Journal of Business Research 133 (2021) 34–50

Contents lists available at ScienceDirect

Journal of Business Research

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

Big data and human resource management research: An integrative review and new directions for future research

Yucheng Zhang, Shan Xu, Long Zhang *, Mengxi Yang

School of Public Administration, Southwestern University of Finance and Economics, Chengdu, China
School of Business Administration, Southwestern University of Finance and Economics, Chengdu, China
Business School, Hunan University, Changsha, China
School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:
Human resource management research
Big data
Integrative review
Inductive and deductive paradigms

ABSTRACT

The lack of sufficient big data-based approaches impedes the development of human resource management (HRM) research and practices. Although scholars have realized the importance of applying a big data approach to HRM research, clear guidance is lacking regarding how to integrate the two. Using a clustering algorithm based on the big data research paradigm, we first conduct a bibliometric review to quantitatively assess and scientifically map the evolution of the current big data HRM literature. Based on this systematic review, we propose a general theoretical framework described as "Inductive (Prediction paradigm: Data mining/Theory building) *vs.* Deductive (Explanation paradigm: Theory testing)". In this framework, we discuss potential research questions, their corresponding levels of analysis, relevant methods, data sources and software. We then summarize the general procedures for conducting big data research within HRM research. Finally, we propose a future agenda for applying big data approaches to HRM research and identify five promising HRM research topics at the micro, meso and macro levels along with three challenges and limitations that HRM scholars may face in the era of big data.

1. Introduction

Big data, as an emerging field of research and a practical method, has stimulated an extensive discussion over the past few years (Mcafee & Brynjolfsson, 2012). The fast-growing technology of the Internet provides new data collection channels for various business and management research areas, including information systems (Schermann et al., 2014), management science (Waller & Fawcett, 2013), healthcare (Wang & Hajli, 2017), marketing (Erevelles et al., 2016), and finance (Fang & Zhang, 2016). A big data approach is a "must have" capability for both human resource management (HRM) researchers and practitioners (Angrave et al., 2016; Shah et al., 2017) and creates an interdisciplinary opportunity for HRM researchers to solve research puzzles that cannot be addressed by traditional sample-based data (Edelman, 2012; George et al., 2014; Groves et al., 2013; Mcafee & Brynjolfsson, 2012). For example, machine learning, especially deep learning via human neural networks, enables more granular and accurate HRM decisions (Blazquez et al., 2018).

However, HRM research has not sufficiently exploited the big data approach (Jiang & Messersmith, in press). The sample-based small data approach still dominates current HRM studies, whereas the advancement of a big data approach (e.g., the velocity or real-time approach) has not been effectively incorporated into the current HRM research paradigm. HRM research faces three major problems and challenges during the paradigm shift from small to big data research.

The first challenge is a methodological problem, that is, how to access HR big data in the workplace and use it to examine academic HRM issues. The paradigm shift from big data includes more competitive data resources and real-time strategic decision making with the technology to process large volumes of data (Tonidandel et al., 2018). Traditional qualitative research methods (such as interviews) and quantitative research methods (such as surveys) have dominated the HRM and industrial psychology research, where most data are either numerical data (from questionnaires) or textual data (derived from interviews) (George et al., 2016). However, with the development of big data technology, other types of data—such as images, videos, and audio generated from mobile phones and social mobile applications—can be collected to address important HRM research questions. For example, research has

* Corresponding author at: Business School, Hunan University, Changsha, China.
E-mail address: lyon.long.zhang@gmail.com (L. Zhang).

https://doi.org/10.1016/j.jbusres.2021.04.019
Received 11 March 2020; Received in revised form 6 April 2021; Accepted 8 April 2021
Available online 5 May 2021

proposed that scholars can better understand employees' behavior and examine how they interact with one another by using geolocation data generated from mobile devices (Haak, 2014).

The second challenge is a technological problem. In practice, leading organizations currently fully use social networking, analytics and even cognitive tools to acquire talent in new ways (Schwartz et al., 2016). In addition, as entire organizations have digitized their records, HRM deploys digital workplaces and workforces (Angrave et al., 2016). However, data science analytic technologies are unfamiliar to HRM researchers; thus, they need to learn new computational skills to manage big data analytical platforms, such as Hadoop, Python, R, and MATLAB (Douthitt & Mondore, 2014).

The third problem is the conventional dichotomy between the deductive and inductive paradigms in HRM research. Traditional small data HRM research is mainly based on a deductive paradigm (Popper, 2005; Zhang et al., 2016, 2017). Although big data offers opportunities for deductive research, big data requires representations of theorization that are more flexible. The emergence of multiple non-structured data instances from big datasets provides opportunities for research based on the inductive paradigm. In contrast to deductive research, most inductive research is data-driven and requires HRM researchers to adopt "reverse thinking" (Mcabee et al., 2017). For example, based on an inductive framework, Chakraborty et al. (2018) explored ways to improve business schools' Master of Business Administration candidate selection by using classification trees and artificial neural networks (ANNs).

In facing these challenges, adopting a big data approach has become increasingly important. In particular, we suggest that three main research gaps in big data HRM research offer future research opportunities and thus deserve more attention. First, big data science techniques provide HRM scholars with greater scope and granularity and more timely and precise results (George et al., 2016; Van Der Vegt et al., 2015). As a result, HRM studies can investigate new research questions that would not otherwise be investigated by using traditional small data analysis (George et al., 2016); additionally, the higher precision in estimating effects could unveil nuances in the effects that have not yet been adequately examined in HRM research. In practice, big data analysis enables predictions that are more precise for the benefit of future HR decision making (Hastie et al., 2009; Varian, 2014).

Second, big data enable HRM researchers to measure factors dynamically to thus establish clearer causal mechanisms (George et al., 2014, 2016; Landers et al., 2016). Therefore, detailed data and further implications instead of purely relational inferences in HRM research are obtainable with big data analysis (Harlow & Oswald, 2016).

Third, by adopting a critical perspective, incorporating big data analysis can supplement the current mainstream approach, enable HRM researchers to reintroduce human elements to HRM, and empower HRM researchers to understand more comprehensively the everyday nature of employees and companies from a broader and more humanized perspective. The current research paradigm has long been criticized as being mechanistic and lacking humanity to some extent (Alvesson & Deetz, 2000). In recent HRM research, scholars have recognized the problem of ignoring the human factor (Braun et al., 2018; Wright & Mcmahan, 2011). Current small data research faces difficulties in capturing the nuances (such as the interpersonal interactions or the everydayness of organizational activities) in human elements (Cheung, 2017). These problems, however, can be solved with the help of big data. For instance, by systematically analyzing employees' tones on social media, such as in Facebook posts, HRM researchers can study the patterns or the changes of the related human element nuances associated with these previously regarded "ambiguous" tones. Therefore, it is important for HRM research to embrace a big data approach.

HRM research is undergoing a paradigm shift from the traditional era of small data to the emerging era of big data. As this trend continues, HRM research will be informed by a significant re-examination and extension of previous research methods and findings (Shah et al., 2017; Sivarajah et al., 2017). In a recent review of the use of big data in HRM research, Mcabee et al. (2017) argued that "overreliance on deductive approaches at the expense of inductive approaches limits our understanding of organizational phenomena" and calls for "a greater integration of deductive and inductive approaches as necessary for the future growth of our field" (p. 278). Thus, this paper aims to help promote the integration of a big data approach with HRM research that is highly inclusive of both the deductive and inductive paradigms. We share Mcabee et al. (2017) belief that "the use of big data analytics offers just such an opportunity" (p. 278). Based on the social science big data research paradigms suggested by Hofman et al. (2017), we hereby propose an "Inductive (Prediction paradigm: Data mining/Theory building) *vs.* Deductive (Explanation paradigm: Theory testing)" research framework to integrate possible HRM research questions that can use big data and their corresponding levels of analysis, analytical methods, data sources and software (see Table 1).

The rest of this paper proceeds as follows (see Fig. 1). We first conduct a systematic review of the current big data HRM literature by using a bibliometric review to visualize and map the literature's developmental trajectory. Second, we review the characteristics of big data and big data research. Third, we compare the emerging big and traditional small data research approaches. Fourth, eleven big data analysis methods that can be applied to HRM research are discussed. Fifth, the general procedures of big data research are introduced. Finally, potential HRM research topics and challenges in the big data era are discussed. Table 1 presents the general framework of this review based on the "Inductive (Prediction paradigm: Data mining/Theory building) *vs.* Deductive (Explanation paradigm: Theory testing)" framework, which includes all potential research questions that we propose throughout this paper and their corresponding analysis levels, methods, data sources and software.

2. Literature review based on a bibliometric analysis of big data hrm research

Using the clustering algorithm based on the big data research paradigm, we first conduct a bibliometric analysis to scientifically visualize the landscape of the current big data HRM literature and to trace its evolution. The literature scope includes the research area distribution and the country co-authorship, institution co-authorship and journal co-citation networks.

We built the datasets for our bibliometric review in two steps (Chen, 2014; Kozlowski et al., 2017). First, we chose Web of Science as our database; Web of Science contains comprehensive academic information resources that comprise more than 8,700 core academic journals (Hanisch & Wald, 2012; Shafique, 2013). In Web of Science, we searched the terms "HRM" paired with "data science", "human resource management" paired with "data science", "human resource management" paired with "big data", "HR analy*", "human resource analy*", "HRM analy*", and "human resource management analy*" without limiting the time spans. To effectively collect related big data HRM research, we further refined the research areas by Web of Science categories that are relevant to HRM big data, such as business and management. In the second step, we scrutinized the titles, abstracts and keywords of each record and deleted all literature that is theoretically or empirically irrelevant to big data HRM. Finally, 84 primary articles and 3,789 secondary references were collected and used in the following analysis.

According to the annual trend chart (Fig. 2), the number of studies in the past 30 years has shown an unstable increasing trajectory. The vertical axis represents the number of publications, and the horizontal axis represents the year. There is a clear difference between the first 21 years of HR big data research (when only a few empirical studies were published each year) and the last 10 years (when the number of published empirical studies increased dramatically). There was a small peak in the publication of empirical studies in 2012, and then (except for a

Table 1
Theoretical framework.

Research paradigms	Research questions	Analysis levels	Methods	Data sources	Software and packages
Inductive (Prediction paradigm: Theory building/Data mining)	Employee selection and recruitment	Micro	Textual Analysis	Daily chat records, such as e-mail or short messaging service	Python: scikit-learn
	Employees' performance assessment	Micro	Data Pattern Extraction	Curriculum vitae data	R: Matrix
	Prediction of HR demand	Macro	Deep Learning	Video data	Python: PyTorch
	Organizational performance prediction	Macro	Classification Approach	Internal data from the company or company website	Python: scikit-learn
Deductive (Explanation paradigm: Theory testing)	Strategic management, succession planning, and organizational design	Macro	Event Study	Data from a social application	R: twitteR
	Cross-country comparative HRM research	Macro	SAM	Company's website or others	R: metaSEM
Hybrid (Both inductive and deductive)	How HRM influences employees' attitudes and behaviours	Micro	SEM Trees	Wearable sensors	R: semtree
	Team social interaction	Meso	Social Network	Social networking platforms, such as Facebook	R: Statnet
	Strategic HRM	Macro	Clustering	Internal data from the company or company website	Cluster Analysis of Sequences

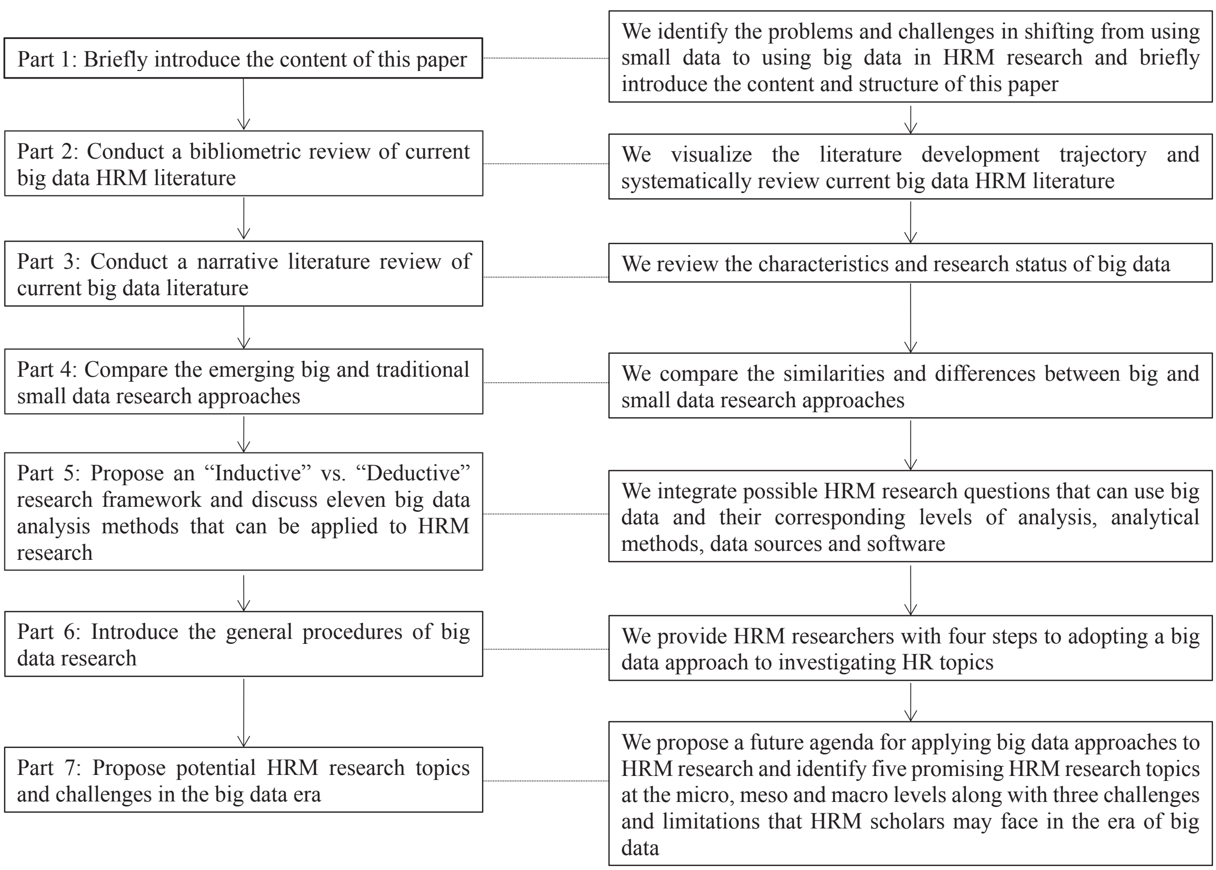

Fig. 1. Research summary.

slight decline in 2014), the number of published empirical studies continued to grow until 2017. Since then, the number of publications has been in a state of fluctuating decrease. In particular, in 2017 alone, a record number of 22 empirical HR big data studies were published. According to Fig. 1, although the amount of HR big data research has increased rapidly during the past three decades, this increase is unstable. Specifically, the number of publications in HR big data has decreased in the last three years.

