

# Big Data in occupational medicine: the convergence of -omics sciences, participatory research and e-health

GUGLIELMO DINI<sup>1,2</sup>, NICOLA LUIGI BRAGAZZI<sup>1</sup>, ALFREDO MONTECUCCO<sup>1</sup>,  
ALESSANDRA TOLETONE<sup>3</sup>, NICOLETTA DEBARBIERI<sup>1,2</sup>, PAOLO DURANDO<sup>1,2</sup>

<sup>1</sup>Department of Health Sciences (DISSAL), Postgraduate School of Occupational Medicine, University of Genoa, Genoa, Italy

<sup>2</sup>Occupational Medicine Unit, Policlinico San Martino Hospital, Genoa, Italy

<sup>3</sup>Occupational Medical Service, Local Health Unit 1, Liguria Regional Healthcare System, Imperia, Italy

**KEY WORDS:** work 4.0; new information and communication technologies; exposomics; occupational medicine

**PAROLE CHIAVE:** lavoro 4.0; nuove tecnologie dell'informazione e della comunicazione; esposomica; medicina del lavoro

## SUMMARY

**Background:** *New occupational hazards and risks are emerging in our progressively globalized society, in which ageing, migration, wild urbanization and rapid economic growth have led to unprecedented biological, chemical and physical exposures, linked to novel technologies, products and duty cycles. A focus shift from worker health to worker/citizen and community health is crucial. One of the major revolutions of the last decades is the computerization and digitization of the work process, the so-called "work 4.0", and of the workplace. Objectives:* To explore the roles and implications of Big Data in the new occupational medicine settings. **Methods:** *Comprehensive literature search. Results:* Big Data are characterized by volume, variety, veracity, velocity, and value. They come both from wet-lab techniques ("molecular Big Data") and computational infrastructures, including databases, sensors and smart devices ("computational Big Data" and "digital Big Data"). **Conclusions:** *In the light of novel hazards and thanks to new analytical approaches, molecular and digital underpinnings become extremely important in occupational medicine. Computational and digital tools can enable us to uncover new relationships between exposures and work-related diseases; to monitor the public reaction to novel risk factors associated to occupational diseases; to identify exposure-related changes in disease natural history; and to evaluate preventive workplace practices and legislative measures adopted for workplace health and safety.*

## RIASSUNTO

**«I Big Data e la Medicina del Lavoro: la convergenza delle scienze omiche, della ricerca partecipativa e dell'e-health».** **Introduzione:** *Nuovi pericoli e rischi occupazionali stanno emergendo a causa di una società sempre più globalizzata, in cui l'invecchiamento, la migrazione, l'urbanizzazione selvaggia e la rapida crescita economica hanno portato a esposizioni biologiche, chimiche e fisiche senza precedenti, insieme all'introduzione di nuove tecnologie, prodotti e cicli lavorativi. È di cruciale necessità una transizione dall'attenzione alla salute dei lavoratori alla salute dei lavoratori/cittadini e delle comunità. Questo cambiamento ha accompagnato il passaggio dalla medicina industriale alla medicina del lavoro del ventesimo secolo, in cui la prevenzione svolge un ruolo importante. Una delle mag-*

Pervenuto il 17.10.2018 - Revisione pervenuta il 4.2.2019 - Accettato il 6.3.2019

Corrispondenza: Dr. Nicola Luigi Bragazzi, MPH, PhD, MD Department of Health Sciences, Postgraduate School of Public Health, University of Genoa, via Antonio Pastore 1, Genoa, 16132 Italy - Tel. +39 010 3538508 - Fax +39 010 3538541

E-mail: robertobragazzi@gmail.com

giori rivoluzioni degli ultimi decenni è data dall'informatizzazione e dalla digitalizzazione del lavoro, il cosiddetto «lavoro 4.0», e del luogo di lavoro. **Obiettivi:** L'attuale revisione mira a esplorare i ruoli e le implicazioni dei Big Data nella medicina del lavoro. **Metodi:** È stata eseguita una ampia ricerca bibliografica. **Risultati:** I Big Data sono caratterizzati da volume, varietà, veridicità, velocità e valore. I Big Data possono provenire sia da tecniche laboratoristiche (“Big Data molecolari”) sia da tecnologie computazionali e informatiche, includendo database, sensori e dispositivi intelligenti (“Big Data computazionali” e “Big Data digitali”). **Conclusioni:** I Big Data molecolari e digitali, insieme a nuovi approcci analitici, possono dare un prezioso contributo alla medicina del lavoro. Gli strumenti computazionali e digitali possono consentire di scoprire nuove relazioni tra esposizioni e malattie professionali, oltre a monitorare la reazione del pubblico ai nuovi fattori di rischio associati alle malattie professionali e ai relativi cambiamenti nella storia naturale della malattia e a verificare l'efficacia delle pratiche e misure legislative preventive adottate per migliorare la salute e sicurezza sul luogo di lavoro.

#### SHIFTING FROM “OCCUPATIONAL SAFETY AND HEALTH” TO “TOTAL WORKER HEALTH”: NEW CHALLENGES FOR OCCUPATIONAL MEDICINE

New occupational risks and hazards are emerging due to an increasingly globalized society, in which ageing, migration, wild urbanization and rapid economic growth have led to unprecedented biological, chemical and physical exposures. This is also accompanied by the new technologies, products and duty cycles being regularly introduced (18, 64). Since the overall population is mostly composed by individuals in working age, health issues arising from workplace conditions will consequently affect the general population health. In order to properly capture these new trends, a shift from a focus on “occupational safety and health” (accidents and injuries, work-related diseases, occupational deaths) to worker/citizen health and communities is crucial (64). Such a shift has paralleled the transition from the industrial medicine, characterized by the industrial revolutions of 1760-1830 to the 20<sup>th</sup> century

occupational medicine, in which prevention played a major role (37).

One of the major revolutions of the last decades was the computerization and the digitization of work, the so-called “work 4.0”, and of the workplace (table 1). This has led to forms of work that were unconceivable before, such as crowd working, in which “digital work” is broken down into a distributed array of smaller tasks, which are performed by “crowd-workers” by means of digital platforms and infrastructures (37).

