The MIT Press

RESEARCH ARTICLE

Bibliometrics-based decision trees (BBDTs) based on bibliometrics-based heuristics (BBHs): Visualized guidelines for the use of bibliometrics in research evaluation

Lutz Bornmann iD

Administrative Headquarters of the Max Planck Society, Division for Science and Innovation Studies, Hofgartenstraße 8, 80539 Munich, Germany

ABSTRACT

Fast-and-frugal heuristics are simple strategies that base decisions on only a few predictor variables. In so doing, heuristics may not only reduce complexity but also boost the accuracy of decisions, their speed, and transparency. In this paper, bibliometrics-based decision trees (BBDTs) are introduced for research evaluation purposes. BBDTs visualize bibliometrics-based heuristics (BBHs), which are judgment strategies solely using publication and citation data. The BBDT exemplar presented in this paper can be used as guidance to find an answer on the question in which situations simple indicators such as mean citation rates are reasonable and in which situations more elaborated indicators (i.e., [sub-]field-normalized indicators) should be applied.

1. INTRODUCTION

Bibliometrics are frequently used in research evaluation. In some situations, peer review and bibliometrics are combined in an informed peer review process. According to Jappe, Pithan, and Heinze (2018), for instance, 11 of 36 assessment units in the 2014 UK Research Excellence Framework (REF) were allowed to see citation benchmarks. Bibliometrics are considered "to break open peer review processes, and stimulate peers to make the foundation and justification of their judgments more explicit" (Moed, 2017, p. 13). In other situations, "desktop bibliometrics" are relied upon. The term *desktop bibliometrics* describes the application of bibliometrics by decision makers (e.g., deans or administrators) without involving experts (i.e., scientists) from the evaluated fields (Leydesdorff, Wouters, & Bornmann, 2016). Another characteristic of "desktop bibliometrics" is the application of inappropriate indicators for measuring performance, since bibliometrics experts are not involved. Informed peer review processes and "desktop bibliometrics" exist side by side in the research evaluation landscape.

Although many literature overviews of bibliometrics and guidelines for their use exist (see, e.g., Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015), the recommendations are frequently not very clearly formulated, leaving ample room for ambiguities. At the same time, research evaluation situations themselves come with ambiguities and uncertainties: What counts as indicator for quality in a given field, which indicators serve to measure quality, and how precise are those measurements in a given field (e.g., machine learning) and for a

given task (e.g., evaluate an individual scientist as opposed to a university)? What are the consequences if certain indicators are used, including politically at the level of the department or university concerned? In practice, often ad hoc decisions, local preferences and politics as well as the demands produced by a specific evaluation situation might dictate how bibliometrics are relied upon.

In the current paper, a decision tree is presented that can be used to decide how to use bibliometrics in research evaluation situations (with or without peer review). The decision tree is precisely formulated and transparent, making it easy to understand, communicate about, and actually use. The tree is grounded in available empirical evidence and applicable across fields.

2. BIBLIOMETRICS-BASED HEURISTICS (BBHS)

The decision trees introduced in this paper are fast-and-frugal heuristics. Such heuristics are simple decision strategies that ignore available information, basing decisions on only a few relevant predictor variables. In so doing, fast-and-frugal heuristics can not only aid reducing complexity but also make fast and transparent decisions; systematically ignoring (irrelevant) information also aids making accurate decisions (Gigerenzer & Goldstein, 1996). Decision trees are grounded in the fast-and-frugal heuristics framework (Gigerenzer, Todd, & ABC Research Group, 1999). That framework, originally developed within the cognitive and decision sciences, has fueled a large number of studies indicating that heuristics can help people make smart decisions in business, law, medicine, and many other task environments. For example, Luan and Reb (2017) show that a significant proportion of managers use fast-and-frugal decision trees (lexicographic heuristics) to make performance-based personnel decisions. In these environments, the strategies achieved performance competitive with more complex approaches (e.g., multiple regression analyses).

Recently, Bornmann and Marewski (2019) extended the fast-and-frugal heuristics framework to research evaluation and formulated a research program for investigating bibliometric-based heuristics (BBHs). BBHs characterize decision strategies in research evaluations based on bibliometrics data (publications and citations). Other data (indicators) besides bibliometrics are not considered. BBHs might be especially qualified for research evaluation purposes, because citations and other bibliometric data are deeply rooted in the research process of nearly every researcher: Researchers are being prompted to make all their results publicly available and embed the results in published research by citing the corresponding publications—research stands on the shoulders of giants (see Merton, 1965).

BBHs may or may not be integrated in peer review processes (Bornmann, Hug, & Marewski, 2019). Initiatives such as the San Francisco Declaration on Research Assessment (DORA; https://sfdora.org) demonstrate that the use of bibliometrics is prevalent in science evaluation.[1] Decision makers in science (e.g., reviewers) do not have unlimited time. Moreover, just like all other humans, scientific decision makers have limited information-processing capacities, putting natural constraints on their ability to tackle computationally demanding evaluation tasks. At the same time, evaluators often have limited knowledge of the subject area at hand, and even if the decision maker is an expert in a field (e.g., decision making) he or she

---

[1] The general recommendation of DORA is not to use "journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions" (see https://sfdora.org/read).

might not be an expert in the target area of research (e.g., decisions with the take-the-best heuristic; see below)—a truism, partially fueled by extreme specialization tendencies in some fields.