Table 2, which shows the frequency and quality of the journal articles, illustrates that the research on HR big data has been published in different publication volumes in different journals. In Table 2, frequency represents the number of HR big data studies published in a journal. Journal quality is the quality of the journal as indicated by the Association of Business Schools (ABS) journal ranking, in which 4* represents the best quality on a scale of 1–4* (4 represents the next-best quality). Field refers to the field that corresponds to each journal; the journals were divided into the following eight categories: not available (NA) (journals in this category were not included in the ABS ranking); human resource management and employment studies (HRM&EMP); ethics and governance (ETHICS-CSR-MAN); international business and area studies (IB&AREA); information management (INFO MAN); operations and technology management (OPS&TECH); operations research and

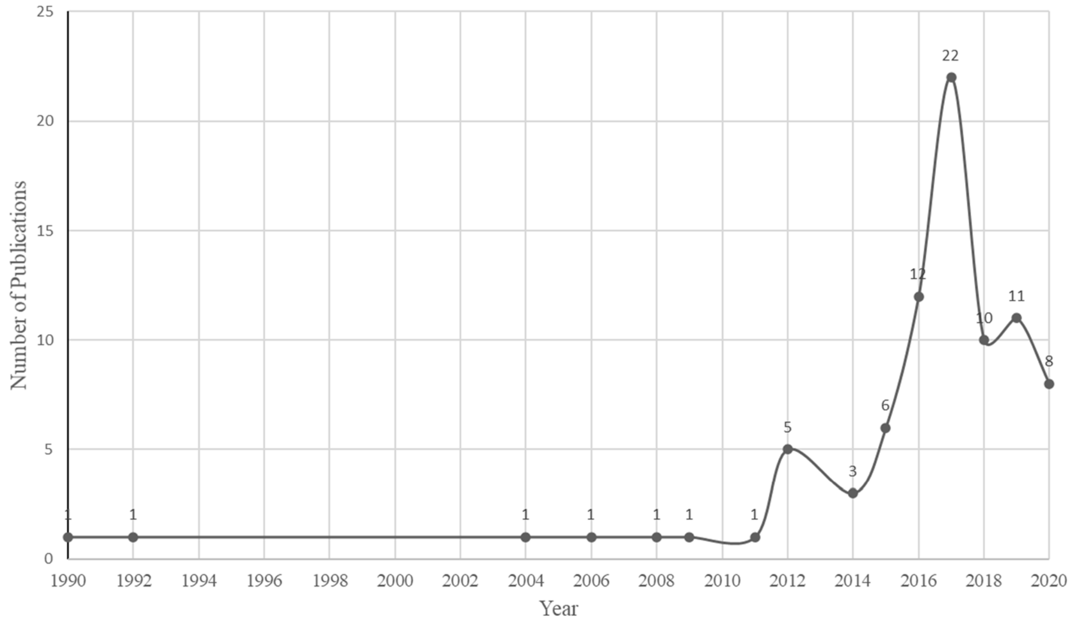

Fig. 2. The annual trends of HRM big data-related publications.

management science (OR&MANSCI); and organization studies (ORG STUD) (see the note in Table 1 for area details). Among the studies published in 2020 and earlier, big data HRM research is published mostly in journals in human resource management and employment studies, including the Journal of Organizational Effectiveness: People and Performance, Human Resource Management, Human Resource Management Review and the Human Resource Management Journal. The field of journals that publish the next-highest number of studies on big data HRM is ethics and governance; these journals include Business Horizons and Management Decision. The table shows that big data HR has received attention in these two fields, while other fields have less research on big data HR. The three journals that have published the most studies on big data HR research are the Journal of Organizational Effectiveness: People and Performance, the International Journal of Information Management, and Human Resource Management. Currently, most of the HR research that uses big data is published in journals ranked 2 or 3. Only one research result is published in a journal (Management Science) ranked 4*. In general, too little research on this topic is published in high-quality journals; this fact may explain why big data HR has not received wider attention.

Three types of bibliographic analyses—including the research area, co-authorship and co-citation—were applied in our research. The research area analyses described the distribution of the research areas of big data research in HRM. Co-authorship detected the relationship of the authors from different countries and institutions; this relationship is essential for achieving intellectual diversity through the complementary skillsets of authors (Liu et al., 2016). Co-citation analyses portrayed the relationship of different studies simultaneously cited by a third study (Chen et al., 2008; Chen, 2004, 2006), and we applied a journal co-citation analysis to detect the landscape of the journals in big data HRM research.

First, the results of the research analysis (Fig. 3) suggest that big data HRM research was distributed in 15 main research areas and was conducted mostly in two fields: business and economics; and management. The numbers of papers published in these two fields are almost the same, and together, they account for approximately 58.7% of the total; these fields were followed by computer science, theory and methods, psychology, applied and social sciences, and others.

Second, according to the country co-authorship analysis (Fig. 4), the United States (43.2%) and China (27.8%) published the most studies on big data HRM; these countries were followed by the United Kingdom, Italy, Germany, and others. The research cooperation situation in big data HRM is very interesting—surprisingly, such cooperation existed only among China, the United States, the United Kingdom, Italy, Belgium and Malaysia. Given the current intensification of globalization, cross-country big data HRM research cooperation is currently limited and requires further attention from HRM scholars globally. Furthermore, as shown in Fig. 5, institution co-authorship for the past 17 years was analyzed with VOSviewer. The results indicate that the top four institutions in big data HRM research publication quantity were the University of Southern California, the University of Texas at San Antonio, the State University of New York at Albany, and the University of Southampton. However, the dispersed pattern of the institute co-authorship network revealed very limited research communication or collaboration among these institutions; this finding, similar to the cross-country collaboration situation, is surprising and rare in the current "global village", where inter-institution research cooperation is prevalent. Thus, more cross-institutional collaborative efforts are also needed in future big data HRM research.

Third, the journal co-citation network (Fig. 6) indicates that in big data HRM research, the most influential academic resource in terms of citations is the Academy of Management Journal, followed by the Academy of Management Review, the Harvard Business Review, the Human Resource Management Journal, and Administrative Science Quarterly. Moreover, the journals in the network were classified into four clusters—HRM (in red), general management (in green), information systems and management science (in blue), and applied psychology (in yellow). The journals included in each cluster are also listed in Table 3. The results indicate that the HRM and general management journals showed the largest interest in big data HRM, whereas information systems and management science journals (which are assumed to have experience with big data research) and applied psychology journals (which are assumed to prefer empirical HRM) paid less attention to big data HRM.

Overall, our scientific mapping showed that on the one hand, the publication numbers of big data HRM research are relatively small and are distributed mainly in business and economics, and management, while other related fields lack academic attention. On the other hand, the big data HRM research network—including cross-country, institutional and discipline research collaboration—has yet to be established. According to these findings, big data HRM research remains an emerging research field that requires further development. However,

Table 2
The Quality of Journal articles.

Journal Title	Frequency	Field	Journal Quality
Journal of Organizational Effectiveness: People and Performance	7	HRM&EMP	2
International Journal of Information Management	5	INFO MAN	2
Human Resource Management	3	HRM&EMP	4
Business Horizons	2	ETHICS-CSR-MAN	2
Human Resource Management Journal	2	HRM&EMP	4
Human Resource Management Review	2	HRM&EMP	3
Management Decision	2	ETHICS-CSR-MAN	2
Personnel Psychology	2	PSYCH (WOP-OB)	4
Baltic Journal of Management	1	IB&AREA	1
Business Process Management Journal	1	OPS&TECH	2
European Journal of Information Systems	1	INFO MAN	3
Evidence-based HRM: A Global Forum for Empirical Scholarship	1	HRM&EMP	1
Human Resource Development Review	1	HRM&EMP	2
International Journal of Human Resource Management	1	HRM&EMP	3
International Journal of Management Reviews	1	ETHICS-CSR-MAN	3
Journal of General Management	1	ETHICS-CSR-MAN	2
Journal of Management Studies	1	ETHICS-CSR-MAN	4
Management Research Review	1	ETHICS-CSR-MAN	1
Management Science	1	OR&MANSCI	4*
MIT Sloan Management Review	1	ETHICS-CSR-MAN	3
Organizational Dynamics	1	ORG STUD	2
Personnel Review	1	HRM&EMP	2
Reliability Engineering and System Safety	1	OR&MANSCI	3
German Journal of Human Resource Management-Zeitschrift Fur Personalforschung	1	NA	NA
IBM Journal of Research and Development	1	NA	NA
Management-Journal of Contemporary Management Issues	1	NA	NA
Revista Empresa Y Humanismo	1	NA	NA
International Journal of Human Capital and Information Technology Professionals	1	NA	NA
Journal of Leadership Studies	1	NA	NA
Brq-Business Research Quarterly	1	NA	NA

Note: NA indicates the journal is not available from the ABS ranking; HRM&EMP represents Human Resource Management and Employment Studies; ETHICS-CSR-MAN represents Ethics and Governance; IB&AREA represents International Business and Area Studies; INFO MAN represents Information Management; OPS&TECH represents Operations and Technology Management; OR&MANSCI represents Operations Research and Management Science; ORG STUD represents Organization Studies; and PSYCH (WOP-OB) represents Psychology.

embracing big data is urgent and integral for HRM research. Therefore, this study aims to help integrate big data approaches into HRM research by systematically discussing the importance, benefits, characteristics, procedures and methods of big data research and the ways that big data approaches can be applied to HRM research.

3. Big data and big data research

Some researchers suspect that "there is no such thing as big data in HR" in practice because very few companies have a sufficiently large body of employees to justify using the special software and tools associated with big data (e.g., Cappelli, 2017, p. 2). HRM research, however, is not typically limited to a single company as its research object; therefore, HRM research is not limited by the number of a single firm's employees. Furthermore, from a multilevel perspective, even small numbers of individuals can generate large amounts of within-person-level data, such as wearable sensor data. More importantly, scholars (such as Angrave et al. (2016, p. 4)) have defined big HR data as "small by the standards of large unstructured data, but big by the standards of the quantitative data sets used in academic social science, and able to generate 'smart' insights by virtue of the longitudinal nature of the data". In this review, we adopt Angrave et al. (2016, p. 2) definition, which focuses on the "smartness" rather than purely on the "size" of the data in defining big data in the HR context. HRM researchers can conduct fine-grained studies with smart data so that HR professionals can better explain and predict employee behaviors (George et al., 2014).

Researchers have identified four "V" characteristics for big data, i.e., "Volume, Velocity, Variety and Veracity" (Tonidandel et al., 2018).[1] The first "V" represents the "big" volume of data (Chen & Wojcik, 2016; George et al., 2016). With a large volume of data, HRM researchers can collect data at the population level to thus solve the methodological problem of sample selection and biases (George et al., 2016). In the era of big data, increasing amounts of data on employers and employees can be collected by HRM researchers. For example, through web crawlers and wearable sociometric sensors (Chaffin et al., 2017), HRM researchers can continually access large databases on employee behavior and organizational performance (Angrave et al., 2016). For instance, based on an online data collection approach, Wang and Cotton (2018) collected a 111-year longitudinal dataset from Major League Baseball (MLB) to examine how experience ties impact team performance. Instead of relying on partial data during a certain period, this study obtained population-level MLB data, which provided robust estimations of the entire population.

The second "V" is data velocity, which refers to the speed with which data are generated and refreshed and the latency of data usage and analysis (Tonidandel et al., 2018). Velocity is particularly important for HRM research and HRM practices. For instance, real-time or nearly real-time information, such as social media (e.g., Twitter) feeds and a company's bulletin board system (BBS; Manyika et al.) postings, can reflect and affect employees' attitudes and behaviors and even a company's HRM policy. These real-time data allow HRM researchers to conduct time-series and causal analyses (Mueen, 2014). For example, through association rule learning, HRM researchers can use data from email exchanges to explore the relationship between pessimism in email exchanges and innovation (Wenzel & Van Quaquebeke, 2018).