Within the European Union, the “Community Strategy for Safety and Health at Work 2002-2006” has asked the “European Agency for Safety and Health at Work” (EU-OSHA) to establish a risk observatory, able to “anticipate new and emerging risks”, and to develop a “genuine culture of risk prevention” (27). Similarly, in the United States of America (USA), due to a constant industrial evolution, there is the need to constantly monitor occupational hazards: the “National Institute for Occupational Safety and Health” (NIOSH) “Health

**Table 1** - The concept of work 4.0

Revolution	Form of work
First industrial revolution	Steam power, mechanization and industrialization of work (work 1.0)
Second industrial revolution	Introduction of electricity and Taylorism (work 2.0)
Information revolution	Skilled work, electronics and automation of work (work 3.0)
Cybernetic and digital revolution	“Global revolution in manufacturing”: Big Data, artificial intelligence, Internet of Things, robotics and cobotics, real-time communication, integration of digital ecosystems and cyber-physical entities, remote sensing, intelligent environments, man-machine interconnection and cooperation, smart manufacturing and digital work (work 4.0)

Hazard Program” has recently recognized the task of mapping occupational risk factors and exposures as strategic and a priority (71).

“Total worker health”, a term introduced by the NIOSH in 2011, is defined “as policies, programs, and practices that integrate protection from work-related safety and health hazards with promotion of injury and illness prevention efforts to advance worker well-being”. According to a systematic review of the literature, total worker health-based interventions are effective, improving workforce health better and more rapidly than other programs (29).

### BIG DATA

Big Data are characterized by different Vs: volume (scale and quantity of data), variety (heterogeneity of data sources), veracity (accuracy, quality and reliability of data), velocity (real-time or near real-time data availability), and value (benefits deriving from collecting huge amounts of data). Big Data can come from both wet-lab techniques (“molecular Big Data”) and computational infrastructures, including databases, sensors and smart devices (“computational Big Data” and “digital Big Data”) (5, 69).

Big Data can be generated and acquired through the different activities performed by occupational physicians, such as risk evaluation and assessment, and biological monitoring (i.e., molecular Big Data) or worker health surveillance programs (i.e., computational and digital Big Data). Worker health surveillance enables a longitudinal, dynamic follow-up of the working population. This wealth of work-related information should be stored in *ad hoc* relational databases designed and implemented for occupational purposes. This should combine clinical (electronic health records, EHRs) and exposure registries. Assessing fitness to work could be an opportunity to perform a global assessment of workers’ health status, collecting data that is not strictly related to professional exposure, useful for identifying individual susceptibility and implementing *ad hoc* preventative measures and interventions and personalized health promotion programs. Moreover, these data could aid in developing mathematical models aimed to identify early predictors of health

effects by the means of advanced analytical techniques, such as Big Data analytics. The development and use of *ad hoc* prediction models is relatively new in occupational health and safety, but it is expected to tremendously advance this field, favoring the shift from OSH to TWH.

### MOLECULAR BIG DATA

Microarrays and wet-lab techniques can produce a wealth of data about occupational diseases. Genomics and proteomics represent two major approaches for identifying differentially expressed genes and proteins in a given disease. For example, Giusti and colleagues (30) examined tissue biopsies from 53 subjects who underwent diagnostic thoracoscopy for the suspicion of malignant pleural mesothelioma. Using differential proteomics techniques, authors identified well-known mesothelioma biomarkers such as calretinin and found novel biomarkers such as prelamin A/C, desmin, vimentin, fructose-bisphosphate aldolase A, myosin regulatory light chain 2, ventricular/cardiac muscle isoform, myosin light chain 3 and myosin light chain 6B. Tooker and collaborators (72) utilized surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometry and “Classification And Regression Trees” (CART) techniques in order to identify serum biomarkers in 35 asbestosis patients with cancer *versus* 35 asbestos-exposed controls without cancer. The authors identified kinesin proteins (namely, KIF18A and KIF5A) as predictors of cancer development. Morr e and coworkers (57) exploited the ONCOblot platform to investigate the role of ecto-nicotinamide adenine dinucleotide oxidase disulfide-thiol exchanger 2 (ENOX2) in asbestosis. The authors identified two mesothelioma-specific ENOX2 protein transcript variants in the serum of asbestos-exposed individuals 4-10 years prior to clinical diagnosis of malignant mesothelioma. Ju and collaborators (41), using 2-dimensional gel electrophoresis (2-DE) and matrix-assisted laser desorption/ionization time of flight tandem mass spectrometry (MALDI-TOF-MS/MS), were able to identify a panel of differentially expressed proteins in a group of 37 patients with asbestosis, 254 workers exposed to asbestos

and 439 healthy controls. These included cytokines,  $\alpha$ 1-AT, and L-ficolin, among others.

Recent breakthroughs and technological advancements have led to methods such as the “Slow Off-rate Modified Aptamers” (SOMAmers)-proteomics arrays, enabling a highly selective and well reproducible protein detection. SOMAmers are short, single stranded deoxyneucleotides, able to bind molecular targets and allow researchers to measure biomarkers, even from small volumes of biological samples (8, 53). Ostroff et al. (61) developed a candidate 13-biomarker panel for the detection of mesothelioma in asbestos-exposed population. Multi-center case-control studies conducted among 117 mesothelioma cases *versus* 142 asbestos-exposed controls led to the identification of new biomarkers involved in different biological pathways, including inflammation/immune response, proliferation and its maintenance, and cell growth regulation. Großerueschkamp and collaborators (32) implemented a novel approach to spatially resolve the heterogeneity of a tumor in a label-free manner integrating both Fourier Transform Infrared Spectroscopy (FTIR) imaging and laser capture micro-dissection (LCM). The authors identified 142 differentially expressed proteins, including five well-known and established protein biomarkers. Hegmans and coworkers (38) studied mesothelioma exosomes, which are small membrane vesicles secreted into the extracellular compartment by exocytosis, putatively playing a major role mediating the interaction between the tumor and the environment. They are believed to be involved in the biological process through which tumors are able to escape immunological response from the host. The authors detected major histocompatibility complex (MHC) class I together with the heat shock proteins HSC70 and HSP90, as well as with annexins and plasmalemmal vesicle associated protein type 1 (PV-1), proteins involved in membrane transport and function, cytoskeleton proteins and their associated proteins, such as ezrin, moesin, actinin-4, desmoplakin, and fascin. Other represented proteins were the molecular motor kinesin-like proteins, and developmental endothelial locus-1 (DEL-1), which is a strong angiogenic factor.

