Safety Management and Human Factors, Vol. 105, 2023, 122–131
https://doi.org/10.54941/ahfe1003077

The Power of Safety Data to Prevent Work Related Accidents: Empirical Evidences From Pilot Projects in Italy

Maria Grazia Gnoni[1], Michele Villa[2], Antonia Donvito[2], Armando Guglielmi[3], Giuseppe Campo[3], Diego De Merich[3], Daniele De Santis[3], Antonio Leva[3], and Michele Pellicci[3]

[1]Department of Innovation engineering, University of Salento, Lecce, Italy
[2]ADAM AI Solutions, Italy
[3]Department of Medicine, Epidemiology, Occupational & Environmental Hygiene, Italian National Institute for Insurance against Accidents at Work (INAIL), Italy

ABSTRACT

A vast amount of safety data is collected every year by public and private organisations for many different reasons. Often this is due to compliance requirements, other times as part of reporting practices or to measure performance towards improvement goals. Although data collection is still a critical activity – as it is usually not characterized by a standard approach –, data analysis now represents the most critical one due to several factors. The present study aims to point out how safety data at different level of aggregation - could support effective continuous improvements activities of companies as well as of institutional organizations. Case studies in the construction sectors have been analysed in order to point out potentialities in analysing structured and unstructured safety data available in a open access way. Results shows the high potential of structured safety data models and tools for acquiring knowledge to improve prevention activities for safety at work.

Keywords: Safety data, Structured and unstructured data, Injury prevention, Digital technologies

INTRODUCTION

A vast amount of safety data is collected every year by public and private organisations for many different reasons. Often this is due to compliance requirements, other times as part of reporting practices or to measure performance towards improvement goals. Although data collection is still a critical activity – as it is usually not characterized by a standard approach –, data analysis now represents the most critical one due to several factors. One example is the loss of knowledge contents for preventing accidents derived by a non-effective (or absent) analysis of audit results, near miss recording, observations, but also accident investigations. Tools and methods derived from digital technologies, starting from artificial intelligence applications to big data and machine learning techniques, through predictive and prescriptive analytics, can provide valuable support for improving effectiveness of accident prevention activities in the public as well as in the private sectors

(De merich et al, 2020). The study focuses on analysing possible contributions of these tools by evaluating pilot research project activities based on analysing safety data at different level of aggregation. Firstly, the potential of using structured and unstructured data to understand real causes of accidents/injuries based on historical data has been analysed trough a quantitative analysis in the construction sector. The analysis looked at several factors ranging from person injured data, nature of the event, circumstances, causes, etc. to find most relevant patterns and elements that could lead to injury. A comparison with similar US databases (derived from Bureau of Labour Statistics, BLS) has been carried out to compare Italian versus US data for the same industrial sector. The work, which included advanced statistics and Natural Language Processing (NLP) applied to event description, highlighted which sub-sectors are more prone to accidents in the construction industry and identified falls, contact with objects and overexertion and bodily reaction as the most frequent type of accident. The Italy/US comparison presented similar conclusion when analysing the datasets. Next, a pilot project in a construction site demonstrating a possible use of structured and unstructured data acquired through digital technologies in real time has been developed. The paper is organized as follows: the first section describes sample features and some basic statistics about the Italian dataset; cross analysis over structured and unstructured data are discussed in section 2; next, in section 3, a content analysis has been developed over two datasets, one about injury occurred at construction sites in Italy and one for USA. Finally, in section 4, a pilot project aiming to collect and analyse safety data in real time is described.

A STATISTICAL LINEAR ANALYSIS OF SAFETY DATA IN THE ITALIAN CONTRUCTION SECTOR

The construction sector is an important part of the Italian workforce, with about 8% of total employees. Despite the downward trend, national statistics data show that the sector is still strongly exposed to accidents, even fatalities. In particular, construction sites are a place of particular risk for workers: about 40% of accidents took place here. The first part of this study aims to point out current level of hazard of the Italian sector by simply analysing safety data available at national level. The first step was to critically analyse two datasets provided by INAIL – the Italian National Institute for Insurance against Accidents at Work - on non-severe (more than 160.000 records) and severe/fatal (almost 600 records) injuries in the construction sector in Italy between 2015 and 2019. INAIL is a public non-profit entity safeguarding workers against physical injuries and occupational diseases. Over the course of the 120 years elapsed since its establishment, INAIL has progressively changed its tasks, providing now an integrated system of protection for victims of work-related injuries or occupational diseases, including preventive and research actions at the workplace, medical services, financial support, rehabilitation and reintegration to social life and work. Furthermore, INAIL is also liable for collecting statistics of work-related injuries and occupational diseases. For this study the data used were extracted from the following INAIL databases (dbs):

