sustaining compensation plan without harming academic missions. *Acad Med* 2017;92:1133–7.
- [2] Schoenfeld AJ, Bhalla A, George J, Harris MB, Bono CM. Academic productivity and contributions to the literature among spine surgery fellowship faculty. *Spine J* 2015;15:2126–31.
- [3] Ence AK, Cope SR, Holliday EB, Somerson JS. Publication productivity and experience: factors associated with academic rank among orthopaedic surgery faculty in the United States. *J Bone Joint Surg Am* 2016;98:e41.
- [4] Zelle BA, Weathers MA, Fajardo RJ, Haghshenas V, Bhandari M. Publication productivity of orthopaedic surgery chairs. *J Bone Joint Surg Am* 2017;99:e62.
- [5] Milone MT, Bernstein J. On track to professorship? A bibliometric analysis of early scholarly output. *Am J Orthop (Belle Mead NJ)* 2016;45:E119–23.
- [6] Bastian S, Ippolito JA, Lopez SA, Eloy JA, Beebe KS. The use of the h-index in academic orthopaedic surgery. *J Bone Joint Surg Am* 2017;99:e14.
- [7] Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 2005;102:16569–72.
- [8] Baseball reference. wRAA for position player WAR explained. Sports Reference, LLC. 2019. Available at: [https://www.baseball-reference.com/about/war\\_explained\\_wraa.shtml](https://www.baseball-reference.com/about/war_explained_wraa.shtml). Accessed March 7, 2019.
- [9] Namavar AA, Loftin AH, Khahera AS, Stavrakis AI, Hegde V, Johansen D, et al. Evaluation of US orthopaedic surgery academic centers based on measurements of academic achievement. *J Am Acad Orthop Surg* 2019;27:e118–26.
- [10] North American Spine Society. Spine fellowship training directory. Burr Ridge, IL: North American Spine Society; 2019. Available at: <https://www.spine.org/Portals/0/Documents/PolicyPractice/Fellowship-Directory.pdf?ver=2018-11-01-155959-633>. Accessed March 3, 2019.
- [11] American Academy of Orthopaedic Surgeons. Postgraduate Orthopaedic Fellowships Directory 2018. Rosemont, IL: American Academy of Orthopaedic Surgeons; 2019. Available at: <https://www7.aaos.org/education/fellowshipbook/fellowshiplisting/ARTHRITISRECON/No%20Preference/0/1?ct=1e2af5ccf8763e1adafe5b512fe638359df5989d67937ff29-d304155abd7720bfe6b7d585882624063cd8fca974d97d9c505025abfbd-bce029aba3f30c302a43>. Accessed March 3, 2019.
- [12] Elsevier. Scopus. Waltham, MA: Elsevier BV; 2019. Available at <http://www.scopus.com>. Accessed March 4, 2019.
- [13] Hegde V, Johansen D, Park HY, Zoller SD, Hamad C, Bernthal NM. The relationship between OREF grants and future NIH funding success. *J Bone Joint Surg Am* 2017;99:e87.