Other predictors can be microRNAs (miRNAs). Cavalleri and collaborators (13) have investigated a

sample of 23 malignant pleural mesothelioma patients and 19 cancer-free subjects with past asbestos exposure. The authors found 55 differentially expressed miRNAs. The most discriminating miRNA panel was given by the combination of miR-103a-3p and miR-30e-3p. Weber et al. (80) explored the expression of miR-103 in human blood in 23 mesothelioma patients, 17 asbestos-exposed controls and 25 controls from the general population. The authors found that miR-103 exhibited good sensitivity and specificity in distinguishing mesothelioma patients from controls. miR-103a-3p as biomarker can also be used in combination with mesothelin, as shown in another study, in which 43 male mesothelioma patients and 52 male controls formerly exposed to asbestos were recruited (81). Motta and colleagues (58) studied miRNAs expression in foundry workers at baseline and after 3 days of work (post-exposure to particulate matter). They found 4 miRNAs were differentially expressed in post-exposure compared with baseline samples, including miR-421, miR-146a, miR-29a, and let-7g.

Other post-genomics specialties are emerging, such as “breathomics” (75). Breathomics enables the systematic investigation of non-invasive metabolic biomarkers, like volatile organic compounds (VOCs) that are produced by virtually all metabolic processes of the body. VOCs are particularly suitable for studying obstructive lung diseases, even though some limitations should be properly addressed by researchers before their introduction in routine clinical practice. Potential limitations include the study of confounding factors (such as co-morbidities), and the need of a rigorous, standardized procedure. Yang and collaborators (84) performed a case-control study recruiting 200 stoneworkers, from which 5 subjects with asthma and 16 subjects under steroids or non-steroidal anti-inflammatory drugs were excluded (25 cases *versus* 154 controls). The authors found that pentane and C5-C7 methylated alkanes represented the major VOCs in the breath of individuals suffering from pneumoconiosis. The breath test was found to have good accuracy for pneumoconiosis diagnosis.

At a more comprehensive and holistic level, the exposome, term coined by Chris Wild (82), can be defined as “the totality of environmental exposures

encountered from birth to death" (40). Exposome can be assessed and quantitatively measured both externally and internally (76). Exposomics comprehensively includes all the previously mentioned omics disciplines, namely: genomics, transcriptomics, proteomics, metabonomics and breathomics, adductomics, and phenomics. Responsomics is the omics super-specialty, which deals with the quantitative measurement of the biological outcomes at a molecular/cellular level. It should be emphasized that the added value of exposomics does not rely only in the wealth of information, but also in its depth at the individual level, potentially capturing "personal exposures" and enabling a tailored, targeted treatment and management rather than "one-size-fits-all" approach (45). Studies such as the "Oxford Street Randomized Trial" (52), the ESCAPE project (62, 67, 74), the RAPTES (76) the TAPAS (83) and the "Health and Environment-Wide Associations based on Large Scale population Surveys" (HEALS) (1) can provide several insights into the effects of air pollutant exposure. Exposome-Explorer (available at <http://exposome-explorer.iarc.fr>) represents the first manually curated database including all known biomarkers of exposure to environmental risk factors, enabling researchers to better design *ad hoc* bio-monitoring/bio-surveillance studies or exposome-wide association investigations (EWAS) (59).

Finally, omics approaches have also been applied in very specific fields of occupational medicine, such as military occupational medicine (4, 21).

#### COMPUTATIONAL BIG DATA

Hamad and colleagues (35), exploiting a claims-based administrative database including a cohort of 14,161 USA workers during the period 1996–2011, developed and validated a "health risk score," termed "DxCG Intelligence tool". This risk score was associated with different outcomes: incident diagnosis of five disorders (namely, diabetes, hypertension, asthma/chronic obstructive pulmonary disease, depression, and ischemic heart disease), healthcare utilization, early retirement and occupational injuries. Furthermore, employee data were linked with the National Death Index, exploring association with mortality. Harber and Leroy (36) utilized basic

natural language processing lexical analysis techniques in order to automatically assess a database of 89,000 Mine Safety and Health Administration free text records. The authors were able to properly demonstrate common exposures, health effects, and exposure–injury relationships, even though many workplace terms were not present in the "Unified Medical Language System" (UMLS) or mapped in an inaccurate and inconsistent way. As such, there is the urgent need to enhance the UMLS vocabulary in order to make it more relevant to the field of occupational medicine.

Drury (24) has recently overviewed the role of Big Data analytics in the field of human factors and ergonomics, showing how data mining techniques can be used to infer meaningful associations between variables and also to quickly and effectively visualize data. Walker and Strathie (78) extensively collected data from the "On-Train Data Recorders" (OTDR) database and analyzed it in order to capture the most important strategic risks currently faced by rail operators and authorities worldwide, using a data-driven approach rather than a hypothesis-based framework. The authors reviewed over 300 ergonomics methods and elaborated nine candidate "Human Factors Leading Indicators", combining psychological knowledge, ergonomics methods and Big Data analytics. Bevilacqua and coworkers (2) applied non-conventional analytical techniques such as classification tree methods (i.e., CART) to data regarding accidents in a medium-sized refinery, to capture the important relationships between the different ergonomic, management and operational variables in a dynamic way. The authors were able to find new cause-effect correlations in occupational safety, previously never described.

Delaunay and colleagues (20) have recently explored the feasibility of incorporating geographical information systems (GISs)-generated information to improve occupational health and safety surveillance, developing both a "macro-approach" (from national to local level) and a "micro-approach" (at a plant/enterprise level). The authors were able to capture new information from the integration of risk assessment data and medical data, and gain a better understanding of the complex interplay between variables.