In short, decisions in science are made in the context of limited information processing capacity, time, and knowledge (see Marewski, Schooler, & Gigerenzer, 2010). To use a term coined by Nobel Laureate Herbert Simon, decision makers' rationality is bounded. Heuristics are models of bounded rationality. They are rules of thumb which perform well under conditions of limited time, knowledge, and information-processing capacity (Katsikopoulos, 2011). They do not use all the information available in a given decision environment but a selection for making reasonable decisions that are ecologically adequate (e.g., bibliometrics data in the case of BBHs). Thus, heuristics "involve partial ignorance" (Mousavi & Gigerenzer, 2017, p. 376). The more redundancy and intercorrelations there are in the complete information, the better are the decisions based on selected information (Marewski et al., 2010).

Heuristics frequently consist of search rules, stopping rules, and decision rules: "A *search rule* that specifies what information (e.g., predictor variables) is searched for and how (i.e., in what order), a *stopping rule* that delineates when information search comes to an end, and a *decision rule* that determines how the acquired information is employed (e.g., combined) to classify objects (e.g., patients)" (Bornmann & Marewski, 2019, p. 424). If we transfer these rules to the area of research evaluation, a one-cue heuristic (BBH) could be as follows: Imagine a funding organization in biomedicine with the goal of selecting exceptional scientists for a group leader position. The organization is especially interested in scientists with an excellent publication record who are selected in an informed peer review process: The reviewers decide based on an extensive bibliometric report and by reading selected publications. Some years ago, however, the organization was confronted with the problem of receiving too many applications. The informed by peer review process did not have the capacity to (properly) review all applications. Thus, the organization decided to introduce the following one-cue BBH for a preselection of applications. A smaller pool of the applications is then reviewed by the peers.

The preselection is based on a single indicator that targets an important goal of the organization: research excellence (expressed by an exceptional publication record). The three building blocks for the one-cue BBH are as follows:

1. *Search rule*: Search all publications (articles and reviews) published by the applicants in Web of Science (WoS, Clarivate Analytics). Download the publications.
2. *Stopping rule*: Send the publication lists to the applicants for validation. If publications are missing, improve the search rules. Add to the validated lists information about whether the publications belong to the 10% most frequently cited publications in the corresponding subject category and publication year (i.e., whether they are highly cited publications).
3. *Decision rule*: Divide the number of highly cited publications by the number of years since publishing the first publication (to generate age-normalized numbers: see Bornmann & Marx, 2014). Sort the applicants by the age-normalized number of highly cited publications in descending order and select the top x% of applicants. These are the applicants for reviewing by the peers.

The organization involved scientometric experts, experts from biomedicine, and representatives of the organization to (empirically) check that the BBH fulfills the desired objective. The BBH is annually evaluated as to whether it should be improved (reformulated) or not.

The application of this and similar BBHs does not mean that decisions based on bibliometrics are recommended for all evaluation contexts. Rather, the fast-and-frugal heuristics research program assumes that any given heuristic is suitable only for certain, but not all, task environments: Heuristics are ecologically rational, not globally rational (Todd, Gigerenzer, & ABC Research Group, 2012). That is not different for BBHs, which can and should only be used in selected situations (e.g., for the selection process of a specific funding organization—see above). These situations can be identified, for instance, by studies that compared assessments based on bibliometrics and peer review. To illustrate this point, Traag and Waltman (2019) analyzed the agreement between metrics and peer review in the UK REF 2014. Their model suggests that "for some fields, the agreement between metrics and peer review is similar to the internal agreement of peer review. This is the case for three fields in particular: Clinical Medicine, Physics, and Public Health, Health Services & Primary Care."

Using the same data set, Rodriguez Navarro and Brito (2019) conclude as follows: "the present results indicate that in many research areas the substitution of a top percentile indicator for peer review is possible". Similar results have been published by Pride and Knoth (2018) and Harzing (2017). Comparing scores from international university rankings and bibliometrics, Robinson-Garcia, Torres-Salinas, Herrera-Viedma, and Docampo (2019) conclude that "ranking scores from whichever of the seven league tables under study can be explained by the number of publications and citations received by the institution (p. 232)." Thus, ranking scores might be substituted by bibliometrics. The results of all these studies point out that simplified decision rules based on bibliometrics may provide fast and accurate decisions that may not deviate much from peers' decisions or university rankings.

From this ecological view, BBHs in general are neither good nor bad. They can be assessed only with respect to the evaluation situation in which they are applied (see Moed, Burger, Frankfort, & van Raan, 1985). The better the functional match is between a certain BBH and its evaluation situation, the higher the level of its ecological rationality (see Mousavi & Gigerenzer, 2017; Waltman, 2018). Specifically, the fast-and-frugal heuristics research program assumes that people select between different heuristics as a function of the task environment at hand. Selecting the adequate heuristic for a given task can aid in making clever decisions. Hence, in addition to describing how people make decisions, models of heuristics can also be interpreted prescriptively: What heuristics should decisions makers, ideally, use in a given environment in order to produce desirable outcomes? In which environment does a certain heuristic perform well with respect to criteria such as accuracy, frugality, speed, or transparency of decision making—or not?