The third "V" is data variety, which refers to "the plurality of structured and unstructured data sources such as text, videos, networks, and graphics among others" (George et al., 2016, p. 1493). Compared with traditional survey methods, multiple types of big data help HRM scholars improve the comprehensiveness of data. Various types of data can be generated from the use of online shopping, smart phones, the

[1] With continually increasing attention being given to big data, the number of "V"s in the definition of big data is also increasing. For example, variability, volatility, viscosity, visualization, value, validity and vulnerability, among others, are also argued to be included in the core characteristics of big data. However, this paper selects only the most well-accepted "4V" model of big data for two reasons. First, we believe that these four characteristics are the defining features that distinguish big from small data. Second, the number of "V"s in this concept can grow infinitely; therefore, a boundary must be drawn in our research.

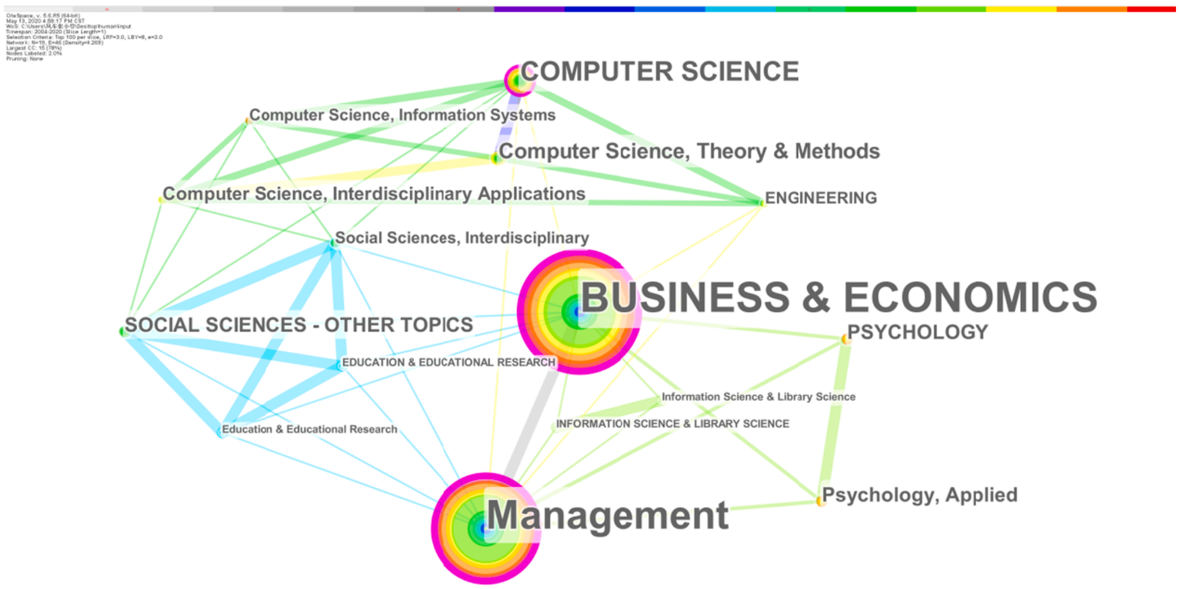

Fig. 3. Category Co-occurrence Network. Note: Each dot represents a research field, and the size of the dot represents the number of articles published in this field. The lines between the dots represent the relevance between two areas. The shape of each dot is similar to a tree ring, representing the years in which the articles were accumulated in this area.

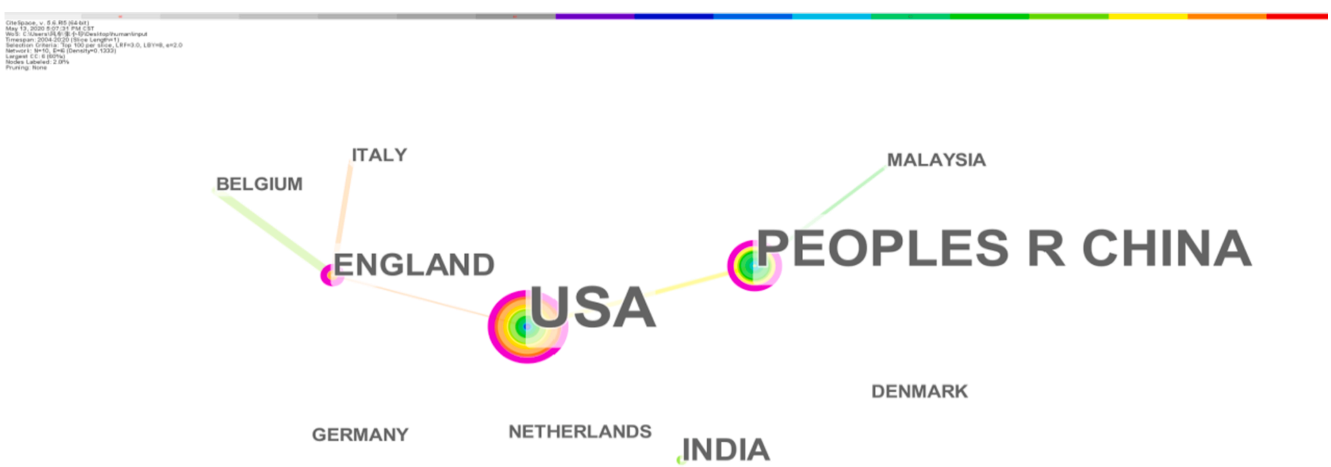

Fig. 4. Country Co-authorship Network. Note: Each dot represents a country, the size of which represents the number of articles published in the country, and the line between the dots represents the cooperation between different countries.

Global Positioning System (GPS), and social network sites (Chaffin et al., 2017). Compared with employing one type of data (as is typically done in traditional HRM research efforts), using data from various sources and methods enhances the robustness of findings. The increasing types of unstructured workplace data on employers and employees include data from email communications, office entry and exit records, radio-frequency identification tagging, wearable sociometric sensors, and phone calls (Chaffin et al., 2017). These data sources are being digitized and made accessible to HRM researchers and practitioners in the era of big data. In addition, through textual analyses, data from email exchanges and social mobile applications can be extracted and combined to measure and analyze employees' moods (Kobayashi et al., 2018); the findings may be helpful in solving the issues of social desirability and measurement errors in investigating how HRM practices affect employees' attitudes.

The fourth "V" is veracity, which refers to the accuracy and precision of the data (Chen & Wojcik, 2016; Gandomi & Haider, 2015). Compared with the traditional small data approach, a big data approach can largely improve data quality (Lazer et al., 2014). For example, the implicit information included in an employee's tone on a company's BBS or an employee's social media platform postings can be highly valuable to HRM researchers and managers, but in traditional survey research, this information is always contaminated because of individual subjectivity. However, this type of information and data can be analyzed through data mining with high precision and reliability. For example, Bogomolov and Lepri et al. (2014, p. 790) used a machine learning method to analyze data generated from employees' mobile phones—including call log, short messaging service and Bluetooth proximity data—to predict employee happiness with high accuracy (80.81%).

4. From small to big data

4.1. Similarities between big and small data research

Small data are used in sample-based studies that "do not have volume, velocity, and variety" characteristics (Chen & Wojcik, 2016, p. 7). Despite the recent prevalence of big data analyses, small data analyses still play a dominant and indispensable role in HRM research (George

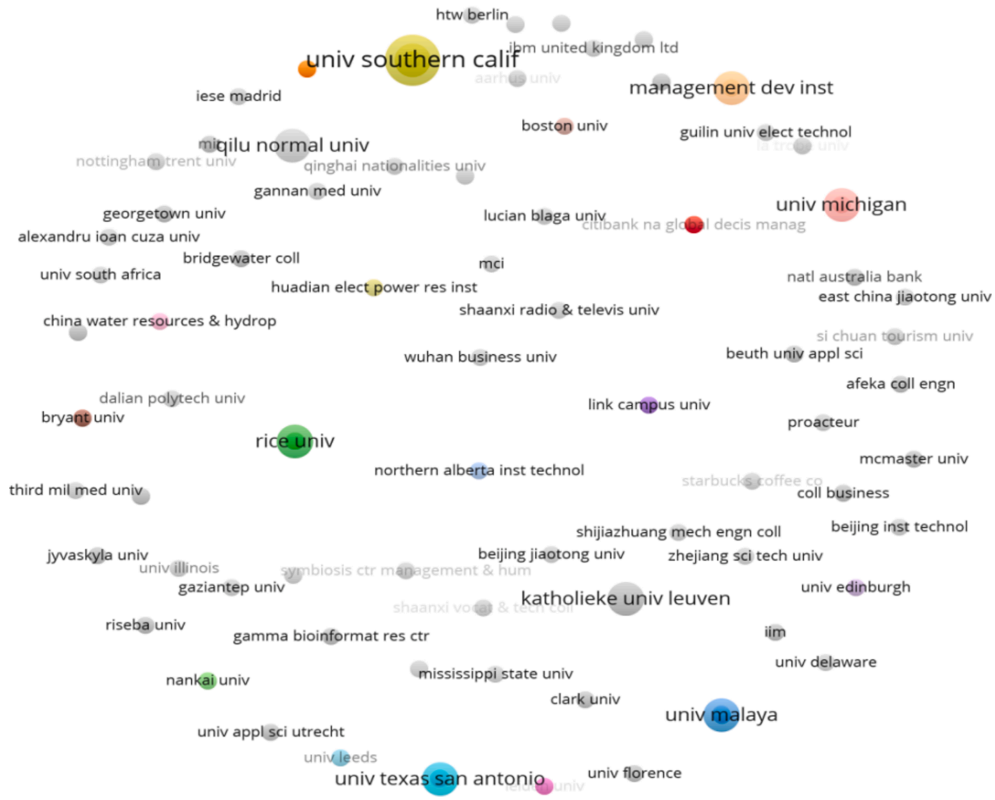

Fig. 5. Institute Co-authorship Network. Note: The dots represent institutes that conduct big data HRM research, and the point-to-point connections represent collaboration between them.

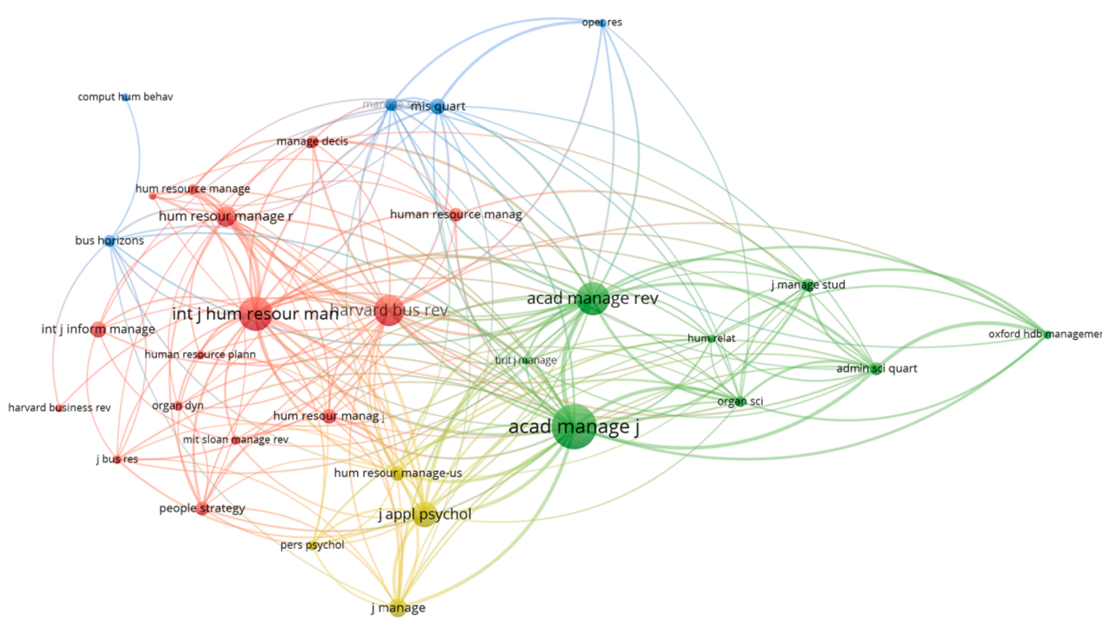

Fig. 6. Journal Co-citation Network. Note: The size of the dot represents the citations of each journal—the bigger the dot is, the more the journal has been cited. The lines between two dots represent the co-citation relationship of these two journals, which are cited simultaneously in the third journal. In addition, nodes with the same colour are placed in the same cluster.

et al., 2014). Big and small data studies have several similarities (Beaton et al., 2016; Constantiou & Kallinikos, 2015). Their primary similarity is that they both aim to identify, extract and refine hidden data patterns from massive sets of unordered and divergent information to determine internal relationships across objects (Chen & Wojcik, 2016). From this perspective, both approaches serve as granular methods of exploring data patterns or possible causal explanations (George et al., 2014).

Second, from a statistical point of view, big and small data studies often use similar algorithms. Algorithms are critical to both small and big data analyses. A wide range of commonly used methods—such as regression analysis, association rule learning, and social network analysis (SNA)—can be applied in both approaches (Manyika et al., 2011).