## DIGITAL DIG DATA

Online reporting systems can be exploited in the field of occupational medicine. For instance, “Monitoring trends in Occupational Diseases and tracing new and Emerging Risks in a NETwork” (MODERNET) represents a tool that enables the monitoring and tracking of new and emerging occupational trends. A recent survey has investigated current European reporting policies and systems for occupational diseases. It has been found that 14 countries (70% of the 20 countries participating in the study) provided information for 33 occupational diseases systems. Eleven of these were compensation-based, whereas additional six countries provided information for non-compensation based systems. The other systems, which reported only occupational diseases from a specific list, were physician-based (11).

In Italy, the pilot network “MALattie e Rischi Emergenti sul Lavoro” (MAREL), which includes five occupational disease consultation centers of university hospitals in central-northern Italy, aims at collecting and detecting emerging disease-exposure associations. As such, it complements the already existing classical surveillance system (MALPROF) (10). In France, the occupational disease surveillance system (RNV3P) adopts advanced data mining techniques in order to capture new meaningful occupational patterns (3).

Guo and coworkers (33) exploited Big Data as a complementary approach to the classical “behavior-based safety” (BBS) approach, in order to systematically observe, analyze and modify workers’ unsafe behaviors. Utilizing a behavioral risk knowledge base as well as images capturing workers’ behavior by means of intelligent video surveillance and mobile application, the authors developed an *ad hoc* platform and a Hadoop Distributed File System (HDFS)-based database. This information system was effective in storing images, quickly retrieving them and automatically extracting semantic information in real-time. Lenderink and colleagues (47) designed, implemented and assessed SIGNAAL (abbreviation of “*Signalering Nieuwe Arbeidsgerelateerde Aandoeningen Loket*”), a Web-based tool that enables to report emerging occupational risks and hazards. The first version of SIGNAAL was experi-

mentally used in the Netherlands and Belgium and, later, extended to other countries, including Italy. During the experimental phase, out of 11 assessed work-related exposures, one resulted as a new occupational entity, four were known but not commonly reported and encountered in the routine practice of occupational medicine, whereas other five cases were reported in new working contexts. Only one case was a well-known, established occupational disease.

“Infodemiology” (a *port-manteau* of the words “information” and “epidemiology”) and “*infoveillance*” (a *port-manteau* of the words “information” and “surveillance”) indicate the new emerging “science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform and improve public health and public policy” (28). Bragazzi and coworkers (6), using infodemiological tools, investigated occupational diseases-related interest at the level of the complex interplay between scientific community, media coverage and web behavior. Silicosis was studied as a classical prototype of occupational disease, characterized by an emerging interest in light of recent acute massive clusters. In the era of the dynamic and highly interactive Web 2.0, people increasingly surf the Internet, looking for health-related information (5). In particular, web-activities were captured by means of Google Trends (GT), whereas the media coverage was analyzed using Google News. Silicosis-related scientific production was explored by mining scholarly databases, like PubMed and Google Scholar, while the Wikipedia traffic was tracked with Wikitrends and the usage of new media by monitoring YouTube and Twitter (in terms of comments, posts, hash-tags, retweets and likes). Authors were able to detect a peak in silicosis-related web searches in the period 2010–2011, when a cluster of acute silicosis was reported, witnessing a marked public interest and engagement. On the other hand, only a small fraction of the posted/uploaded material was found to contain accurate, reliable scientific information.

In another study, Bragazzi and coauthors (7) focused on silicosis-related digital behavior in the USA, in the period 2004–2010. GT-based data were compared with “real-world” epidemiological

data of silicosis mortality obtained from the Centers for Disease Control and Prevention (CDC). Authors found statistically significant correlations with epidemiological data of silicosis ( $r=0.805$ ,  $p$ -value  $<0.05$ ), with both the temporal and the geo-spatial trends strongly correlating with the epidemiological data.

Dini et al. (23) systematically searched the term “*silicosi*” (silicosis, in the Italian language) using five top search engines and screening the first three pages of results by means of two validated readability tools (namely, the Gulpease and the ReadIt DyLanLab grade level scores). The authors found that the readability scores differed among the types of websites: academic websites differed from institutional websites, as well as encyclopedia/dictionary pages from institutional sites. Approximately half of the websites were intended/designed for workers, with only 1.4% of websites adhering to Health on the Net Foundation Code of Conduct.

Zhang and collaborators (85) utilized infodemiology approaches to monitor media reaction and coverage to the Accreditation Council for Graduate Medical Education (ACGME)’s 2017 Relaxed Resident Duty-Hour Restrictions. This allowed first-year residents to work 24-hour shifts, and as such affecting the internship experience of radiology residents (22). Systematically mining Google News, the authors found 36 relevant news articles (58% national, 22% local, and 20% medical) published over a 4-week period after the announcement. Most unfavorable articles were from national news sources, underlining that the decision would have led to sleep deprivation, medical errors, car accidents, thus compromising residents’ health and work-life balance. On the other hand, supporting data were less likely to be provided by unfavorable articles when compared with favorable news.

Only few articles have addressed the role of Big Data for capturing public understanding towards environmental issues (26, 51), suggesting that sniffing social media and social networks may be a useful proxy measurement for pollution, as a form of “electronic nose”. According to a recent survey carried out by Nissan (60), the popularity of search terms like “best air purifier” and “air quality index” both rose by a factor of 750% from August 2006 to

August 2016 whilst the phrase “air pollution facts” climbed 350% over the same period. Furthermore, the interest for “best air purifier” saw a 10-year spike in June 2016. Tao and colleagues (70) investigated how public perceptions of air quality responded to changing pollutant levels, correlating particulate matter measurements from four Chinese megacities (namely, Beijing, Shanghai, Guangzhou, and Chengdu) with 112 million posts on Weibo (a popular Chinese micro-blogging system) from corresponding days in 2011–2013. Authors were able to construct and validate an Air Discussion Index for estimating daily particulate matter based on the content of Weibo posts. In Beijing, a strong correlation ( $r=0.88$ ) between the index and measured particulate matter could be found. Similarly, Jiang and coworkers (39) analyzed the spatiotemporal trends in geo-targeted social media messages with comprehensive Big Data filtering procedures, extracting data from Sina Weibo (Chinese Twitter). The authors correlated the daily Air Quality Index released by the China’s Ministry of Environmental Protection with the social media messages during 2012. They found strong correlations.