Bibliometrics-based decision trees (BBDTs), which are introduced in this study, are visualized BBHs (heuristics are usually formulated as text only). BBDTs consist of a sequence of nodes with questions which are answered for a specific evaluation situation (see Katsikopoulos, 2011). Exits at the nodes lead to appropriate bibliometrics for the situation. BBDTs can be seen as adaptive bibliometrics toolboxes (including selected BBHs) which functionally match with certain evaluation tasks. The goal of BBDTs is prescriptive by recommending when (and how) one should use which BBH or bibliometric indicator, respectively (see Raab & Gigerenzer, 2015).

The bibliometrics literature provides many hints as to which bibliometrics data and indicators should be applied (or not) in certain environments. The following sections draw on that literature and explain decision trees (BBDTs), which can be used to decide on the proper use of appropriate heuristics (BBHs) in concrete evaluation situations. Although these BBDTs try to include only rules that might be standard in the bibliometrics field, the standards are frequently questioned (e.g., Opthof & Leydesdorff, 2010) and lead to internal discussions (e.g., van Raan, van Leeuwen, Visser, van Eck, & Waltman, 2010).

3. DECISION TREES

Decision trees can be defined as visualized lexicographic heuristics that are used for the categorization of cases (Kurz-Milcke & Gigerenzer, 2007; Martignon, Vitouch, Takezawa, & Forster, 2011). The term "lexicographic" has its roots in the term "lexicon," in which entries are sorted by the order of their letters. Kelman (2011) defines lexical decisions as follows: "A decision is made lexically when a subject chooses A over B because it is judged to be better along a single, most crucial dimension, without regard to the other 'compensating' virtues that B might have relative to A. Thus, for instance, one would have chosen some restaurant A over B in a lexical fashion if one chose it because it were cheaper and did not care about other traits, like its quality, proximity, level of service, and so on" (p. 8). Take-the-best heuristics are typical representatives of lexicographic heuristics that work with the following building blocks (see Scheibehenne & von Helversen, 2009):

Search rule: The most important cue is searched among the available cues.
Stopping rule: The search is stopped if the option with the highest cue value is found.
Decision rule: This option is selected. If several options with equal cue values exist, the next important cue is considered (see Marewski et al., 2010; Raab & Gigerenzer, 2015).

In a series of computer simulations with 20 real-world data sets, Czerlinski, Gigerenzer, and Goldstein (1999) showed that lexicographic heuristics outperform more complex inference methods, such as multiple linear regressions. At the same time, lexicographic take-the-best heuristics can be explained well with the use of bibliometrics reports on single scientists in research evaluation. Suppose the report includes the results of many analyses concerning productivity, citation impact, collaboration, and theoretical roots. Because the decision makers are interested in an outstanding scientist (most important cue) rooted in a specific theoretical tradition (second most important cue) with frequent international collaborations (third most important cue), the decision makers select the scientists with the most papers belonging to the 1% most frequently cited within their field and publication year ($P_{top\ 1\%}$, targeting the most important cue). Since two scientists perform similarly, the decision makers select the scientist who is active in the desired theoretical tradition (and reject the other who is not). Because the consideration of the theoretical tradition allows to discriminate between the scientists, the selection process finishes (and does not consider the third aspect, international collaborations).

Decision trees are processes that can be described in terms of tree-structured decision rules (Martignon et al., 2011). Decision trees consists of three elements: (a) *Nodes* represent cue-based questions, (b) *branches* represent answers to these questions, and (c) *exits* represent decisions and leavings of the tree (Phillips, Neth, Woike, & Gaissmaier, 2017).[2] Decision trees are always lexicographic; once a decision has been made based on a certain piece of information, no other information is used in the decision-making process. In the example given above (the selection of a single scientist), no indicators are used besides citation impact and publication output (Phillips et al., 2017). Fast-and-frugal decision trees (FFTs)—as a subgroup of decision trees—are defined by Martignon, Katsikopoulos, and Woike (2008) as trees that have exactly two branches from one node, whereby one or both branches lead to exits.

The root node is the starting point in the use of decision trees. Various further levels follow at which one cue is processed at each level. Two types of nodes exist: (a) the first type is branch oriented. A node contains a question about the evaluated case. The answer then leads

---

[2] Phillips et al. (2017) define decision trees as "supervised learning algorithms used to solve binary classification tasks." In this study, another kind of decision tree is proposed that does not accord with this definition.

to another node at further levels. (b) The second type is exit-oriented. The evaluated case is categorized and the decision process stops. In principle, decision trees may consist of dozens of nodes within a complex network of branches, which might be complicated to read. Katsikopoulos (2011) recommends therefore that "for trees to be easy for people to understand and apply, they should not have too many levels, nodes, or attributes" (p. 13).