Table 3
Journal cluster analysis.

cluster1		cluster2		cluster3		cluster4	
label	Citations	label	Citations	label	Citations	label	Citations
harvard bus rev	66	acad manage j	95	bus horizons	25	hum resour manage-us	31
harvard business rev	17	acad manage rev	70	comput hum behav	17	j appl psychol	54
hum resour manag j	31	admin sci quart	27	manage sci	26	j manage	39
hum resour manage r	43	brit j manage	15	mis quart	33	pers psychol	22
hum resource manage	22	hum relat	17	oper res	18		
human resource manag	29	j manage stud	27				
human resource plann	19	organ sci	23				
int j hum resour man	72	oxford hdb managemen	18				
int j inform manage	34						
j bus res	19						
manage decis	26						
mit sloan manage rev	19						
organ dyn	21						
people strategy	29						
pers rev	16						

However, the traditional statistical techniques that are commonly used in small data analysis need to be improved to reduce their temporal and spatial complexity before they can be used on large-scale data (George et al., 2016). In particular, when the size of a dataset is larger than the amount of storage capacity available on a standard computer, a big data approach should be applied.

Third, both big and small data offer important empirical grounds for HRM strategy and behavioral studies. For example, both HRM strategy and behavioral scholars have examined their theories and models in organizations with small datasets to explore the mechanisms of the relationships among variables in explaining why certain HR practices are more effective than others (Van Beurden et al., 2020). Big data research based on a data-mining approach can also determine the HR practices that best predict employee performance (Bharadwaj et al., 2013).

4.2. Differences between big and small data research approaches

In this section, five notable differences between small and big data approaches are discussed, including differences in reliance on theory (Chen & Wojcik, 2016), data structures and content (Berman, 2013), analyses (Tonidandel et al., 2018), measurements (Shin, 2016) and reproducibility (Constantiou & Kallinikos, 2015). The differences between small and big data research approaches are summarized in Table 4.

Table 4
Differences between small and big data research.

Differences	Small data research	Big data research
Reliance on theory	Answer specific research questions based on existing theories.	Does not necessarily rely on a specific theory at the start.
Data source and content	Highly structured data based on carefully designed data collection.	Both structured and unstructured data can be applied and are generally by-products of machine and Internet activity.
Analysis	All data can be put in a dataset and analysed together simultaneously based on traditional psychometric or econometric approaches.	Data are typically extracted, reviewed, cleaned, structuralized, visualized, interpreted, and analysed using methodologies rooted in computer science.
Measurements	Typically, HRM constructs are measured with data from public archives or surveys in small data research.	A single rule cannot define a universal protocol of measurement because data varieties are too numerous.
Replication	The entire project, including data collection, should be able to be replicated.	Replication of big data research is hardly feasible or necessary.

The first difference between small and big data research is the reliance on theory. According to Hofman et al. (2017), big data research in social science can be categorized as either deductive (which focuses on explanation) or inductive (which focuses on prediction). Big data enable HRM researchers to conduct both deductive research (i.e., theory testing) and inductive research (i.e., data mining/theory building). For example, big data methods, such as SAM (split/analyze/*meta*-analyze), can be used to test existing theories (Cheung & Jak, 2016; Zhang et al., 2019), whereas clustering (Kobayashi et al., 2018) and deep learning techniques (Lecun et al., 2015) can be used to develop new theories and to enhance prediction accuracy in HRM research. The small data approach is typically used to address specific research questions that must be built on solid theories (Constantiou & Kallinikos, 2015). However, big data analysis does not necessarily rely on theory because a big data approach can be used to identify and mine specific patterns in data (Constantiou & Kallinikos, 2015) and does not require a theory to guide the modelling process (Berman, 2013).[2] As Mcabee et al. (2017) advocated, adopting a big data research approach based on an inductive paradigm can facilitate the advancement of HRM research.

The second difference lies in data structure and content. While traditional empirical HRM research normally processes only numerical data, a big data approach can help HRM scholars and practitioners manage unstructured data (Chen & Wojcik, 2016). Small data are usually highly structured based on a carefully designed data collection strategy. The scope of the data is often restricted to a single discipline (Kaufman, 2010), and the data are generally in the form of uniform records in an ordered spreadsheet; an example is survey data based on Likert-type scales (Chen & Wojcik, 2016). However, big data can include both structured and unstructured data, which are generally by-products of machine and Internet activity. Therefore, big data analysis can incorporate unstructured data, which refer to "raw, unstructured digital information, such as text or images" (Chen & Wojcik, 2016, p. 459). Unstructured data are characterized by a lack of structural organization, which makes numerical analysis difficult. In addition, unstructured data are generated in natural settings that are not produced for research purposes.

Third, big and small data are significantly different in terms of data analysis, and a big data approach can help HRM research avoid the issue of false correlation to a certain extent. Small data analysis typically uses a "P" value to measure the significance of the relationships among

[2] Although big data research is highly data-driven, we suggest that an important goal of HRM research is to be able to explain why relationships occur; the data or type analysis is not necessarily more important than this ability. Therefore, big data research can be an important initial step for theory building.

variables. However, using traditional statistical tools to analyze big data can lead to spurious relationships because the large volume of big data may produce many significant yet meaningless relationships (George et al., 2014). Therefore, instead of "P" values, directly providing effect sizes is more reasonable since even a tiny difference can likely be significant in very large sample sizes. Concerning big data, practical significance is more important than statistical significance. In addition, in small data analysis, all data can be entered into a single datasheet—such as in Excel, Stata or SPSS—and analyzed simultaneously based on traditional psychometric or econometric approaches. However, big data analyses contain vast numbers of unstructured data, which are impossible to analyze based on traditional analytical platforms (Cheung & Jak, 2016). Some "big" datasets are simply too large to be analyzed by using conventional desktop software and require a big data analytical platform that can store raw data and conduct analyses (e.g., Hadoop). Such datasets are often accessed through an application programming interface (API) and stored with a database management system, such as MySQL. Thus, for big data analysis, the data are typically extracted, reviewed, cleaned, structuralized, visualized, interpreted, and analyzed by using methodologies rooted in computer science (Berman, 2013). In several large analytical platforms, such as R and Python, packages are available for conducting analyses based on big data approaches (Wenzel & Van Quaquebeke, 2018).

The fourth difference between big and small data is the measurements used; big data can help to solve the issues of measurement error and social desirability in HRM research. In traditional small data research, HRM researchers apply well-established scales based on previous research (Bainbridge et al., 2017). However, no standardized measurement rule exists in big data research because big data come in various types—including textual, video, and audio forms—and a single rule cannot restrict diversified methods of variable operationalization (George et al., 2016). The advantage of applying big data for measurement is the ability to reveal employees' "real" preferences at work rather than their stated preferences, attitudes and intentions, which is usually captured by survey research. For example, employees' affect receives considerable attention in HRM research. However, most current HRM studies use self-report surveys to measure employees' positive/negative affect. This method relies mainly on employees' self-rating on "levels of positive and negative affect over a particular period of time or to make a judgement of their overall life quality" (Bogomolov, Lepri, & Pianesi, 2013, p. 790). Big data, such as the data generated from smart phone usage, weather conditions and individuals' personality traits, can better capture employees' true states of happiness, thus improving the accuracy and reliability of employees' affect measurements (Bogomolov, Lepri, & Pianesi, 2013).

The fifth difference is the need for replication. Traditional HRM studies value replication because it helps solve the issue of sampling errors, whereas another benefit of big data in HRM research is that sampling errors are less likely to be a problem. In small data research, the findings must be replicable (Berman, 2013). In particular, sample-based research is highly influenced by sampling errors. Therefore, if data are qualified in terms of reliability and validity, both the data collection and the entire project should be replicable. However, the replication of big data research projects is unnecessary because they are often based on population-level rather than on sample data (Harlow & Oswald, 2016). The results based on big data are less likely to be influenced by sampling errors (Yahav et al., 2016).

5. Methods of big data analyses in hrm research

Management research is undergoing a paradigmatic shift that necessitates a rigorous re-examination of HRM research methods based on the big data paradigm. On the one hand, big data approaches enhance the precision of HR managers' decisions (Angrave et al., 2016). On the other hand, big data provides HRM researchers with better insights to examine some important research questions (Kosinski et al., 2016). From different disciplines, we identified eleven among various big data methods that can be used effectively in HRM research; these eleven methods are structural equation modelling (SEM) trees (Brandmaier et al., 2016), classification analysis (Chakraborty et al., 2018), genetic algorithms (Strandmark et al., 2020), prescriptive analytics (Berk et al., 2019), latent semantic analysis (LSA) and differential language analysis (DLA) (Chen & Wojcik, 2016; Kern et al., 2016), singular value decomposition (SVD) (Kosinski et al., 2016), the experimental approach of event studies (Jones et al., 2016), SNA (Zhang et al., 2016), the SAM approach (Cheung & Jak, 2016; Zhang et al., 2019), clustering techniques (Kobayashi et al., 2017) and deep learning techniques (Lecun et al., 2015). These techniques and methods have been selected for three reasons. First, these methods and techniques are currently underutilized in HRM research (e.g., Zhang et al., 2019). Second, these methods can be applied to examine critical HRM topics, such as the social capital of a company (e.g., Zhang et al., 2016). Third, these methods are especially appropriate for use with big datasets and therefore serve as bridges to integrate big data analyses with HRM research (Kern et al., 2016). In this section, we illustrate key HRM research questions that can be investigated in big data research and describe relevant big data analysis methods and software.

5.1. *Methods in the inductive research paradigm*

5.1.1. *Textual analyses: LSA, DLA, and sentiment analyses—Research question: Employee selection and recruitment*

Textual analyses are an important method in big data analyses, as text is the most common type of unstructured data. In textual analyses, LSA (Chen & Wojcik, 2016) and DLA are powerful tools that have been widely adopted (Rohrer et al., 2017). The assumption of LSA is that words are not randomly distributed because words with similar meanings are more likely to appear in the same document; the method used in LSA is based on information retrieval; and the search algorithms used in LSA aim to find similar terms that express similar meanings in documents (e.g., webpages). Additionally, the DLA-based model is used to identify text that relates to specific characteristics (Kern et al., 2016). Moreover, text can be divided into different subgroups based on each topic. LSA and DLA have great potential in HRM research. For example, traditional vocational tests require HR managers to spend substantial time matching appropriate employees to their jobs (Wang & Wanberg, 2017). When LSA and DLA are used to analyze the public linguistic information of employees, such as job reports and work diaries, the cost of matching employees to jobs decreases rapidly, and the precision of employee-job matches increases (Anderson et al., 2012; Gill, 2004; Kern et al., 2016). Employee characteristics drawn from LSA and DLA are more objective, accurate and natural than tests based on questionnaires since the former are generated in a natural setting. Various analytical packages, such as scikit-learn in Python, can be used to conducted textual analyses (Chen & Wojcik, 2016).

5.1.2. *Data pattern extraction: SVD and latent Dirichlet allocation (LDA)—Research question: Employees' performance assessment*

SVD can increase the interpretability of data by reducing its dimension (Kosinski et al., 2016). This dimensionality reduction technique is commonly used in various contexts, such as computational social sciences and machine learning (Wall et al., 2003). LDA, which is another interpretable dimensionality reduction technique, can be used to detect patterns in text (Kosinski et al., 2016). Both SVD and LDA are critical tools for extracting patterns from large amounts of data by reducing data dimensionality. In HR research, factor and principal component analyses are traditional methods of pattern extraction. Factor analyses include both exploratory and confirmatory factor analyses (EFA and CFA, respectively), which are used mainly for models with latent variables in HRM research (Bainbridge & Broady, 2017). EFA is applied in the scale development stage to identify the underlying relationships between items and to reduce redundant items (Wang et al., 2019). CFA is used to confirm whether the items on a scale are consistent with a theoretical structure of this construct (Kline, 2015). Principal component analyses are similar to EFA in terms of their function of reducing redundant data and are used mainly for variable reduction to identify highly correlated variables rather than individual items (Alavi et al., 2020). The major difference among SVD, LDA, factor analyses and principal component analyses is that factor and principal component analyses mainly analyze numerical data, but SVD and LDA can be used to analyze both numerical and textual data. In HRM, evaluating employees by using multiple performance indicators is difficult (Den Hartog et al., 2013). SVD and LDA can effectively identify key performance indicators that are important for organizations and can be conducted by using the analytical package Matrix in R.

5.1.3. *Deep learning techniques: Multi-layer perceptron (MLP)—Research question: Prediction of HR demand*

Deep learning techniques represent a broad class of different machine learning algorithms based on ANNs. Various disciplines, such as speech recognition, machine translation, and image recognition, have applied deep learning techniques (Lecun et al., 2015). In deep learning techniques, MLP, also known as feed-forward neural networks, is especially relevant for HR research because of its effectiveness in prediction research. MLP is a class of ANNs whose neurons form an acyclic graph in which information moves in only one direction from input to output; thus, MLP is extensively used in pattern recognition. In HR research, Gu and Zhen (2018) applied MLP to examine the predictive power of companies' investment in HRM on companies' performance by using an international dataset from companies in more than 100 countries. Góes and De Oliveira (2020) further discussed the implication of MLP in HR performance evaluation. MLP can also be applied to investigate variables, such as labor demand, that are usually characterized by the time series of the company's development stage and product demands. For prediction research, MLP is an important deep learning technique for enhancing prediction effectiveness. PyTorch, which is written in Python, is commonly used for deep learning.