Chen et al. (14) exploited the discipline of Web Science, as a novel tool for exploring public interest towards smog disaster and related health hazards. The authors defined and calculated ad hoc Individual Public Health Indexes for smog caused health hazard quantification, to understand people’s behavior and concerns and to inform government’s strategy design for disaster mitigation. Wang and collaborators (79), based on extensive social media and social network analysis (Sina micro-blog and Baidu Tieba) from January 2015 to June 2016, developed and computed an ad hoc Environmental Quality Index (EQI) in order “to measure and represent people’s overall attitude and sentiment towards an area’s environmental quality at a specific time; it includes not only metrics for water and food quality but also people’s feelings about air pollution”. A high sentiment analysis and classification precision of 85.67% could be obtained.

Liu and coworkers (49) used a Facebook page about Citizens’ Observatories (COs), and reviewed indicators for evaluating public interest in social media content, in terms of users’ engagement. The au-

thors found that environmental health content appealed to adults between 35-44 years of age, equally balanced between men and women. Carducci and collaborators (12) evaluated the citizen awareness and interest towards air pollution together with their positive behaviors, using different information sources (press coverage, Google searches, tweets and a survey from a sample of 598 subjects) from September 2015 to March 2016. The authors found that the media coverage concerning air pollution was very high from the end of 2015 to the beginning of 2016, this was also the case for Internet searches and Twitter messages. However, the parallel study of mass media information and people's attitudes and behaviors seemed to indicate that the high media coverage was not followed by a high motivation towards pro-environmental behaviors. Covolo and colleagues (17) investigated whether the presentation of information on some environmental health topics ("nuclear energy", "electromagnetic waves", "air pollution", "waste", and "radon") differed among various search engines (Google, Yahoo!, Bing, Ask, and AOL). The authors reported variable results when surfing the Internet on different environmental health topics. Furthermore, Big Data-based analysis can also verify the impact of environmental campaigns. For instance, Leas and coworkers (46) evaluated how Leonardo DiCaprio's 2016 Oscar acceptance speech citing climate change, motivated global English language news, social media and information seeking about climate change. The authors found that tweets including the terms "climate change" or "global warming" increased by 636% (the so-called "DiCaprio effect"), even surpassing the daily average effect of the 2015 Conference of the Parties and the Earth Day effect by a factor of 3.2 and 5.3, respectively. At the same time, Google searches for "climate change" or "global warming" increased by 261% and by 210%.

Social media could play an important role in occupational medicine, even though their use is rather limited and overlooked. Harber and Leroy (36) have recently reviewed the potential roles of social media in the field of occupational safety and health (with a focus on occupational lung diseases), in terms of information dissemination, peer-to-peer communication, survey research data collection, participatory

research, exposome data acquisition, assessment and monitoring of public concerns, and knowledge generation.

#### **BIG DATA AT A CITIZEN LEVEL: THE ROLE OF PARTICIPATORY RESEARCH IN THE FIELD OF OCCUPATIONAL SAFETY AND HEALTH**

The role of citizens is growing in both the post-modern society and science. Community-based participatory research (CBPR) or shared research or citizen science is a bridge between producers and users of knowledge, which is being constantly co-created and (re-)used (42). For instance, in the field of occupational safety and health, the "Safe-cast group" (details available at <http://blog.safecast.org/>) is an example of a citizen sensing project, the purpose of which is the collection of environmental data after the 2011 earthquake in Japan that led to the Fukushima radiation fallout (63). Moreover, CBPR can be effective in identifying emerging risk and hazard factors (34) and influencing the process of formulating policies (31, 43, 55, 56). For instance, the NIOSH launched a project aimed at improving health and safety for low-income elderly and disabled persons and their in-home care workers. Community partners and various stakeholders participated in focus groups, stakeholder interviews, and meetings; they played multiple roles including identifying organizational policy changes the partners could initiate immediately, as well as broader public policy goals (31). These represent valuable examples of community-based participatory research.

#### **DISCUSSION**

In the field of occupational medicine, Big Data-based approaches can be used to collect, store, distribute, manage, and analyze a high volume of work-related epidemiological, clinical, and administrative data. These approaches can extract relevant pattern and information and convert Big Data into smart, actionable data. Big Data are expected to provide opportunities that will have a positive impact on occupational health (15).

They can be generated by different work-related sources. Workplace health informatics deals with



the collection and handling of this mass of data, concerning the interface between the work environment and the general health status, both at the level of the individual worker and of the working population overall. Integrating this large volume of data could, indeed, provide important insights into the interactions between workers and workplace conditions, including organizational, environmental and psycho-social aspects, and their effects on health outcomes.

Databases represent a major source of Big Data: the use of sophisticated techniques such as data mining or text mining enables researchers to extract relevant information, identifying statistically significant and regular patterns (Big Data analytics). Proper predictive algorithms derived from artificial intelligence and deep learning can enable simulation of future trends (data now-casting/forecasting), in principle allowing the design and implementation of *ad hoc* training and preventive interventions.

Big Data can also be used to monitor the impact and the effectiveness of occupational policies, as well as knowledge concerning work-related health and safety measures and initiatives. Moreover, they can be exploited to inform and design *ad hoc* health prevention programs paving the way for “precision occupational medicine”.

Big Data, either molecular, digital or computational, have some strengths but also suffer from a number of limitations (44, 68). Among the opportunities, they provide quick, reliable and accurate assessment of the effects of exposure to environmental and occupational hazards and risks (19, 77). For example, systems biology, and computational toxicology approaches can adequately inform the occupational exposure limit setting process. Furthermore, new reporting and surveillance systems can complement the classical ones, which are, often, plagued by inconsistencies, missing or inaccurate data and reporting delay.

Among the drawbacks, Big Data and Big Data sources are highly heterogeneous and should be effectively integrated and harmonized together. In this regard, recently, the “European Centre for Ecotoxicology and Toxicology of Chemicals” (ECE-TOC) has issued guidelines on a Good-Laboratory Practice (GLP)-like context for collecting, storing,

curating, processing and integrating omics data (9). Further multi-center network initiatives can foster the usage of Big Data, also establishing new standards by leveraging and enhancing the current ones.