Because decision trees are visualizations with clearly understandable rules for application, they are useful tools for situations in which application errors must be avoided or where it is important that all stakeholders are aware of the decision process (e.g., candidates for tenure should know, in advance, on what dimensions and how they will be evaluated). Along the same lines, one might also argue that decision trees are useful tools for making decisions that come with important consequences. Finally, trees are also particularly suitable for situations of time pressure, because they simplify decision processes and hence allow speeding them up. For FFTs, specific software has been developed—written in the open-source R language—for creating, visualizing, and evaluating decision trees (Phillips et al., 2017).

According to Phillips et al. (2017), FFTs have three important advantages: "First, FFTs tend to be both fast and frugal as they typically use very little information … FFTs are heuristics by virtue of ignoring information … Second, FFTs are simple and transparent, allowing anyone to easily understand and use them … FFTs can make good predictions even on the basis of a small amount of noisy data because they are relatively robust against a statistical problem known as overfitting" (p. 347). As is usual for heuristics in general, decision trees provide an adaptive toolbox whereby each decision tree is appropriate for a given evaluation situation. Thus, it is necessary to specify for each decision tree the relevant situations, which means the environments in which it allows successful decisions (Marewski et al., 2010).

4. METHODS

The BBDT exemplar presented in this study has been developed based on literature overviews of bibliometrics (e.g., de Bellis, 2009; Todeschini & Baccini, 2016; van Raan, 2004; Vinkler, 2010). The author of this paper works as professional bibliometrician in research evaluation whereby the development of the BBDT is based on his practical experiences. He also has extensively published standards of good practice for using bibliometric data in research evaluation (e.g., Bornmann, in press; Bornmann et al., 2014; Bornmann & Marx, 2014; Bornmann, Mutz, Neuhaus, & Daniel, 2008). Some of these standards have been used to develop the decision tree. It is the general idea of the BBDT to guide the use of bibliometric indicators in specific situations of research evaluation.

5. RESULTS

The BBDT presented in the following is a tool for making decisions in bibliometrics-based research evaluations. The development of the BBDT followed the key question raised by Phillips et al. (2017) in the context of decision making: "How to make good classifications, and ultimately good decisions, based on cue information" (p. 345)? The BBDT is prescriptive and oriented toward ecological rationality (see Marewski et al., 2010). In the process of developing the BBDT, the experience has been that this process was not only interesting in view of the application by the later user but was also interesting for the developer, because he had to think about the evaluation situation, available indicators, evaluation goals, etc. These aspects are usually not considered in processes of new indicator developments in scientometrics, because they focus on technical improvements of indicators.

The BBDT focuses on the use of citation impact indicators in research evaluation. Much research in bibliometrics deals with the development of field-normalized indicators (Waltman, 2016). These indicators have been introduced because citation rates are dependent on not only the quality of papers but also field-specific publication and citation practices (Bornmann, in press). By considering expected field-specific citation rates as reference standards, field-normalized citation indicators provide information on citation impact that is field-specifically standardized and can be used in cross-field comparisons (e.g., for the comparison of different countries). For example, the $PP_{top\ 10\%}$ indicator—the recommended field-normalized indicator for institutional citation impact measurements (Hicks et al., 2015; Waltman et al., 2012)—is the proportion of papers (published by an institution) that belong to the 10% most frequently cited papers in the corresponding subject categories (and publication years). Since many units in science publish research in different fields, field-normalized indicators are very necessary. The most important disadvantages of field-normalized indicators are, however, their complexity (because they are more complex than simple citation rates, their results are more difficult to interpret) and their lost link to the underlying citation impact data (the field-normalized data can be different from the citation data that can be found in the Scopus or WoS databases).

In a recent conference paper, Waltman (2018) argues for using the difference between the micro- and macro levels to decide whether field-normalized indicators should be used or not. Only field-normalized indicators would have the necessary validity to be used at the macro level on which experts view the world exclusively through indicators. The most important validity criterion mentioned by Waltman (2018) is the question of whether the units are active in multiple fields (or not). This question focuses on not only the research of different units but also the research within the units. The use of simple citation rates for research evaluation might lead to distorted world views if units are active in multiple fields. For example, universities with a focus on biomedical research might have an advantage in research impact measurements against universities with other focuses (e.g., engineering and technology), simply because they are active in fields with high publication activity and—on average—many cited references listed in the papers (and not because of their high quality of papers).

Figure 1 visualizes a BBDT which considers both aspects—(1) micro/macro level and (2) research orientation in single or multiple fields—to decide whether field-normalized or nonnormalized indicators should be used in a research evaluation situation. Because we can assume that evaluations at the country and university levels always target research which has been done in multiple fields, field-normalized indicators should be used in all situations (based on multidisciplinary classification systems such as journal sets proposed by Clarivate Analytics or Elsevier). Research-focused institutions and single researchers can be active in single or multiple fields. Some research-focused institutions are specialized in research topics of certain fields (e.g., the European Molecular Biology Laboratory, EMBL), and others have a broader research spectrum. The same is true with single scientists—to a limited extent—since scientists are as a rule focused on research in a single field. Thus, for both units the decision in the concrete evaluation situation must be made on whether the focus is on one field only or on several fields.