5.1.4. *Classification analyses—Research question: Organizational performance prediction*

Big data research has multiple methods for classification, such as classification trees and support vector machines (SVMs). Classification trees were developed based on decision trees. Classification trees are used for inference tasks or classification problems (Bertsimas & Dunn, 2017). For instance, Chakraborty et al. (2018) combined classification trees and ANNs to identify qualified students in a business school. Another classification approach is SVMs, which are rooted in machine learning. They address mainly issues of binary classification. Specifically, as "a large set of observations with known labels (training set), SVM finds a maximum margin function that separates the observations into two classes where each observation is a point in a multidimensional space of feature measurements" (Ghaddar & Naoum-Sawaya, 2018, p. 994). HRM researchers can apply SVMs along with other machine learning techniques and factor analyses (EFA and CFA) to compare the influence of various predictive models of HRM bundles on organizational performance. Predictive models comparison research has been conducted in studies on organizational performance prediction (e.g., Delen et al., 2013). The analytical package scikit-learn in Python can be used for classification analyses.

5.1.5. *Genetic algorithms—Research question: Employee scheduling*

Strohmeier and Piazza (2015) pointed out that artificial intelligence (AI) can improve employee scheduling systems based on genetic algorithms. Genetic algorithms consider both employees' biological processes and organization arrangement; consequently, genetic algorithms can generate solutions according to specified objective functions and problem constraints. Employee scheduling (also called "staff rostering") is an application scenario that uses genetic algorithms in HRM; this application enables employees to deliver optimal assignment generations (Gong et al., 2019). Genetic algorithms take the fit between employees' biological condition and working schedule as an optimization problem, thereby providing feasible rosters that consider a multitude of constraints. Genetic algorithms, i.e., selection, crossover and mutation, can produce improved rosters and the roster with the best fitness value (Strandmark et al., 2020).Genetic algorithms can automate the rostering tasks and evaluate valid rosters, which outperforms manual scheduling.

5.1.6. *Prescriptive analytics: Robust and adaptive optimization—Research question: Human resource planning optimization*

In past decades, the technique of optimization formulation has been used in predicting the HR planning process (Berk et al., 2019). For example, Kawas et al. (2013) proposed a nonlinear optimization formulation model to predict the sales in a sales team. Additionally, Santos et al. (2013) implied a two-stage integer optimization to assign employees to incoming projects. One popular optimization technology, namely, stochastic optimization, is widely used in the research on HR decision making in uncertainty environments (Bertsimas & Kallus, 2020; Birge & Louveaux, 2011). However, stochastic optimization cannot address the "curse of dimensionality" in multiperiod decision making in HR planning. Recently, Berk et al. (2019) proposed "affinely adaptive robust optimization", which combines robust and adaptive optimization, to identify the uncertainties of HR planning and its dynamic nature. The authors adapted the data from Sapient Corporation to verify robust and adaptive optimization, which is useful in hiring advice and staffing recommendations.

5.2. *Methods in the deductive research paradigm*

5.2.1. *Experimental approach: Event study—Research question: Strategic management, succession planning, and organizational design*

Using the traditional small data approach to study the impacts of events on an organization is often time-consuming and costly (Jones et al., 2016). A big data approach may open a new path for researching the impacts of events. Instead of working in the field, researchers who conduct event studies in the big data era can focus mainly on online news, comments and even real-time reporting about an event. The explosion of big data generated by the Internet, particularly on social media platforms, thus becomes the basis for experimental event studies. Event studies include four steps: comparative organization selection, data collection, creating measures, and data analysis. First, control and experimental groups must be selected. The second step is data collection, in which researchers obtain data from social media platforms' APIs or organizations' BBSs. In traditional event studies, different stakeholders are difficult to identify. A big data-based event study can effectively collect data from multiple stakeholders, such as employees, consumers and governments. Third, identifying keywords to evaluate an event's influence is critical to creating measures (Haas et al., 2015). Fourth, in the data analysis, data from control and experimental groups are used to test the impacts of the event. This approach can be used in examining the impact of organizational change on companies' HRM systems; for example, this approach can be used to evaluate how companies' mergers and acquisitions affect companies' strategic HR choices. The analytical package twitteR in R can be used to collect big data for event studies.

5.2.2. *The SAM approach—Research question: Cross-country comparative HRM research*

The SAM approach was recently proposed by Cheung and Jak (2016). It follows the split/apply/combine method of big data computation by incorporating a *meta*-analysis to help behavioral researchers integrate psychometric analyses (Wickham, 2011). This approach converts a big dataset, which is much larger than the storage capacity of a computer, into many manageable sub-datasets and then combines the results by using a *meta*-analysis. Zhang et al. (2019) further developed this approach by integrating a multilevel framework that can be used to conduct big data analyses with multilevel modelling. For example, the multilevel SAM (MSAM) approach can be applied to investigate how countries' HR systems affect the relationship between companies' HRM investments and firm financial performance (Zhang et al., 2019). Specifically, researchers first split the big data datasets into several small datasets according to the country of the sample. Second, a regression analysis is applied within each sub-dataset. Third, a *meta*-analysis integrates all the results that draw from each sub-dataset into one result. Fourth, a mixed model is applied to examine the effect of country-level factors. This approach combines the advantages of big data analysis with the traditional psychometric approaches that are familiar to most HRM researchers. This approach can be used for both theory-driven HRM research and data mining. The metaSEM R package available at the Comprehensive R Archive Network (CRAN) is usually used for SAM and MSAM analyses.

5.3. Methods in the hybrid (both inductive and deductive) research paradigm

5.3.1. SEM trees—Research question: How HRM influences employees' attitudes and behaviors

SEM trees combine the strengths of SEM and the decision tree paradigm (Brandmaier et al., 2016). The advantage of this method is that it can solve both theory- and data-driven research questions. Specifically, SEM proposes a fundamental framework to test the existing theory, and decision trees can be used to conduct an exploratory analysis to modify the theory (Jacobucci et al., 2017). The decision tree approach aims to explore the associations between the potential indicator effects and the specific outcomes (Brandmaier et al., 2013). SEM trees help researchers discover subgroup heterogeneity. Specifically, SEM trees can reveal the interactive relationships among variables in multivariate statistical analyses based on a theoretical model. In particular, researchers first use SEM to propose or test a prior theory and then use a decision tree to refine the theory. The decision tree serves as a search function to find the best explanation for how potential indicators predict an outcome. For instance, SEM trees can examine the different change trends of work performance for employees in different subgroups. The analytical package semtree in CRAN is designed for SEM tree analysis.

5.3.2. SNA—Research question: Team social interaction

SNA is used to examine social exchanges among individuals or groups (Liu et al., 2015). It is also applied to examine how individuals are embedded within a social structure and how social structures emerge from the micro-relationships among individuals. Social network research, as an important branch of sociology, has been applied to big data research in management (Zhang et al., 2016), education (Isba et al., 2017), and political science (Maireder et al., 2017). For example, Zhang et al. (2016) used SNA to analyze the relationships among various companies based on the social network platform Facebook. Two sophisticated methods in social network research include exponential random graph models (ERGMs) and auto-logistic actor attribute models (ALAAMs). ERGMs can explain different network ties simultaneously, including the presence and absence of network ties, while ALAAMs can examine the network effects of national-, community- and individual-level factors on individual consequences. ERGMs are used to predict the emergence of different social network ties, whereas ALAAMs, an extension of ERGMs (Letina, 2016), can be used to predict actor characteristics from social networks at the micro level and the interdependencies among individuals at the meso level (Daraganova & Robins, 2013; Lusher et al., 2013). Both methods can be used to examine within-team social interaction. Various analytical packages, such as Statnet in R, can be used to conduct textual analyses.

5.3.3. Clustering techniques—Research question: Strategic HRM

Clustering functions group a set of objects into the same cluster, which has a representative prototype that represents objects in the cluster according to some specific rules or principles (Kobayashi et al., 2017). Objects in a cluster are more similar to each other regarding some characteristics than to objects in other clusters (Kobayashi et al., 2017). In clustering analyses, K-means clustering is an efficient method that is often applied in management research at a very early stage. For example, Brentan et al. (2018) employed K-means clustering to examine management effectiveness in water distribution systems. The principle of K-means clustering is to minimize the total mean-squared error between the training samples and their representative prototypes (Novikov, 2019). Procedurally, K-means clustering includes four steps. First, an initial partition with K clusters must be specified. Second, each pattern is assigned to its closest representative prototypes. Third, new representative prototypes are identified. Fourth, stateable representative prototypes are identified by repeating steps 2 and 3. For HR research, K-means clustering can be effective for identifying "bundles" of HR practices. A cluster analysis of sequences is often used for a cluster analysis.

We summarize the big data analysis methods used in HRM research in Table 5 and Fig. 7.

6. General procedures for big data research

The general procedures for big data research consist of four steps. The first step is data collection. Data collection or acquisition is the bedrock of big data research. Methods such as sensors, web scraping, web traffic, and communications monitoring are efficient for collecting large-scale and multitudinous data (e.g., Chaffin et al., 2017; Landers et al., 2016; Zhang et al., 2016). The second step is data exploration and pre-processing. Methods such as correlation graphs, summary statistics, and visualization techniques can be used for data exploration in big data research. Pre-processing the data involves two steps: (1) extracting the data by detecting and correcting them and (2) reshaping the clean data into the format needed for an analysis (Tonidandel et al., 2018). The third step is data analysis. There are various techniques, rooted in computer science, for big data analysis (Chen & Wojcik, 2016). Several general categories are classification, regression, clustering and association analyses. These different techniques are used in different contexts according to distinct research topics. The fourth step is reporting the results (Chen & Wojcik, 2016). Because of the volume and variety of big data, determining how to report the results becomes a critical challenge. On the one hand, the volume of big data may lead to spurious conclusions because all the relationships between pairs of variables tend to be significant when using a traditional significance test due to the high statistical power of extremely large sample sizes (Simonsohn et al., 2019). On the other hand, the variety in big data leads to difficulty in presenting the various data sources in a uniform format. Additionally, the steps and processes used in pre-processing and merging data must be reported and discussed (Haas et al., 2015). For example, visualization technology provides vivid image displays so that audiences can clearly understand the data processing results (Tu et al., 2017). As data science develops further, the number of new visualization tools (such as BIME, Qlik Sense, and Tableau) that can be used in HRM research will continue to increase.

Following these four steps, HRM researchers can adopt a big data approach to investigate HR topics. The general procedures for big data research are summarized in Fig. 8.

7. Implications for hrm research in the era of big data

The result of combining our narrative review and bibliometric analysis suggests that big data HRM research is an emerging field that awaits further study. However, the rapid development of big data offers both opportunities and challenges for HRM research. To help integrate big data approaches into HRM research, we suggest five potential big data HRM topics according to the levels of analysis and discuss three challenges and limitations that HRM research could face in the era of big

Table 5
Methods of big data analysis for HRM research.

Method	The Application Context	Data Type	Empirical Research
Textual analysis	A latent semantic analysis (LSA) and a differential language analysis (DLA) can both be used to perform textual analyses. The assumption of LSA is that the words are not randomly distributed because words with similar meanings are more likely to appear in the same document, whereas a DLA-based model is used to identify texts that relate to specific characteristics.	Textual data	Chen and Wojcik (2016) Kern et al. (2016)
Prescriptive analytics: Robust and adaptive optimization	This method combines both robust and adaptive optimization to identify both the uncertainty and the dynamic nature of human resource planning.	Both quantitative and qualitative data	Berk et al. (2019)
Genetic algorithms	Genetic algorithms fit the task by addressing its very nature as an optimization problem and by providing feasible rosters considering a multitude of constraints. Using a genetic algorithm, i.e., selection, crossover and mutation, leads to generation of improved rosters, and selection of the roster with the best fitness value.	Both quantitative and qualitative data	Strandmark et al. (2020)
Data pattern extraction	Singular value decomposition (SVD) and latent Dirichlet allocation are two commonly used dimensionality reduction techniques. They can be used in detecting the patterns of both large quantitative and qualitative datasets.	Both quantitative and qualitative data[a]	Rohrer et al. (2017)
Deep learning techniques	In deep learning techniques, multi-layer perceptron (MLP), which is also known as a feedforward neural network, is especially relevant for HR research because of its effectiveness in prediction research. MLPs are a class of artificial neural networks whose neurons form an acyclic graph in which information moves in only one direction from input to output; MLPs are extensively used in pattern recognition.	Quantitative data	Góes and De Oliveira (2020)
Classification approach	Classification approaches include both classification trees and support vector machines. Classification trees were developed based on decision trees. They are used for inference tasks or classification problems in HRM. Another classification approach is support vector machines, which are rooted in machine learning. This approach mainly concerns issues of binary classification.	Both quantitative and qualitative data	Kosinski et al. (2016)
Experimental approach: Event study	An experimental event study is used to analyse how employees react to a particular event (i.e., how the organization is affected by events) by collecting and analysing big data related to the event through social media platforms for control and experiment groups.	Both quantitative and qualitative data	Oreg et al. (2018)
Split/analyse/*meta*-analyse (SAM) approach	SAM follows the split/apply/combine method of big data computation by incorporating *meta*-analyses to help behavioural researchers integrate psychometric analyses. This approach converts a big dataset, which is much larger than the amount of random-access memory of a computer, into many manageable sub-datasets and then combines the results by using a *meta*-analysis.	Quantitative data	Y. E. Zhang et al. (2019) Cheung and Jak (2016)
Structural equation modelling (SEM) trees	SEM trees combine the strengths of SEM and the decision tree paradigm. They can be used with an auto-regression model, a regression model, a longitudinal growth curve model, or a standard factor model.	Quantitative data	Brandmaier et al. (2016)
Social network analysis (SNA)	SNA is used to examine the social exchange among individuals or groups. SNA is applied to examine the ways individuals are embedded within a social structure and the ways social structures emerge from the micro-relationships among individuals.	Both quantitative and qualitative data	Jacobucci et al. (2017)
Clustering techniques	The goal of cluster analysis, such as K-means clustering, is to group a set of objects with similar characteristics into the same group (called a cluster) by some specific rules. Objects in the same cluster are more similar to each other than to those in other clusters.	Both quantitative and qualitative data	K. Zhang et al. (2016)

Note. a. Qualitative data must be transformed into quantitative data for an SVD approach.