Patient and information privacy represents another major challenge: indeed, significant efforts should be done in order to preserve and protect privacy, confidentiality and identity. The emerging field of “Big Data Ethics” is trying to address all these issues (48, 54).

Another important technical/computational issue is to transform the enormous wealth of occupational medicine-related data from “passive Big Data” into “active, dynamic Smart Data” (25, 73). This requires new and sophisticated technologies for data transfer and physical storage space, as well as properly trained human resources to implement the transition from Big Data-based knowledge to action. This implies, at least in the short term, a substantial financial investment, even though, in the long run, more efficient and effective healthcare systems will contribute to cost reduction and savings.

Occupational clinicians should become aware of the topics most frequently searched by patients and proactively address these concerns during the medical examination. Institutional bodies and organisms should be more present and active in digital tools and media to disseminate and communicate scientifically accurate information. In this way, the goals of promoting good health in the workplace, keeping people healthy and active for longer and, thus, positively impacting on productivity and competitiveness, can be achieved (16, 50).

The networking of Occupational Physicians and of the Occupational Health Services (OHS) can enable the collection and sharing of occupational health and safety related data in a systematic and homogeneous way, with promising and challenging perspectives also for applied research. In this context, scientific Societies must play a major role in order to guarantee appropriate quality and analysis of collected data and their dissemination. This could, in turn, allow the optimal utilization of advanced analytical approaches, which are essential in order to provide both a wide and reliable epidemiological picture, updated over time, of working population health, and useful insights regarding interac-

tion between workers and the work “environment”. Furthermore, well-designed occupational health and safety strategies, developed on a national basis, can also provide meaningful contributions towards reaching important public health goals (65, 66).

NO POTENTIAL CONFLICT OF INTEREST RELEVANT TO THIS ARTICLE WAS REPORTED BY THE AUTHORS

## REFERENCES

- Baldacci S, Maio S, Cerrai S, et al: Allergy and asthma: Effects of the exposure to particulate matter and biological allergens. *Respir Med* 2015; 109: 1089-1104
- Bevilacqua M, Ciarapica FE, Giacchetta G: Industrial and occupational ergonomics in the petrochemical process industry: A regression trees approach. *Accid Anal Prev* 2008; 40: 1468-1479
- Bonnetterre V, Bicout DJ, Larabi L, et al: Detection of emerging diseases in occupational health: usefulness and limitations of the application of pharmacosurveillance methods to the database of the French National Occupational Disease Surveillance and Prevention network (RNV3P). *Occup Environ Med* 2008; 65: 32-37
- Bradburne C, Graham D, Kingston HM, et al: Overview of ‘Omics Technologies for Military Occupational Health Surveillance and Medicine. *Mil Med* 2015; 180 (10 Suppl): 34-48
- Bragazzi NL: From P0 to P6 medicine, a model of highly participatory, narrative, interactive, and “augmented” medicine: some considerations on Salvatore Iaconesi’s clinical story. *Patient Prefer Adherence* 2013; 7: 353-359
- Bragazzi NL, Dini G, Toletone A, et al: Leveraging Big Data for Exploring Occupational Diseases-Related Interest at the Level of Scientific Community, Media Coverage and Novel Data Streams: The Example of Silicosis as a Pilot Study. *PLoS One* 2016; 11: e0166051
- Bragazzi NL, Dini G, Toletone A, et al: Infodemiological data concerning silicosis in the USA in the period 2004-2010 correlating with real-world statistical data. *Data Brief* 2016; 10: 457-464
- Brody E, Gold L, Mehan M, et al: Life’s simple measures: unlocking the proteome. *J Mol Biol* 2012; 422: 595-606
- Buesen R, Chorley BN, da Silva Lima B, et al: Applying ‘omics technologies in chemicals risk assessment: Report of an ECETOC workshop. *Regul Toxicol Pharmacol* 2017; 91 Suppl 1: S3-S13
- Campo G, Papale A, Baldasseroni A, et al: The surveillance of occupational diseases in Italy: the MALPROF system. *Occup Med (Lond)* 2015; 65: 632-637
- Carder M, Bensefa-Colas L, Mattioli S, et al: A review of occupational disease surveillance systems in Modern countries. *Occup Med (Lond)* 2015; 65: 615-625
- Carducci A, Donzelli G, Cioni L, et al: Air pollution: a study of citizen’s attitudes and behaviors using different information sources. *Epidemiology Biostatistics and Public Health* 2017; 14: e12389
- Cavalleri T, Angelici L, Favero C, et al: Plasmatic extracellular vesicle microRNAs in malignant pleural mesothelioma and asbestos-exposed subjects suggest a 2-miRNA signature as potential biomarker of disease. *PLoS One* 2017; 12: e0176680
- Chen J, Chen H, Zheng G, et al: Big smog meets web science: smog disaster analysis based on social media and device data on the web. In: *Proceedings of the companion publication of the 23rd international conference on World wide web companion, 2014*: 505-510
- Ciarapica FE: Big Data Analytics as a Tool for Reducing Ergonomics Risk. *J Ergonomics* 2016; 7: e164
- Copello F, Garbarino S, Messineo A, et al: Occupational Medicine and Hygiene: applied research in Italy. *J Prev Med Hyg* 2015; 56: E102-10
- Covolo L, Filisetti B, Mascaretti S, et al: A web search on environmental topics: what is the role of ranking? *Telemed J E Health* 2013; 19: 967-972
- Cullinan P, Muñoz X, Suojalehto H, et al: Occupational lung diseases: from old and novel exposures to effective preventive strategies. *Lancet Respir Med* 2017; 5: 445-455
- DeBord DG, Burgoon L, Edwards SW, et al: Systems Biology and Biomarkers of Early Effects for Occupational Exposure Limit Setting. *J Occup Environ Hyg* 2015; 12 Suppl 1: S41-54
- Delaunay M, Van der Westhuizen H, Godard V, et al: Use of GIS in visualization of work-related health problems. *Occup Med (Lond)* 2015; 65: 682-692
- DelRaso NJ, Chan VT, Mauzy CA, et al: Air Force Research Laboratory Integrated Omics Research. *Mil Med* 2015; 180 (10 Suppl): 67-75
- DeRienzo CM, Frush K, Barfield ME, et al: Handoffs in the era of duty hours reform: a focused review and strategy to address changes in the Accreditation Council for Graduate Medical Education Common Program Requirements. *Acad Med* 2012; 87: 403-410
- Dini G, Bragazzi NL, D’Amico B, et al: A reliability and readability analysis of silicosis-related Italian websites: implications for occupational health. *Med Lav* 2017; 108: 167-173
- Drury CG: Human factors/ergonomics implications of Big Data analytics: Chartered Institute of Ergonomics and Human Factors annual lecture. *Ergonomics* 2015; 58: 1-15
- Duong TH, Nguyen HQ, Jo GS: Smart Data: Where