In Figure 1, another distinction is made between research (at research-focused institutions or by single researchers) that is done in multiple subfields or not (given that these units do research in only one field). For example, economists can be active in various subfields, such as financial economics or industrial organization (see Bornmann & Wohlrabe, 2019). These subfields can be concerned by different citation rates, why Bornmann, Marx, and Barth (2013) propose to consider these differences in research evaluation by calculating subfield-normalized citation rates (see Narin, 1987). In recent years, subfield-normalized citation rates have been calculated based on the following monodisciplinary classification systems: Medical Subject Headings (MeSH)
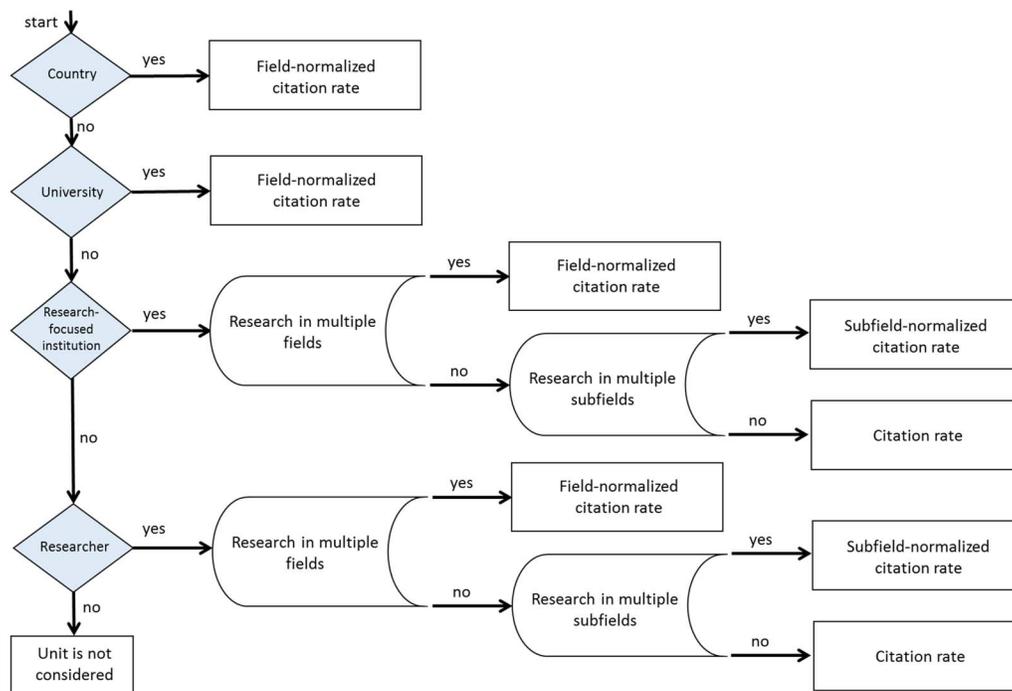
Figure 1. Decision tree (BBDT) for selecting (sub-)field-normalized or nonnormalized citation impact indicators in an evaluation situation.

(Boyack, 2004), Physics and Astronomy Classification Scheme (PACS) (Radicchi & Castellano, 2011), sections of the American Sociological Association (ASA) (Teplitskiy & Bakanic, 2016), and Journal of Economic Literature (JEL) codes (Bornmann & Wohlrabe, 2019).

6. DISCUSSION

The "royal road" in research evaluation can be summarized as follows: Judgments are based on peer review processes that include complete search and processing of information in decision-making. All information about a unit (e.g., an institution or a research group) is made available to decision makers who use all the information to make a preliminary recommendation or final decision (e.g., on funding or hiring). The information is usually weighted according to its predictive value for the evaluation task. The problem with these processes is, however, that they occasion high costs and absorb the valuable time of evaluating researchers, reviewers, and decision makers. For example, the 2014 UK REF panel for physics consisted of 20 members. According to Pride and Knoth (2018), 6,446 papers have been submitted as outputs from the universities. Because each paper should be read by two reviewers—which increases the number of papers for reading to 6,446 × 2 = 12,892 paper instances—more than 600 papers had to be read on average within less than one year.

According to Hertwig and Todd (2003), the tendency to use complex procedures including all information seems to follow a certain belief: "Scientific theorizing, visions of rationality and common wisdom alike appear to share a mutual belief: the more information that is used and the more it is processed, the better (or more rational) the choice, judgment or decision will be" (p. 220). However, the successful actions of people in daily life question this belief: "On a daily basis, people make both important and inconsequential decisions under uncertainty, often under time pressure with limited information and cognitive resources. The fascinating

phenomenon is that most of the time, a majority of people operate surprisingly well despite not meeting all requirements of optimization, be they internal (calculative abilities) or external (information access)" (Mousavi & Gigerenzer, 2017). Hertwig and Todd (2003) conclude therefore that "making good decisions need not rely on the standard rational approach of collecting all available information and combining it according to the relative importance of each cue—simply betting on one good reason, even one selected at random, can provide a competitive level of accuracy in a variety of environments" (p. 223).