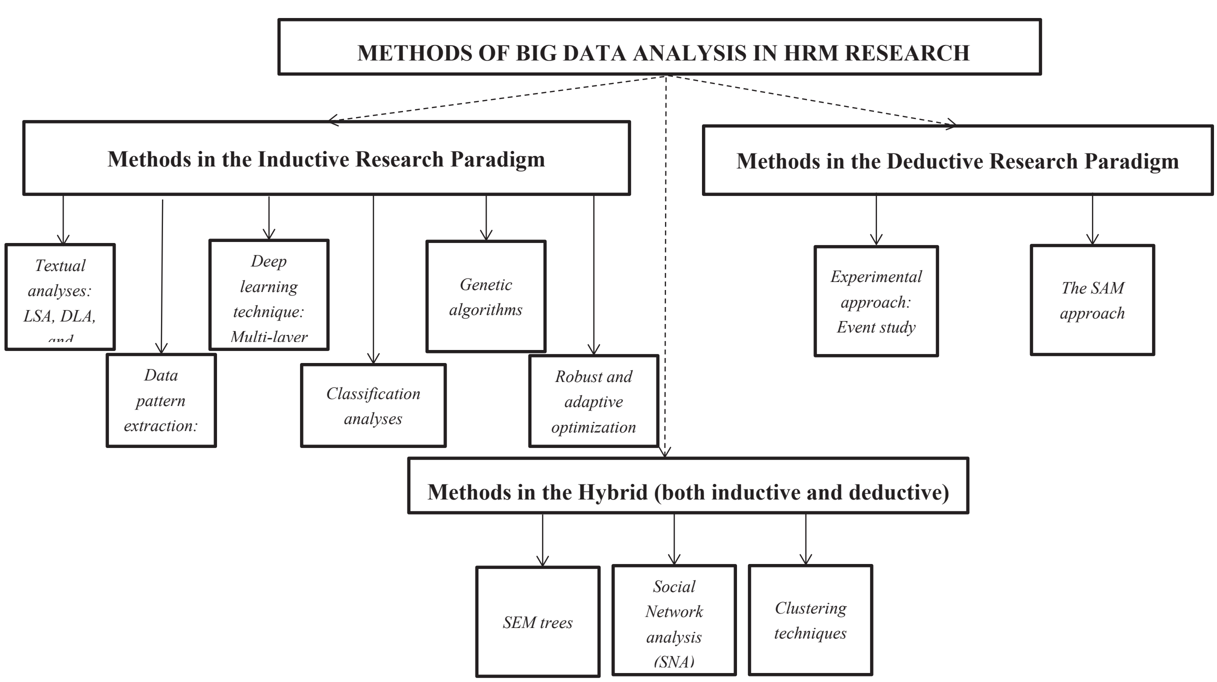

Fig. 7. Methods of Big Data Analysis in HRM Research.

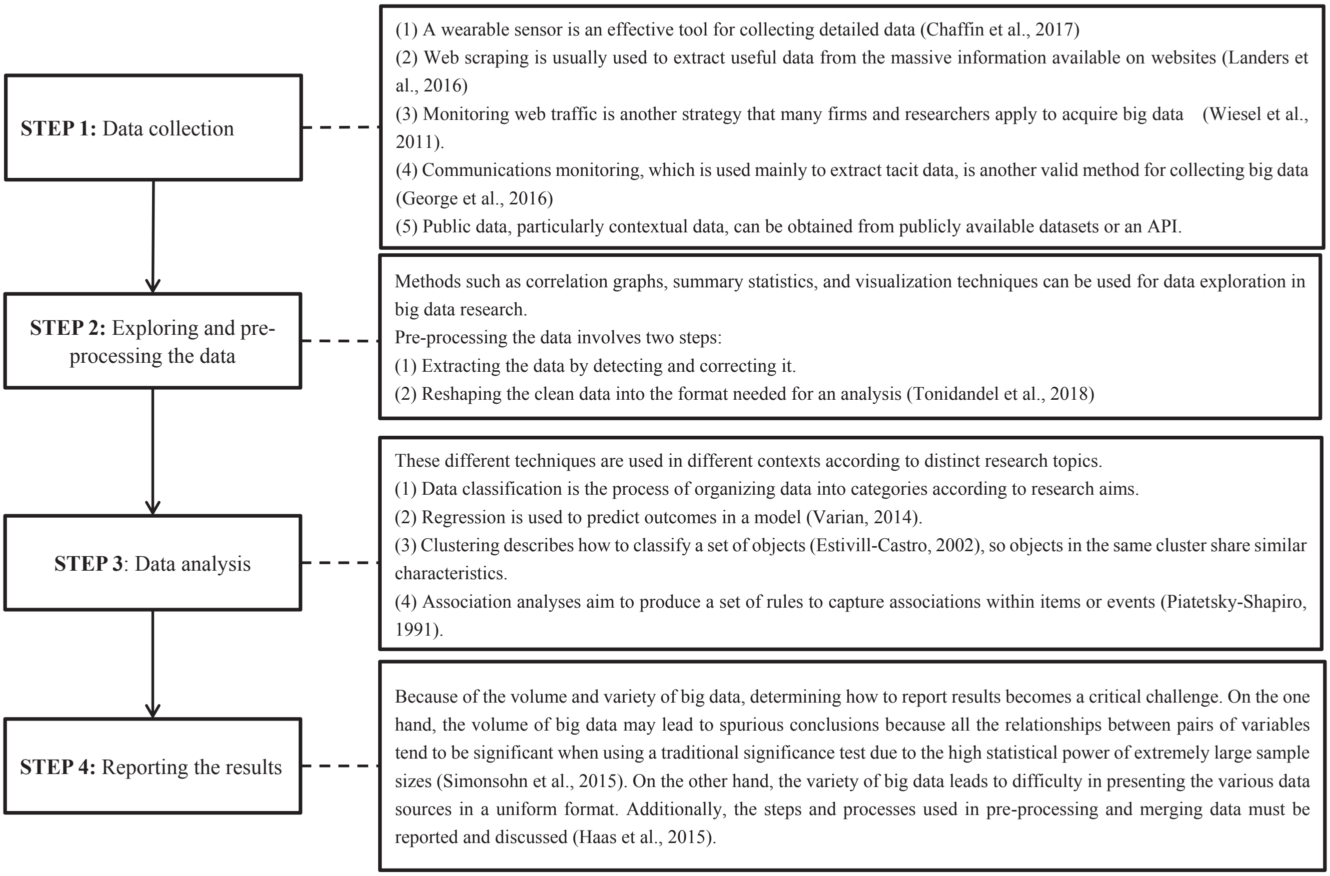

Fig. 8. General Procedures for Big Data Research.

data.

7.1. Opportunities for HRM research in the era of big data

In this section, we propose five promising HRM topics that may benefit from the application of big data at the micro, meso and macro levels.

7.1.1. Micro-level topics

7.1.1.1. Applying big data collection techniques to HRM—employee attitudes and behaviors research. The micro level here refers to the between- and within-individual levels (Baron & Greenberg, 1990). In HRM research, this level of research focuses mainly on employees' attitudes and behaviors (such as employee's job satisfaction and in-role performance) in an organization. Compared with small data approaches, such as the survey method, big data collection techniques can greatly improve the validity of measurements for employees' attitudes and behaviors in HRM research. For example, the application of wearable sensors can reduce the measurement error of subjectivity when measuring employee-level attitudes and behaviors. A recent study that compared data based on surveys and wearable sensors indicated that the latter approach is more likely to yield valid results (Barber et al., 2017) partly because the findings based on survey data are more easily contaminated due to social desirability (Podsakoff et al., 2003). Further, while traditional survey approaches in HRM research can rarely collect real-time longitudinal data, wearable sensors can enable real-time monitoring. Thus, wearable sensors can collect real-time longitudinal data and track changes in employees' attitudes and behaviors. This capability enables addressing research questions about how HR practices influence employees' attitudes and behavioral dynamics, especially when integrating the experience sampling method (ESM) with wearable sensors (Beal, 2015).[3] For instance, employee physical engagement, as an important type of work engagement (Kahn, 1990), receives less attention than cognitive and emotional engagement largely because of the lack of valid measurement approaches. By applying wearable sensors (such as in actigraphy) in an ESM study, physical engagement can be effectively measured because the number of steps and the energy expenditure of employees in the workplace are recorded. Similarly, with the help of wearable sensors and smart phone applications, micro-level research topics, such as employee interactions (Chaffin et al., 2017) and daily stress (Bogomolov & Lepri et al., 2014) have been better (empirically) measured and studied.

7.1.2. Meso-level topics

7.1.2.1. Integrating multiple types of big data in HRM—team interaction research. At the meso level, which refers to research at the working-team level (Baron & Greenberg, 1990), the combination of companies' internal and social media data provides promising future advancement for HRM team-related research. Currently, instant messenger applications, such as LINE and WeChat, are widely used in team interactions (Wang et al., 2015), and these applications generate a large amount of textual data. By adopting inductive textual analysis methods, such as LSA and DLA, such instant messenger applications provide opportunities for HRM scholars to better investigate how HR practices affect intra- and

[3] For example, ExperienceSampler is an experience sampling smartphone application that is specifically designed for ESM research (Thai & Page-Gould, 2018); this application can be used to integrate wearable sensors with ESM studies. This application can be downloaded through the web at http://www.experiencesampler.com/.

inter-team interactions, such as the communication between a team leader and members, thereby further affecting distal team outcomes, such as team performance. The data can be collected from either formal team interactions (such as conference information and documents) or informal communications (such as daily chat records). However, collecting team interaction data may lead to high privacy and scientific misconduct risks, which is a problem that has become increasingly salient in recent years (Chan & Perrig, 2003). Therefore, a major challenge for collecting team interaction data in HRM research is to determine the boundary between public and private information and to protect employees' privacy. We further discuss privacy and ethics issues in the following sections.

7.1.3. Macro-level topics

7.1.3.1. People analytics in HRM research. In HRM, the macro-level refers to the organization level. Using a big data approach, we can better understand people analytics, also called HR analytics, which is defined as "the systematic identification and quantification of the people drivers of business outcomes" (Van den Heuvel & Bondarouk, 2017, p. 130). In particular, the following three streams of research questions in people analytics can be better addressed by applying big data methods. First, how can organizations more effectively recruit and select high-potential talent? Second, how can managers track the development of skills within the organization to promote work design? Third, how can employee attitudes and behaviors that are traditionally considered micro-level constructs be understood in new ways in terms of the level of analysis?

First, personnel recruitment research can be improved by using a big data approach, such as analyzing online interviews (Chamorro-Premuzic et al., 2017) and social networks (Garcia-Arroyo & Osca, 2019). Second, work design research can be further advanced by using employees' digital data, such as data from computer and mobile use (Liu et al., 2017), to investigate how knowledge workers arrange their workday and how they allocate their time to different tasks. Third, by integrating a big data approach, traditionally defined micro-level individual employee attitudes and behaviors can be investigated at the collective level (Bowen & Ostroff, 2004), thereby enabling HRM researchers to explore new research questions about the dynamic relationships between HR practices and collective-level employee attitudes and behaviors. For instance, natural language processing methods, such as linguistic or sentiment analyses of job-related texts that employees post on social media platforms, can be used to predict collective-level employee attitudes and behaviors (Chung & Pennebaker, 2012; Hernandez et al., 2016).

7.1.3.2. AI implications in the employee recruitment and selection and employee turnover research. The recent development of AI offers emerging opportunities for recruitment and turnover management research. First, by incorporating AI algorithms with candidate résumés and other sources of data, new theoretical models can be developed to revise and even reform recruitment. For example, Khosla et al. (2016) discussed how human-robot interaction modelling can be used in employee recruitment and retention. Both verbal and non-verbal data can be collected from robots and then used to track and interpret the candidate's emotions. Further, a candidate's emotional state can be used to help evaluate his or her commitment to answering job interview questions. Second, with the implications of ANNs in employee turnover research, new types of antecedents of employee turnover (e.g., social network-related antecedents, such as social ties) can be better studied empirically. Recently, HRM researchers have suggested that with the help of a network-based textual analysis, online data (including textual data) can be a complementary data source to traditional small quantitative data in HRM research (Pérez-Campdesuñer et al., 2018).

7.1.3.3. HR planning and cross-country HRM comparative research. Organizations' HR planning refers to an HRM function that recognizes an organization's HR needs and plays a crucial role that connects HRM and firm strategic management (Jiang & Messersmith, in press; Kaufman, 2010). HR planning can be better investigated by using the prediction approach of big data (Mcabee et al., 2017). Traditional small data methods used in the HR planning research include regression and trend analyses (e.g., Cook & Ferris, 1986), which suffer from methodological limitations, such as sampling and measurement errors (Yahav et al., 2016). A big data approach with large volumes of information and better algorithms can address these limitations and thereby aid in developing better HR planning models and eventually more accurate predictions of HR demands. For example, by using a machine learning technique (namely, the least absolute shrinkage and selection operator) to analyze the predictors of HR demand, the precision of HR planning (regarding, e.g., the proportion of internal training, promotion and external recruitment) can be greatly improved (Varian, 2014).