• The so *called "Statistical db",* which provides, in an aggregate manner, data and information on multiple aspects of the injuries phenomenon collected in Italian Companies. The statistical data formation chain originates from the collection throughout the national territory of administrative data relating to each individual case of injuries (Campo et al, 2020), which are then validated, certified and subsequently uploaded to the db;

• The so called *Infor.Mo db, the Italian national surveillance system for occupational fatal injuries*, which provides information about root causes of the specific injury. It is developed by INAIL in cooperation with Regions and Autonomous Provinces, and Local Health & Safety Departments (LHSDs), which are the centres of administrative operations related to public healthcare in Italy under the National Healthcare Service. The database is based on information derived from the injury investigation activity carried out by LHSDs when a fatal injury occurs at the workplace. The reconstruction of the accident dynamics follows the classic backward path used in the judicial investigative process. Starting from the last event in chronological order (i.e. the biological damage), the multi-factorial model identifies the accident that occurred (e.g. a fall from a height of the worker or the overturning of a work vehicle) and the related causes (i.e. risk factors) of the specific injury.

By overlapping information from the two databases, it is possible to obtain both quantitative and detailed information on injuries, allowing a most effective and reliable evaluation of their actual causes. Results due to this activity are described as follows. Firstly, collected data about non-severe injuries have been analysed; data refer to the Italian construction sector. Some relevant features characterized the analysed sample are reported as follows:

- 97% of the injured person are male; only 3% are female;
- At the time of the accident, about 60% of injured workers were between the ages of 30 and 60. In detail, the group most exposed to accidents is the 40–50 age group (30% of injured people); followed by the 50–60 age group (about 25%).

Next, an analysis of specific activity carried out by people involved in the injury has been developed: this is a relevant topic as the aim is to point out most hazardous activities in order to plan specific initiatives aiming at reducing most critical risks related to these activities. Results are reported in Table 1.

Next, a crossing analysis about the specific activity carried out when the injury occurred has been developed. Data outline interesting results: 30% of total accidents occurred during the performance of **manual activities** - such as picking up, grasping, tearing, holding in the hand, laying - on a **horizontal plane**, or **working with hand tools**, or in **handling activities** (e.g. moving heavy loads, handling objects). Thus, these activities could represent the most hazardous ones for workers as they represent very common activities in the construction sector. Next, in 20% of data, the worker was engaged in a movement activity (walking, running, going up, getting off, etc.), 7% refers to accidents occurred both driving of a means of transport or an motorized handling equipment and during vertical transportation, lifting, raising, bringing down an object from hight.

Table 1. Distribution of data about injury events based on type of profession.

Professional Identification code	Occurred injury events evaluated in the total sample	Additional notes
Bricklayer	18%	Very critical activities refer to using specific materials as brick, stone or concrete materials
Manual workers	10%	A critical group is construction workers, who carried out make up 87% of total data in this cluster
Electrician	9%	A critical activity highlighted is outdoor installations, which contributes to 52% total data in this cluster
Carpenter	6%	A critical activity highlighted is construction, which contributes to 73% total data in this cluster
Plumbing	5%	-

CROSSING STRUCTURED AND UNSTRUCTURED SAFETY DATA TO UNDERSTAND REAL CAUSES OF INJURIES

After analysing single data about injuries (reported in the previous section), a multi-variate analysis looked at possible combinations of factors to identify specific "hidden patterns" has been carried out aiming to point put actual as well as interconnected causes of occurred accidents. These results could represent a first attempt towards predicting what can possibly happen when a combination of factors is detected and, consequently, prevent accidents before they occur.

The first cross analysis has been carried out by overlapping data about specific sub-sector inside the construction one. In detail, three main sub-sectors in the have been analysed, such as building construction, civil Engineering and specialised Construction Work: results outline as walking, running, climbing, descending, are the most frequent activities (about 22% of cases) that workers were carrying out when the accident occurred. Next, a deeper analysis about most frequent activities in each specific subsector have been carried out. In specialized construction work and civil engineering, the second most critical activity is an operation developed with a transport means or a specific movement equipment. On the other hand, in the construction of buildings, the vertical transport of an object is the second most frequent activity (8% of data). In addition, in civil engineering, accidents during manual use of tools are significantly less frequent than in the construction of buildings and specialized construction work.