Bibliometrics combines methods and data that can be used to make performance decisions in science by focusing on only a part of the available information. Since the introduction of bibliometrics decades ago, it has become more and more popular in research evaluation. For example, the results of Hammarfelt and Haddow (2018) show that about one-third of their respondents in a survey stated that "they had used metrics for evaluation or self-promotion in applications and CVs" (p. 927). This large share of respondents is a surprising result, since they identified themselves as being in the humanities, where the missing coverage of the literature in bibliometrics databases makes the use of bibliometrics indicators problematic. It seems that even in environments in which the use of bibliometrics is highly problematic, its use is popular.

Because bibliometrics is based on partial and nonrandom ignorance of other data or indicators and can be applied in a time-efficient and effortless way, Bornmann and Marewski (2019) made a connection between bibliometrics and heuristics and introduced BBHs. Heuristics are defined as "simple decision-making strategies, also called 'rules of thumb', that make use of less than complete information … more and more researchers are beginning to realise, especially in fundamentally uncertain domains such as medicine, that expertise and good decision making involve the ignoring of some information" (Wegwarth, Gaissmaier, & Gigerenzer, 2009, p. 722). The heuristics research program introduced by Gigerenzer, Todd, and ABC Research Group (1999) has already studied the use of heuristics in various fields, including psychology, demography, economics, health, transportation, and biology. The program is based on the bounded rationality view by Simon (1956, 1990), who argues that people use simple strategies in situations where resources are sparse. Simon's view of problem solving is known as satisfying: People search for real-world solutions and avoid complex solutions (which are time consuming and difficult to apply; see Tuccio, 2011).

BBHs are decision strategies that are solely based on publication and citation data. These strategies ignore information about performance (e.g., amount of third-party funds raised, assessments of single publications by experts), which allows quick decisions in research evaluation. An "ideal" BBH is an empirically validated strategy for performance decisions on certain units (e.g., researchers or papers) in a specific evaluation environment using a clearly defined set of bibliometric indicators integrated in certain (search, stopping, and decision) rules. The introduction of BBHs by Bornmann and Marewski (2019) should not be understood as a general push for using bibliometrics in research evaluation. In contrast, it is a call to investigate the evaluative use of bibliometrics more extensively. Answers are needed on the following (and similar) questions: In which situations is it reasonable to use bibliometrics? When should bibliometrics be combined with peer review? Are there situations in which bibliometrics should be avoided? For example, if one has identified situations in which bibliometrics comes to the same results as peer review, one might think about the replacement of peer review by bibliometrics (e.g., in the UK REF). It does not help to demonize or push the use of bibliometrics in general. We need research that shows in which situations it is useful and in which not.

This study introduces an exemplar BBDT that can be used in specific research evaluation situations. Decision trees are prototypical noncompensatory algorithms that are applied

sequentially until a final decision is reached. They are graphical representations of a set of rules that operate as lexicographic classifiers (see Martignon et al., 2011). Decision trees consist of nodes (cue-based questions), branches (answers to questions), and exits (decisions) (Phillips et al., 2017). According to Luan and Reb (2017), empirical studies in many domains have shown that people often decide with noncompensatory strategies that are similar to decision trees. One reason might be that compensatory strategies work better in "small world" situations in which much information is known and everything is calculable. However, many decisions (especially in research evaluation) are not "small world" situations but "are characterized by unclear utilities, unknown probabilities, and often multiple goals. These conditions severely restrict the effectiveness of compensatory strategies in finding the optimal solutions" (Luan & Reb, 2017, p. 31). "Ideal" BBDTs are appropriate visualizations of one or several "ideal" BBHs (as explained above).

The BBDT presented in this study is practically oriented for deciding, in specific evaluation situations, which indicator should be used. The decision tree can be characterized as a kind of checklist, which is—according to Tuccio (2011)—"the simplest form of a heuristic as they specify a procedure or rule" (p. 42). The proposed BBDT can be used as guidance to find an answer to the question of in which situations simple indicators such as mean citation rates are reasonable and in which situations more elaborate indicators (normalized indicators on the field or subfield level) should be applied.

Most BBDTs are valid only for a certain time period. New bibliometrics indicators are proposed that improve established indicators, and more appropriate statistical methods are proposed for analyzing bibliometric data. Thus, the improvement of BBDTs is an ongoing task that should involve as many professional bibliometricians (and scientists concerned) as possible (Marewski et al., 2010). In my opinion, the generation and continuous revision of BBDTs could be handled by the International Society for Informetrics and Scientometrics (ISSI)—the international association of scholars and professionals active in the interdisciplinary study science of science, science communication, and science policy (see www.issi-society.org). ISSI could implement BBDTs in software tools with well-designed interactive user interfaces that guide the user through the different choices that need to be made.

To be clear, BBDTs are not introduced in this paper as new tools for application in "desktop bibliometrics." As outlined in section 1, the term *desktop bibliometrics* describes the application of bibliometrics by decision makers (e.g., deans or administrators) with the help of "click the button tools" without involving experts in scientometrics and the evaluated fields. In general, it should be the goal to include these experts in the processes of developing and establishing BBDTs to decide what is to be assessed, what analyses make sense, and what data sources should be used. In principle, both groups could also be involved in the interpretation of the results of BBDTs. BBDTs are intended to structure available standards in the field of scientometrics and to facilitate the application of these standards.