In addition to inductive research, cross-country comparative deductive research that focuses on identifying country-level boundary conditions can be improved by using big data methods (Cooke & Bartram, 2015). For example, based on the MSAM method, Zhang et al. (2019) examined how a country-level HRM system impacts the relationship between a company's investment in HRM and the company's financial performance. Zhang et al. (2019) divided the cross-country big dataset "OSIRIS", which contains 620,000 firm records for all public companies worldwide from 1983 to 2016, into 75 sub-samples according to sample location, analyzed the model in each sub-sample, and further tested the cross-country moderation effects by using a mixed-effect model *meta*-analysis.

7.2. Challenges and limitations for HRM research in the era of big data

7.2.1. Ethical issues

Although a big data approach provides promising opportunities for HRM research, this approach also creates new challenges. First, big data collection suffers from high ethical risks (Menon & Sarkar, 2016; Tonidandel et al., 2018). Most big data regarding employees' online behaviors, such as online searches and transactions, are by-products of employees' everyday lives in the Internet era and are not generated for research purposes. Further, anonymity in the dataset might not be guaranteed because identity can be detected from information found in the digital footprint (Kosinski et al., 2016). Therefore, in big data analysis, protecting employee privacy is the first challenge for HRM scholars. HRM researchers must be cautious about using big data to investigate HRM issues and must avoid violating individuals' or organizations' privacy. Current privacy protection policies and ethical standards developed for big data HRM research alone may be insufficient to protect employees' privacy. The application of privacy protection techniques can be equally important. For example, to protect individuals' privacy when using transactional data, Menon and Sarkar (2016) developed scalable approaches that can be applied in further studies about employee remuneration.

7.2.2. Ontological challenges

This new data-driven research paradigm underscored by a big data approach is difficult for traditional theory-driven (i.e., hyper-positivist ontological framework) HRM researchers to understand and manage. To facilitate data-driven research in HRM, we suggest that HRM scholars apply "reverse thinking" to big data HRM research, i.e., research using inductive-oriented approaches (Popper, 2005). Mcabee et al. (2017) advocate that the HRM field requires a big data shock to better apply inductive research. However, the dominant deductive research mindset of HRM researchers may impede applying a big data approach.

7.2.3. Analytical and methodological difficulties

Although HRM methodology researchers have made some efforts to develop easy-to-use big data methods that fit HRM researchers' backgrounds (Zhang et al., 2019), most big data approaches are rooted in computer science; therefore, these approaches differ from most methods with which HRM scholars are familiar (Cheung & Jak, 2016). HRM researchers with knowledge of traditional psychometrics and econometrics must learn new programming skills to query and analyze big data. However, using data science analytic technologies is difficult for HRM researchers and requires familiarity with big data analytical platforms, such as Hadoop, Python, R, and MATLAB (Douthitt & Mondore, 2014). The high threshold of programming skills for big data research may decelerate the paradigm shift from sample-based empirical research to big data research in HRM. Therefore, on the one hand, the training system for HR practitioners and researchers should be updated by providing more education in advanced mathematical, statistical and programming knowledge. On the other hand, as Tonidandel et al. (2018) suggest, multidisciplinary collaboration is critical for statisticians and computer scientists who can provide analytical support for HRM researchers. Moreover, HRM researchers need to learn and adopt research methods from big data psychology research (Harlow & Oswald, 2016). For example, Cheung and Jak (2016) developed a SAM approach to analyze big data by splitting data into sub-datasets for analysis and then integrating the results of the sub-datasets with a *meta*-analytical approach; this hybrid approach can help psychology scholars to conduct big data research by using traditional psychometric approaches without studying complex data science methods.

Finally, the ways to evaluate and select valid and reliable data are challenging for HRM researchers due to the variety of big data types. Instead of mainly numerical data, various types of data, including text and video, are used in big data HRM studies. For example, textual data, especially from social media, can be "highly uncertain in both its expression and content"; therefore, the need to work with non-numerical data may impede HRM researchers from achieving reliable results (Claverie-Berge, 2012, p. 3). Thus, it is especially difficult for HRM researchers who are used to working with numerical data to detect potential biases in data collection and analysis processes or to guarantee validity and reliability (Lazer et al., 2014).

8. Conclusion

Empirical research in both the natural and social sciences is undergoing a paradigmatic shift from the traditional small data era to the emerging big data era. As this trend continues, HRM research will be informed by a rigorous re-examination and extension of previous research methods and findings. Embracing the big data approach is becoming increasingly important for HRM research. Therefore, the purpose of this study is to help integrate big data approaches into HRM research by systematically reviewing current big data HRM studies and identifying new directions for future research.

Based on a quantitative review that uses a bibliometric analysis to visualize and map the evolution of big data HRM research, this paper systematically discusses the importance, benefits, and characteristics of big data research and explores the ways that big data techniques and approaches can be applied to HRM research. This paper proposes an emerging "Inductive (Prediction paradigm: Data mining/Theory building) *vs.* Deductive (Explanation paradigm: Theory testing)" research framework for big data HRM research; this framework includes potential HRM research questions that can use big data and their corresponding levels of analysis, analytical methods, data sources and software. In proposing this framework, this paper conducts an integrative review of the characteristics of big data, compares the characteristics of big data with those of traditional small data, outlines the general procedure for conducting big data research, and describes the eleven methods that can be applied in big data HRM studies.

Finally, this study further discusses five specific, promising HRM research topics at the micro, meso and macro levels along with three challenges and limitations that HRM scholars may face in the era of big data. Future HRM research could benefit from applying big data approaches in five ways. At the micro level, employee attitude and behavior research can benefit from applying big data collection techniques. At the meso level, team interaction research can benefit from integrating multiple types of big data. At the macro level, HRM research can benefit from conducting people analytics, analyzing the implication of AI in employee recruitment and selection and employee turnover research, and assessing HR planning and cross-country HRM comparative research. However, ethical issues, ontological challenges, and analytical and methodological difficulties become three critical challenges and limitations of integrating big data approaches into HRM research, which warrant future research attention. We hope that systematically reviewing and discussing the current state of the science and the future research agenda can offer a roadmap for applying big data approaches to HRM studies and inspire future big data HRM research.

Acknowledgement

This study was funded by National Natural Science Foundation of China (Grant No. 71972065, 71802077, 71602163 and 71790593).

References

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Exploratory factor analysis and principal component analysis in clinical studies: Which one should you use? *Journal of Advanced Nursing, 76*(8), 1886–1889.
Alvesson, M., & Deetz, S. (2000). *Critical overview of quantitative and conventional qualitative methodology' Doing critical management research, sage, london* (pp. 49–80). London: Sage Publications Ltd..
Anderson, B., Fagan, P., Woodnutt, T., & Chamorro-Premuzic, T. (2012). Facebook psychology: Popular questions answered by research. *Psychology of Popular Media Culture, 1*(1), 23–37.
Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). Hr and analytics: Why hr is set to fail the big data challenge. *Human Resource Management Journal, 26*(1), 1–11.
Bainbridge, H. T., & Broady, T. R. (2017). Caregiving responsibilities for a child, spouse or parent: The impact of care recipient independence on employee well-being. *Journal of Vocational Behavior, 101*, 57–66.
Bainbridge, H. T., Sanders, K., Cogin, J. A., & Lin, C. H. (2017). The pervasiveness and trajectory of methodological choices: A 20-year review of human resource management research. *Human Resource Management, 56*(6), 887–913.
Barber, L. K., Taylor, S. G., Burton, J. P., & Bailey, S. F. (2017). A self-regulatory perspective of work-to-home undermining spillover/crossover: Examining the roles of sleep and exercise. *Journal of Applied Psychology, 102*(5), 753–763.
Baron, R. A., & Greenberg, J. (1990). *Behavior in organizations*. Boston: Allyn and Bacon.
Beal, D. J. (2015). Esm 2.0: State of the art and future potential of experience sampling methods in organizational research. In F. P. Morgeson (Ed.), Annual review of organizational psychology and organizational behaviour (Vol. 2(2), pp. 383–407).
Beaton, D., Dunlop, J., & Abdi, H. (2016). Partial least squares correspondence analysis: A framework to simultaneously analyze behavioral and genetic data. *Psychological Methods, 21*(4), 621–651.
Berk, L., Bertsimas, D., Weinstein, A. M., & Yan, J. (2019). Prescriptive analytics for human resource planning in the professional services industry. *European Journal of Operational Research, 272*(2), 636–641.
Berman, J. J. (2013). *Principles of big data: Preparing, sharing, and analyzing complex information*. Waltham, MA: Elsevier.
Bertsimas, D., & Dunn, J. J. M. L. (2017). Optimal classification trees. *Machine Learning, 106*(7), 1039–1082.
Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science, 66*(3), 1025–1044.
Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly, 37*(2), 471–482.
Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.
Blazquez, D., Domenech, J. J. T. F., & Change, S. (2018). Big data sources and methods for social and economic analyses. *Technological Forecasting and Social Change, 130*, 99–113.
Bogomolov, A., Lepri, B., & Pianesi, F. (2013). Happiness recognition from mobile phone data. Paper presented at the Social Computing (SocialCom) 2013 International Conference on.
Bogomolov, A., Lepri, B., Kessler, F. B., Pianesi, F., & Pentland, A. S., Orlando, F. L. (2014). Daily stress recognition from mobile phone data, weather conditions and individual traits. Paper presented at the The 22nd ACM International Conference in Multimedia, Orlando, FL.
Bowen, D. E., & Ostroff, C. (2004). Understanding hrm-firm performance linkages: The role of the "strength" of the hrm system. *The Academy of Management Review, 29*(2), 203–221.
Brandmaier, A. M., Prindle, J. J., Mcardle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods, 21*(4), 566–582.