- the Big Data Meets the Semantics. *Comput Intell Neurosci* 2017; 2017: 6925138
26. Durmuşoğlu ZDU: Using Google trends data to assess public understanding on the environmental risks. *Human and Ecological Risk Assessment: An International Journal* 2017; 23: 1968-1977
  27. European Agency for Safety and Health at Work: Outlook 1 – New and emerging risks in occupational safety and health. 2009, Luxembourg: Office for Official Publications of the European Communities.
  28. Eysenbach G: Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009; 11: e11
  29. Feltner C, Peterson K, Palmieri Weber R, et al: The Effectiveness of Total Worker Health Interventions: A Systematic Review for a National Institutes of Health Pathways to Prevention Workshop. *Ann Intern Med* 2016; 165: 262-269
  30. Giusti L, Da Valle Y, Bonotti A, et al: Comparative proteomic analysis of malignant pleural mesothelioma evidences an altered expression of nuclear lamin and filament-related proteins. *Proteomics Clin Appl* 2014; 8: 258-268
  31. Gong F, Baron S, Ayala L, et al: The role for community-based participatory research in formulating policy initiatives: promoting safety and health for in-home care workers and their consumers. *Am J Public Health* 2009; 99 Suppl 3: S531-8
  32. Großerueschkamp F, Bracht T, Diehl HC, et al: Spatial and molecular resolution of diffuse malignant mesothelioma heterogeneity by integrating label-free FTIR imaging, laser capture microdissection and proteomics. *Sci Rep* 2017; 7: 44829
  33. Guo SY, Ding LY, Luo HB, et al: A Big-Data-based platform of workers' behavior: Observations from the field. *Accid Anal Prev* 2016; 93: 299-309
  34. Gutberlet J, Baeder AM, Pontuschka NN, et al: Participatory research revealing the work and occupational health hazards of cooperative recyclers in Brazil. *Int J Environ Res Public Health* 2013; 10: 4607-4627
  35. Hamad R, Modrek S, Kubo J, et al: Using "Big Data" to capture overall health status: properties and predictive value of a claims-based health risk score. *PLoS One* 2015; 10: e0126054
  36. Harber P, Leroy G: Feasibility and Utility of Lexical Analysis for Occupational Health Text. *J Occup Environ Med* 2017; 59: 578-587
  37. Harrison J, Dawson L: Occupational Health: Meeting the Challenges of the Next 20 Years. *Saf Health Work* 2016; 7: 143-149
  38. Hegmans JP, Bard MP, Hemmes A, et al: Proteomic analysis of exosomes secreted by human mesothelioma cells. *Am J Pathol* 2004; 164: 1807-1815
  39. Jiang W, Wang Y, Tsou MH, et al: Using social media to detect outdoor air pollution and monitor air quality index (AQI): a geo-targeted spatiotemporal analysis framework with Sina Weibo (Chinese Twitter). *PLoS One* 2015; 10: e0141185
  40. Johnson CH, Athersuch TJ, Collman GW, et al: Yale school of public health symposium on lifetime exposures and human health: the exposome; summary and future reflections. *Hum Genomics* 2017; 11: 32
  41. Ju L, Chen JQ, Jiang ZQ, et al: Screening of differentially expressed serum proteins in patients with asbestosis. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi* 2012; 30: 408-412
  42. Jull J, Giles A, Graham ID: Community-based participatory research and integrated knowledge translation: advancing the co-creation of knowledge. *Implement Sci* 2017; 12: 150
  43. Katz-Kimchi M, Manosevitch I: Mobilizing Facebook users against Facebook's energy policy: The case of Greenpeace Unfriend Coal campaign. *Environmental Communication* 2015; 9: 248-267
  44. Kruse CS, Goswamy R, Raval Y, et al: Challenges and Opportunities of Big Data in Health Care: A Systematic Review. *JMIR Med Inform* 2016; 4: e38
  45. Larkin A, Hystad P: Towards Personal Exposures: How Technology Is Changing Air Pollution and Health Research. *Curr Environ Health Rep* 2017; 4: 463-471
  46. Leas EC, Althouse BM, Dredze M, et al: Big Data Sensors of Organic Advocacy: The Case of Leonardo DiCaprio and Climate Change. *PLoS One* 2016; 11: e0159885
  47. Lenderink AF, Keirsbilck S, van der Molen HF, et al: Online reporting and assessing new occupational health risks in SIGNAAL. *Occup Med (Lond)* 2015; 65: 638-641
  48. Lipworth W, Mason PH, Kerridge I: Ethics and Epistemology of Big Data. *J Bioeth Inq* 2017; 14: 485-488
  49. Liu HY, Eleta I, Kobernus M, et al: Analysis of Public Interest in Environmental Health Information: Fine Tuning Content for Dissemination via Social Media. In Satsiou A, et al. (eds): *Collective Online Platforms for Financial and Environmental Awareness*. IFIN 2016, ISEM 2016. Lecture Notes in Computer Science, vol 10078. Springer, Cham
  50. Manzoli L, Sotgiu G, Magnavita N, et al: Evidence-based approach for continuous improvement of occupational health. *Epidemiol Prev* 2015; 39: 81-85
  51. McCallum ML, Bury GW: Google search patterns suggest declining interest in the environment. *Biodiv Conserv* 2013; 22: 1355-1367