Because bibliometrics-based research evaluations refer to a broad spectrum of data, analyses, and tasks, further BBDTs should be developed for the practical application of bibliometrics (or BBHs). For instance, BBDTs could be developed that help a user choose which bibliometric data source (WoS, Scopus, Dimensions, PubMed, Google Scholar, Microsoft Academic, Crossref, etc.) to work with. Another option is to think of a BBDT that helps a user choose, based on the aim one has, the type of unit one is focusing on, and the field in which this unit is active, how to carry out a research evaluation (e.g., based only on bibliometrics, based only on peer review, or based on some combination of the two). A third option could be a BBDT that helps a user to choose a bibliometric indicator for evaluating the impact of

journals (2-year journal impact factor, 5-year journal impact factor, Eigenfactor, Article Influence Score, CiteScore, Source Normalized Impact per Paper, SCImago Journal Rank, etc.). A fourth option could be a BBDT for the choice of indicators that can be used for the evaluation of individual researchers (see Bornmann & Marx, 2014).

DATA AVAILABILITY

Not applicable.

REFERENCES

Bornmann, L. (in press). Bibliometric indicators—methods for measuring science. In R. Williams (Ed.), *Encyclopedia of Research Methods*. Thousand Oaks, CA, USA: Sage.

Bornmann, L., Bowman, B. F., Bauer, J., Marx, W., Schier, H., & Palzenberger, M. (2014). Bibliometric standards for evaluating research institutes in the natural sciences. In B. Cronin & C. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (pp. 201–223). Cambridge, MA, USA: MIT Press.

Bornmann, L., Hug, S., & Marewski, J. N. (2019). Bibliometrics-based heuristics: What is their definition and how can they be studied? Retrieved November 5, 2019, from https://arxiv.org/abs/1810.13005

Bornmann, L., & Marewski, J. N. (2019). Heuristics as conceptual lens for understanding and studying the usage of bibliometrics in research evaluation. *Scientometrics*, *120*(2), 419–459.

Bornmann, L., & Marx, W. (2014). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. *Scientometrics*, *98*(1), 487–509. https://doi.org/10.1007/s11192-013-1161-y

Bornmann, L., Marx, W., & Barth, A. (2013). The normalization of citation counts based on classification systems. *Publications*, *1*(2), 78–86.

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Use of citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, *8*, 93–102. https://doi.org/10.3354/esep00084

Bornmann, L., & Wohlrabe, K. (2019). Normalisation of citation impact in economics. *Scientometrics*, *120*(2), 841–884. https://doi.org/10.1007/s11192-019-03140-w

Boyack, K. W. (2004). Mapping knowledge domains: Characterizing PNAS. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 5192–5199.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & ABC Research Group. (Eds.), *Simple heuristics that make us smart* (pp. 97–118). Oxford, UK: Oxford University Press.

de Bellis, N. (2009). *Bibliometrics and citation analysis: From the Science Citation Index to cybermetrics*. Lanham, MD, USA: Scarecrow Press.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295x.103.4.650

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford, UK: Oxford University Press.

Hammarfelt, B., & Haddow, G. (2018). Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators. *Journal of the Association for Information Science and Technology*, *69*(7), 924–935. https://doi.org/10.1002/asi.24043

Harzing, A.-W. (2017). Running the REF on a rainy Sunday afternoon: Do metrics match peer review? Retrieved August 5, 2018, from https://harzing.com/publications/white-papers/running-the-ref-on-a-rainy-sunday-afternoon-do-metrics-match-peer-review; https://openaccess.leidenuniv.nl/handle/1887/65202

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (pp. 213–231). Hoboken, NY; USA: Wiley.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431.

Jappe, A., Pithan, D., & Heinze, T. (2018). Does bibliometric research confer legitimacy to research assessment practice? A sociological study of reputational control, 1972–2016. *PLOS ONE*, *13*(6), e0199031. https://doi.org/10.1371/journal.pone.0199031

Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice. *Decision Analysis*, 8(1), 10–29. https://doi.org/10.1287/deca.1100.0191

Kelman, M. (2011). *The heuristics debate*. Oxford, UK: Oxford University Press.

Kurz-Milcke, E., & Gigerenzer, G. (2007). Heuristic decision making. *Marketing*, 3(1), 48–56.

Leydesdorff, L., Wouters, P., & Bornmann, L. (2016). Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators – a state-of-the-art report. *Scientometrics*, 109(3), 2129–2150. https://doi.org/10.1007/s11192-016-2150-8

Luan, S. H., & Reb, J. (2017). Fast-and-frugal trees as noncompensatory models of performance-based personnel decisions. *Organizational Behavior and Human Decision Processes*, 141, 29–42. https://doi.org/10.1016/j.obhdp.2017.05.003

Marewski, J. N., Schooler, L. J., & Gigerenzer, G. (2010). Five principles for studying people's use of heuristics. *Acta Psychologica Sinica*, 42(1), 72–87. https://doi.org/10.3724/SP.J.1041.2010.00072

Martignon, L., Katsikopoulos, K. V., & Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, 52(6), 352–361. https://doi.org/10.1016/j.jmp.2008.04.003

Martignon, L., Vitouch, O., Takezawa, M., & Forster, M. (2011). Naive and yet enlightened: From natural frequencies to fast and frugal decision trees. In G. Gigerenzer, R. Hertwig, & T. Pachur (Eds.), *Heuristics: The Foundations of Adaptive Behavior* (pp. 4–15). Oxford, UK: Oxford University Press.