Brandmaier, A. M., Von Oertzen, T., Mcardle, J. J., & Lindenberger, U. (2013). Structural equation model trees. *Psychological Methods, 18*(1), 71–86.
Braun, M. T., Kuljanin, G., & Deshon, R. P. (2018). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods, 21*(3), 633–659.
Brentan, B., Meirelles, G., Luvizotto, E., Jr, Izquierdo, J. J. E.m., & Software. (2018). Hybrid som+ k-means clustering to improve planning, operation and management in water distribution systems. *Environmental Modelling & Software, 106*, 77–88.
Cappelli, P. (2017). There's no such thing as big data in hr. *Harvard Business Review, 2*.
Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2017). The promise and perils of wearable sensors in organizational research. *Organizational Research Methods, 20*(1), 3–31.
Chakraborty, T., Chattopadhyay, S., & Chakraborty, A. K. (2018). A novel hybridization of classification trees and artificial neural networks for selection of students in a business school. *Opsearch, 55*(2), 434–446.
Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (2017). The datafication of talent: How technology is advancing the science of human potential at work. *Current Opinion in Behavioral Sciences, 18*, 13–16.
Chan, H., & Perrig, A. (2003). Security and privacy in sensor networks. *Computer, 36*, 103–105.
Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences, 101*(1), 5303–5310.
Chen, C. (2006). Citespace ii: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359–377.
Chen, C., Song, I.-Y., Yuan, X., & Zhang, J. (2008). The thematic and citation landscape of data and knowledge engineering (1985–2007). *Data & Knowledge Engineering, 67*(2), 234–259.
Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods, 21*(4), 458–474.
Chen, C. (2014). Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace [Available at <8-emerging trends in regenerative.Pdf>].
Cheung, M. W.-L., & Jak, S. (2016). Analyzing big data in psychology: A split/analyze/meta-analyze approach. *Frontiers in Psychology, 7*(738). https://doi.org/10.3389/fpsyg.2016.00738.
Cheung, L. (2017). Integrating human capital analysis into small business lending in china. (Ph.D.), School of Management, UNSW Business School, Australia.
Chung, C. K., & Pennebaker, J. W. (2012). Linguistic inquiry and word count (liwc): Pronounced "luke",... And other useful facts. In *Applied natural language processing: Identification, investigation and resolution* (pp. 206–229). IGI Global.
Claverie-Berge,I. (2012). Solutions big data ibm. [Available at http://www05.ibm.com/fr/events/netezzaDM_2012/Solutions_Big_Data.pdf].
Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology, 30*(1), 44–57.
Cook, D. S., & Ferris, G. R. (1986). Strategic human resource management and firm effectiveness in industries experiencing decline. *Human Resource Management Journal, 25*(3), 441–457.
Daraganova, G., & Robins, G. (2013). Auto logistic actor attribute model. In D. Lusher, J. Koskinen, & G. Robins (Eds.), *Exponential random graph models for social networks: Theory, methods and applications* (pp. 102–114). Cambridge: Cambridge University Press.
Delen, D., Zaim, H., Kuzey, C., & Zaim, S. (2013). A comparative analysis of machine learning systems for measuring the impact of knowledge management practices. *Decision Support Systems, 54*(2), 1150–1160.
Den Hartog, D. N., Boon, C., Verburg, R. M., & Croon, M. A. (2013). Hrm, communication, satisfaction, and perceived performance: A cross-level test. *Journal of Management, 39*(6), 1637–1665.
Douthitt, S., & Mondore, S. (2014). Creating a business-focused hr function with analytics and integrated talent management. *People and Strategy, 36*(4), 16.
Edelman, B. (2012). Using internet data for economic research. *The Journal of Economic Perspectives, 26*(2), 189–206.
Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897–904.
Fang, B., & Zhang, P. (2016). *Big data in finance*. Springer International Publishing.
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144.
Garcia-Arroyo, J., & Osca, A. (2019). Big data contributions to human resource management: A systematic review. *The International Journal of Human Resource Management*, 1–26.
George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal, 57*(2), 321–326.
George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal, 59*(5), 1493–1507.
Ghaddar, B., & Naoum-Sawaya, J. (2018). High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research, 265*(3), 993–1004.
Gill, G. (2004). Sumo and ubiquitin in the nucleus: Different functions, similar mechanisms? *Genes & Development, 18*(17), 2046–2059.
Góes, A. S. D. O., & De Oliveira, R. C. L. (2020). A process for human resource performance evaluation using computational intelligence: An approach using a combination of rule-based classifiers and supervised learning algorithms. *IEEE Access, 8*, 39403–39419.
Gong, X., Wang, S., & Jiao, R. (2019). An efficient 2d genetic algorithm for optimal shift planning considering daily-wise shift formats: A case of airport ground staff scheduling. Paper presented at the 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM).
Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'big data 'revolution in healthcare. McKinsey Quarterly. http//www payer fusion 14/02/The-big-data-revolution-in-healtheare-l. pdf. 2013-1-15/2015-10-09.
Gu, X., & Zhen, Y. (2018). Supervised machine learning methods: A general analytical framework and illustration in hrm research. Working paper.
Haak, T. (2014). Escaped from the hr analytics lab [Available at http://hrtrendinstitute.com/2014/2010/2021/escaped-from-the-hr-analytics-lab/ (accessed 2027 November 2014)].
Haas, M. R., Criscuolo, P., & George, G. (2015). Which problems to solve? Online knowledge sharing and attention allocation in organizations. *Academy of Management Journal, 58*(3), 680–711.
Hanisch, B., & Wald, A. (2012). A bibliometric view on the use of contingency theory in project management research. *Project Management Journal, 43*(3), 4–23.
Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology: Introduction to the special issue. *Psychological Methods, 21*(4), 447.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). Overview of supervised learning. In *The elements of statistical learning* (pp. 9–41). Springer.
Hernandez, I., Newman, D. A., & Jeon, G. (2016). Twitter analysis: Methods for data management and a word count dictionary to measure city-level job satisfaction. In S. Tonidandel, E. B. King, & J. M. Cortina (Eds.), *SIOP organizational frontier series. Big data at work: The data science revolution and organizational psychology* (pp. 64–114). Routledge/Taylor & Francis Group.
Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science, 335*(6324), 486–488.
Isba, R., Woolf, K., & Hanneman, R. (2017). Social network analysis in medical education. *Medical Education, 51*(1), 81–88.
Jacobucci, R., Grimm, K. J., & Mcardle, J. J. (2017). A comparison of methods for uncovering sample heterogeneity: Structural equation model trees and finite mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(2), 270–282.
Jiang, K., & Messersmith, J. (in press). On the shoulders of giants: A meta-review of strategic human resource management. International Journal of Human Resource Management, 1-28.
Jones, N. M., Wojcik, S. P., Sweeting, J., & Silver, R. C. (2016). Tweeting negative emotion: An investigation of twitter data in the aftermath of violence on college campuses. *Psychological Methods, 21*(4), 526.
Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *The Academy of Management Journal, 33*(4), 692–724.
Kaufman. (2010). Shrm theory in the post-huselid era: Why it is fundamentally misspecified. *Industrial Relations: A Journal of Economy and Society, 49*(2), 286–313.
Kawas, B., Squillante, M. S., Subramanian, D., & Varshney, K. R. (2013). Prescriptive analytics for allocating sales teams to opportunities. Paper presented at the 2013 IEEE 13th International Conference on Data Mining Workshops.
Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods, 21*(4), 507–525.
Khosla, R., Chu, M.-T., & Nguyen, K. (2016). Human-robot interaction modelling for recruitment and retention of employees. Paper presented at the International Conference on HCI in Business, Government, and Organizations.
Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational Research Methods, 21*(3), 733–765.
Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods, 21*(4), 493–506.
Kozlowski, S. W., Chen, G., & Salas, E. (2017). One hundred years of the Journal of Applied Psychology: Background, evolution, and scientific trends. *Journal of Applied Psychology, 102*(3), 237–253.
Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the internet for use in psychological research. *Psychological Methods, 21*(4), 475–492.
Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science, 343*(6176), 1203–1205.
Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
Letina, S. (2016). Network and actor attribute effects on the performance of researchers in two fields of social science in a small peripheral community. *Journal of Informetrics, 10*(2), 571–595.
Liu, C., Gan, B., & Zhang, Y. C. (2015). Why 'they' occupies the critical network positions? *Management Decision, 53*(1), 100–123.
Liu, C., Olivola, c., & Kovacs, B. (2016). Coauthorship trends in the field of management: Facts and perceptions. *Academy of Management Learning and Education, 16*(4), 509–530.
Liu, C.-H., Wang, J. S., & Lin, C.-W. (2017). The concepts of big data applied in personal knowledge management. *Journal of Knowledge Management, 21*(1), 213–230.
Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.
Maireder, A., Weeks, B. E., Gil De Zúñiga, H., & Schlögl, S. (2017). Big data and political social networks: Introducing audience diversity and communication connector bridging measures in social network theory. *Social Science Computer Review, 35*(1), 126–141.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity. 2011*. KY: Lexington: McKinsey & Company.

Mcabee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review, 27*(2), 277–290.
Mcafee, A., & Brynjolfsson, E. (2012). Big data: The management revolution: Exploiting vast new flows of information can radically improve your company's performance. But first you'll have to change your decision making culture'[2012] Harvard Business Review.
Menon, S., & Sarkar, S. (2016). Privacy and big data: Scalable approaches to sanitize large transactional databases for sharing. *MIS Quarterly, 40*(4), 963–982.
Mueen, A. (2014). Time series motif discovery: Dimensions and applications.*Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, 4*(2), 152–159.
Novikov, A. V. (2019). Pyclustering: Data mining library. *Journal of Open Source Software, 4*(36), 1230–1235.
Oreg, S., Bartunek, J. M., Lee, G., & Do, B. (2018). An affect-based model of recipients' responses to organizational change events. *Academy of Management Review, 43*(1), 65–86.
Pérez-Campdesuñer, R., De-Miguel-Guzmán, M., Sánchez-Rodríguez, A., García-Vidal, G., & Martínez-Vivar, R. (2018). Exploring neural networks in the analysis of variables that affect the employee turnover in the organization. *International Journal of Engineering Business Management, 10*, 1–11.
Podsakoff, P. M., Mackenzie, S. B., Jeong-Yeon, L., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
Popper, K. (2005). *The logic of scientific discovery*. Routledge.
Rohrer, J. M., Brümmer, M., Schmukle, S. C., Goebel, J., & Wagner, G. G. (2017). " What else are you worried about?"–integrating textual responses into quantitative social science research. *PLoS ONE, 12*(7), 182–1156.
Santos, C., Gonzalez, T., Li, H., Chen, K.-Y., Beyer, D., Biligi, S., … Ramanujam, R. (2013). Hp enterprise services uses optimization for resource planning. *Interfaces, 43* (2), 152–169.
Schermann, M., Hemsen, H., Buchmueller, C., Bitter, T., Krcmar, H., Markl, V., & Hoeren, T. (2014). Big data: An interdisciplinary opportunity for information systems research. *Business & Information Systems Engineering, 6*(5), 261–266.
Schwartz, B. S., Gretczko, M., & Sloan, N. (2016). The gig economy: Distraction or disruption. . In P. B. & J. Schwartz (Eds.), Global human capital trends (pp. 105–110).
Shafique, M. (2013). Thinking inside the box? Intellectual structure of the knowledge base of innovation research (1988–2008). *Strategic Management Journal, 34*(1), 62–93. https://doi.org/10.1002/smj.2002.
Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an hr context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research, 70*, 366–378.
Shin, D.-H. (2016). Demystifying big data: Anatomy of big data developmental process. *Telecommunications Policy, 40*(9), 837–854.
Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2019). Specification curve: Descriptive and inferential statistics on all reasonable specifications. Available at SSRN 2694998.
Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research, 70*, 263–286.
Strandmark, P., Qu, Y., & Curtois, T. (2020). First-order linear programming in a column generation based heuristic approach to the nurse rostering problem. *Computers & Operations Research, 104945*.
Strohmeier, S., & Piazza, F. (2015). Artificial intelligence techniques in human resource management—a conceptual exploration. In *Intelligent techniques in engineering management* (pp. 149–172). Cham: Springer.
Thai, S., & Page-Gould, E. (2018). ExperienceSampler: An open-source scaffold for building smartphone apps for experience sampling. *Psychological Methods, 23*(4), 729–739.
Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods, 21*(3), 525–547.
Tu, C., He, X., Shuai, Z., & Jiang, F. (2017). Big data issues in smart grid–A review. *Renewable and Sustainable Energy Reviews, 79*, 1099–1107.
Van Beurden, J., Van De Voorde, K., & Van Veldhoven, M. (2020). The employee perspective on HR practices: A systematic literature review, integration and outlook. *The International Journal of Human Resource Management*, 1–35.
Van den Heuvel, S., & Bondarouk, T. (2017). The rise (and fall?) of HR analytics. Journal of Organizational Effectiveness: People and Performance. Article submitted for the 2nd HR Division International Conference (HRIC) on February 20-22, 2016 in Sidney, Australia.
Van Der Vegt, G. S., Essens, P., Wahlström, M., & George, G. (2015). Managing risk and resilience. *Academy of Management Journal, 58*(4), 971–980.
Varian, H. R. (2014). Big data: New tricks for econometrics. *The Journal of Economic Perspectives, 28*(2), 3–27.
Wall, M. E., Rechtsteiner, A., & Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis* (pp. 91–109). Boston, MA: Springer.
Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics, 34*(2), 77–84.
Wang, L., & Cotton, R. (2018). Beyond moneyball to social capital inside and out: The value of differentiated workforce experience ties to performance. *Human Resource Management, 57*(3), 761–780.
Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research, 70*, 287–299.
Wang, Y.-S., Tseng, T. H., Wang, Y.-M., & Chu, C.-W. J. I. R. (2019). Development and validation of an internet entrepreneurial self-efficacy scale. *Internet Research, 30*(2), 653–675.
Wang, M., & Wanberg, C. R. (2017). 100 years of applied psychology research on individual careers: From career management to retirement. *Journal of Applied Psychology, 102*(3), 546–563.
Wang, R., Xue, M., Liu, K., & Qian, H. (2015). *Data-driven privacy analytics: A wechat case study in location-based social networks*. Cham: Springer.
Wenzel, R., & Van Quaquebeke, N. (2018). The double-edged sword of big data in organizational and management research: A review of opportunities and risks. *Organizational Research Methods, 21*(3), 548–591.
Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software, 40*(1), 1–29.
Wright, P. M., & Mcmahan, G. C. (2011). Exploring human capital: Putting 'human'back into strategic human resource management. *Human Resource Management Journal, 21* (2), 93–104.
Yahav, I., Shmueli, G., & Mani, D. (2016). A tree-based approach for addressing self-selection in impact studies with big data. *MIS Quarterly, 40*(4), 819–A819.
Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-scale network analysis for online social brand advertising. *MIS Quarterly, 40*(4), 849–A812.
Zhang, Y. C., Liu, S., Xu, S., Yang, M., & Zhang, J. (2019). Integrating split/analyze/meta-analyze (sam) approach and multilevel framework to advance big data research in psychology: Guidelines and an empirical illustration via human resource management investment-firm performance relationship. *Zeitschrift für Psychologie-Journal of Psychology, 226*(4), 274–283.
Zhang, L., Zhang, Y., Jiang, H., Yang, M. M., Huang, Y. Y., & Li, S. J. (2017). Customer identification in the healthcare industry. *International Journal of Market Research, 59* (6), 803–822.
Zhang, Y., Zhang, L., Lei, H., Yue, Y., & Zhu, J. (2016). Lagged effect of daily surface acting on subsequent day's fatigue. *The Service Industries Journal, 36*(15–16), 809–826.

Yucheng Zhang (PhD, University of New South Wales) is a Professor at School of Public Administration, Southwestern University of Finance and Economics, Chengdu, China. He has published papers in journals such as Journal of Applied Psychology, Journal of Management, Journal of Business Research, Tourism Management, Asia Pacific Journal of Management, International Journal of Human Resource Management, and Journal of Business Ethics. His current research interests include abusive supervision, dual career workers, and meta-SEM (structural equation modeling) analysis. ORCID iD: https://orcid.org/0000-0001-9435-6734.

Shan Xu (PhD, Southwestern University of Finance and Economics) is an Associate Professor at School of Business Administration, Southwestern University of Finance and Economics in China. Shan Xu has published papers in Psych Journal, International Journal of Human Resource Management, Cornell of Hospitality Quality and Journal of Vocational Behaviour, among others. Her research interests include human resource management and organizational behaviour. ORCID iD: https://orcid.org/0000-0002-0184-4470.

Long Zhang (PhD, The University of New South Wales) is an Associate Professor at Business School, Hunan University, Changsha, China. His research has been published in journals such as Journal of Management, Journal of Business Ethics, Asia Pacific Journal of Management, International Journal of Human Resource Management, Employee Relations, and Cornell Hospitality Quarterly. His research interests include human capital, strategic human resource management, organizational behavior, and leadership. ORCID iD: https://orcid.org/0000-0001-9704-2976

Mengxi Yang (PhD, Tsinghua University) is an Assistant Professor at School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China. His work has been published in journals such as Human Relations and Business Ethics: A European Review. His research interest lies in leadership and human resource management. ORCID iD: https://orcid.org/0000-0001-7082-2956.