52. McCreanor J, Cullinan P, Nieuwenhuijsen MJ, et al: Respiratory effects of exposure to diesel traffic in persons with asthma. *N Engl J Med* 2007; 357: 2348-2358
53. Mehan MR, Ostroff R, Wilcox SK, et al: Highly multiplexed proteomic platform for biomarker discovery, diagnostics, and therapeutics. *Adv Exp Med Biol* 2013; 735: 283-300
54. Milton CL: The Ethics of Big Data and Nursing Science. *Nurs Sci Q* 2017; 30: 300-302
55. Minkler M, Lee PT, Tom A, et al: Using community-based participatory research to design and initiate a study on immigrant worker health and safety in San Francisco's Chinatown restaurants. *Am J Ind Med* 2010; 53: 361-371
56. Moir S: Ideological Influences on Participatory Research in Occupational Health and Safety: A Review of the Literature. *New Solutions* 2005; 15: 15-28
57. Morr  DJ, Hostetler B, Taggart DJ, et al: ENOX2-based early detection (ONCOblot) of asbestos-induced malignant mesothelioma 4-10 years in advance of clinical symptoms. *Clin Proteomics* 2016; 13: 2
58. Motta V, Angelici L, Nordio F, et al: Integrative Analysis of miRNA and inflammatory gene expression after acute particulate matter exposure. *Toxicol Sci* 2013; 132: 307-316
59. Neveu V, Moussy A, Rouaix H, et al: Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res* 2017; 45: D979-D984
60. Nissan: Internet searches on air pollution up 750% in last decade as consumers show more concern for air quality, Nissan reveals. Available at: <https://newsroom.nissan-europe.com/uk/en-gb/media/pressreleases/148821/internet-searches-on-air-pollution-up-750-in-last-decade-as-consumers-show-more-concern-for-air-qual>
61. Ostroff RM, Mehan MR, Stewart A, et al: Early detection of malignant pleural mesothelioma in asbestos-exposed individuals with a noninvasive proteomics-based surveillance tool. *PLoS One* 2012; 7: e46091
62. Plusquin M, Guida F, Polidoro S, et al: DNA methylation and exposure to ambient air pollution in two prospective cohorts. *Environ Int* 2017; 108: 127-136
63. Reis S, Seto E, Northcross A, et al: Integrating modelling and smart sensors for environmental and human health. *Environ Model Softw* 2015; 74: 238-246
64. Sepulveda MJ: From worker health to citizen health: moving upstream. *J Occup Environ Med* 2013; 55 (12 Suppl): S52-57
65. Societ  Italiana di Medicina del Lavoro e Igiene Industriale (SIMLII): Linee guida per la sorveglianza sanitaria. Piacenza, Italy: Nuova Editrice Berti, 2017
66. Societ  Italiana di Medicina del Lavoro e Igiene Industriale (SIMLII): Technical assessment. Promozione della salute nei luoghi di lavoro. Piacenza, Italy: Nuova Editrice Berti, 2011
67. Steenhof M, Janssen NA, Strak M, et al: Air pollution exposure affects circulating white blood cell counts in healthy subjects: the role of particle composition, oxidative potential and gaseous pollutants - the RAPTES project. *Inhal Toxicol* 2014; 26: 141-165
68. Stieb DM, Boot CR, Turner MC: Promise and pitfalls in the application of Big Data to occupational and environmental health. *BMC Public Health* 2017; 17: 372
69. Stoyanov SR, Hides L, Kavanagh DJ, et al: Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR Mhealth Uhealth* 2016; 4(2): e72
70. Tao Z, Kokas A, Zhang R, et al: Inferring Atmospheric Particulate Matter Concentrations from Chinese Social Media Data. *PLoS One* 2016; 11(9): e0161389
71. Tomasi SE, Fechter-Leggett ED, et al: Niosh Health Hazard Evaluation Program as a sentinel surveillance system for novel occupational lung diseases. *American Journal of Respiratory and Critical Care Medicine* 2017; 195: A3868
72. Tooker BC, Newman LS, Bowler RP, et al: Proteomic detection of cancer in asbestosis patients using SELDI-TOF discovered serum protein biomarkers. *Biomarkers* 2011; 16: 181-191
73. Trifir  G, Sultana J, Bate A: From Big Data to Smart Data for Pharmacovigilance: The Role of Healthcare Databases and Other Emerging Sources. *Drug Saf* 2018; 41: 143-149
74. Turner MC, Nieuwenhuijsen M, Anderson K, et al: Assessing the Exposome with External Measures: Commentary on the State of the Science and Research Recommendations. *Annu Rev Public Health* 2017; 38: 215-239
75. van der Schee MP, PaffT, Brinkman P, et al: Breathomics in lung disease. *Chest* 2015; 147: 224-231
76. Vineis P, Chadeau-Hyam M, Gmuender H, et al: The exposome in practice: Design of the EXPOsOMICS project. *Int J Hyg Environ Health* 2017; 220: 142-151
77. Vlaanderen J, Moore LE, Smith MT, et al: Application of OMICS technologies in occupational and environmental health research; current status and projections. *Occup Environ Med* 2010; 67: 136-143
78. Walker G, Strathie A: Big Data and ergonomics methods: A new paradigm for tackling strategic transport safety risks. *Appl Ergon* 2016; 53 Pt B: 298-311
79. Wang Z, Ke L, Cui X, et al: Monitoring Environmental Quality by Sniffing Social Media. *Sustainability* 2017; 9: 85.
80. Weber DG, Johnen G, Bryk O, et al: Identification of

- miRNA-103 in the cellular fraction of human peripheral blood as a potential biomarker for malignant mesothelioma--a pilot study. *PLoS One* 2012; 7: e30221.
81. Weber DG, Casjens S, Johnen G, et al: Combination of MiR-103a-3p and mesothelin improves the biomarker performance of malignant mesothelioma diagnosis. *PLoS One* 2014; 9(12): e114483
  82. Wild CP: Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005; 14: 1847-1850
  83. Yang M: A current global view of environmental and occupational cancers. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 2011; 29: 223-249
  84. Yang HY, Shie RH, Chang CJ, et al: Development of breath test for pneumoconiosis: a case-control study. *Respir Res* 2017; 18: 178
  85. Zhang Z, Krauthamer AV, Rosenkrantz AB: The Media Response to the ACGME's 2017 Relaxed Resident Duty-Hour Restrictions. *J Am Coll Radiol* 2018; 15 (3 Pt A): 452-457