Merton, R. K. (1965). *On the shoulders of giants*. New York, NY, USA: Free Press.

Moed, H. F. (2017). *Applied evaluative informetrics*. Heidelberg, Germany: Springer.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14(3), 131–149.

Mousavi, S., & Gigerenzer, G. (2017). Heuristics are Tools for Uncertainty. *Homo Oeconomicus*, 34(4), 361–379. https://doi.org/10.1007/s41412-017-0058-z

Narin, F. (1987). Bibliometric techniques in the evaluation of research programs. *Science and Public Policy*, 14(2), 99–106. https://doi.org/10.1093/spp/14.2.99

Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics*, 4(3), 423–430.

Phillips, N. D., Neth, H., Woike, J. K., & Gaissmaier, W. (2017). FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgment and Decision Making*, 12(4), 344–368.

Pride, D., & Knoth, P. (2018). Peer review and citation data in predicting university rankings, a large-scale analysis. In E. Méndez, F. Crestani, C. Ribeiro, G. David, & J. Lopes (Eds.), *Digital Libraries for Open Knowledge. TPDL 2018*. Lecture Notes in Computer Science, vol. 11057. (pp. 195–207). Basel, Switzerland: Springer.

Raab, M., & Gigerenzer, G. (2015). The power of simplicity: A fast-and-frugal heuristics approach to performance science. *Frontiers in Psychology*, 6. https://doi.org/10.3389/fpsyg.2015.01672

Radicchi, F., & Castellano, C. (2011). Rescaling citations of publications in physics. *Physical Review E*, 83(4). https://doi.org/10.1103/PhysRevE.83.046116

Robinson-Garcia, N., Torres-Salinas, D., Herrera-Viedma, E., & Docampo, D. (2019). Mining university rankings: Publication output and citation impact as their basis. *Research Evaluation*, 28, 232–240. https://doi.org/10.1093/reseval/rvz014

Rodriguez Navarro, A., & Brito, R. (2019). Like-for-like bibliometric substitutes for peer review: Advantages and limits of indicators calculated from the ep index. Retrieved June 26, 2019, from https://arxiv.org/abs/1903.11119

Scheibehenne, B., & von Helversen, B. (2009). Useful heuristics. In T. Williams, K. Samset, & K. J. Sunnevåg (Eds.), *Making Essential Choices with Scant Information; Front-end Decision Making in Major Projects* (pp. 194–211). Basingstoke, UK: Palgrave Macmillan.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138. https://doi.org/10.1037/h0042769

Simon, H. A. (1990). Invariants of human-behavior. *Annual Review of Psychology*, 41, 1–19. https://doi.org/10.1146/annurev.ps.41.020190.000245

Teplitskiy, M., & Bakanic, V. (2016). Do peer reviews predict impact? Evidence from the *American Sociological Review*, 1978 to 1982. *Socius*, 2, 1–13. https://doi.org/10.1177/2378023116640278

Todd, P. M., Gigerenzer, G., & ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York, NY, USA: Oxford University Press.

Todeschini, R., & Baccini, A. (2016). *Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research*. Weinheim, Germany: Wiley.

Traag, V. A., & Waltman, L. (2019). Systematic analysis of agreement between metrics and peer review in the UK REF. *Palgrave Communications*, 5(1).

Tuccio, W. A. (2011). Heuristics to improve human factors performance in aviation. *Journal of Aviation/Aerospace Education & Research*, 20(3), 39–53.

van Raan, A. F. J. (2004). Measuring science. Capita selecta of current main issues. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems* (pp. 19–50). Dordrecht, The Netherlands: Kluwer Academic Publishers.

van Raan, A. F. J., van Leeuwen, T. N., Visser, M. S., van Eck, N. J., & Waltman, L. (2010). Rivals for the crown: Reply to Opthof and Leydesdorff. *Journal of Informetrics*, 4, 431–435.

Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Oxford, UK: Chandos Publishing.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391.

Waltman, L. (2018). Responsible metrics: One size doesn't fit all. In P. Wouters (Ed.), *Proceedings of the Science and Technology Indicators Conference 2018 in Leiden "Science, Technology and Innovation Indicators in Transition"* (pp. 526–531). Leiden, The Netherlands: University of Leiden.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., … Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419–2432.

Wegwarth, O., Gaissmaier, W., & Gigerenzer, G. (2009). Smart strategies for doctors and doctors-in-training: Heuristics in medicine. *Medical Education*, 43(8), 721–728. https://doi.org/10.1111/j.1365-2923.2009.03359.x