The second cross analysis has been carried out by evaluating the three most exposed professions previously outlined (i.e. bricklayers, manual workers and electricians). Results outline that the most frequent activity carried out at the time of the accident is

- Construction for brick layer;
- Excavation, construction, maintenance and demolition of buildings for manual workers ;
- Moving activities- with or without means of transport - for electricians.

Finally, an analysis of potential deviations from planned procedures that have contributed to injury occurrence has been developed in order to evaluate if common paths could be outlined over different sectors. Results shows that similarities as for bricklayers, manual workers as well as electricians deviations mainly referrers to "uncoordinated and unexpected movements". However, it must be noted that the second most frequent outlined deviation is "slipping or tripping with a fall in person at the same level" which is outlined only for bricklayers and manual workers. Instead, for electricians, more frequent outlined deviations refer to falls from height, loss of control of vehicles / equipment, slips without falling and only later falling to the same level.

APPLYING NATURAL LANGUAGE PROCESSING TECNIQUES FOR EVALUTING SAFETY DATA: ITALIAN VERSUS USA INJURY DATA

After analysis Italian data, a comparison between this information and similar one referring to USA has been carried out in order to enlarge the sample in analysis, and consequently try to validate in a more reliable way obtained results. A comparison with similar US databases has been carried out to compare Italian versus US data for the same industrial sector, i.e., the construction sector.

The analysis has been developed by using advanced statistics and Natural Language Processing (NLP) applied to each event descriptions in the specific dataset. In recent years, NLP has been applied for effectively analysing large set of data about injuries at workplace. Tixier et al., (2016) proposed a content analysis for extracting precursors and outcomes from unstructured injury reports in the construction sector. Baker et al, (2020) applied NLP together with other tools (like Machine learning techniques) to point out the textual patterns useful to evaluate valid injury precursors. Liu et al.. (2021) applied NLP models to identify contributing factors under specific failure types in pipelines accidents. The activity has been developed by using the Infor.Mo db for Italy: the time period in analysis is from 2015 and 2019, and 589 accidents regarding fatal injuries are included. For USA, databases derived from Bureau of Labour Statistics (BLS) has been used (OSHA, 2021): data are analysed rage from 1/1/2015 to 11/30/2021.

NLP has been applied to text description of each event in the Italian db in order to better understand its actual causes and related dynamics. In detail, the most frequent words detected are "earth, fall, height, ladder, floor, construction site, roofing, protection, scaffolding". By evaluating word combinations, the most frequent are "He lost balance", "Fracture of multiple sites", "Fall from height", Risk of falling from height", "Trauma", "Construction site", "Civil dwelling", "Seat belt", "Reinforced concrete", "Construction company", "He fell to the ground". In addition, a crossing

analysis between most frequent words with respect to the parameter "place", is developed. For construction sites, "to fall" and "fall", "earth", "dash", "roof" and "scaffolding" are the most outlined frequent words. For other places, "to fall" and "fall", "height", "stairs", "metallic", "electric" are the most reported frequent words. The analysis by words combination allows for a deeper analysis of the events and the understanding of their real root causes.

By comparing data regarding USA with Italian ones, some common features are emerging even if the organizational models of the two countries in the construction sector are not so similar. In detail, similarly to Italy, a strong prevalence of accidents due to falls emerges, towards lower levels. This is confirmed by the emergent use of expressions such as "fall from above", "fall to the ground", "fall", in the textual description of Italian accidents dynamics. In the USA dataset, a similar emergent topic is "fall to lower level"; in addition, it is very widespread the word "fall" in the descriptive section of the accident as it is used in 46% of the descriptions. Similar condition is outlined for "contact with objects", in particular "crushing due to falling objects". Less relevant terms but representative in both datasets are accidents due to direct or indirect contact with electricity. Another interesting result refers to the keywords "roof" and "ladders", which are characterized by similar frequencies in two different datasets. In detail, the word roof is used in 15% of Italian accident descriptions compared to 9% of USA ones. The word "ladders" is used in the description of 17% and 13% respectively for Italian and USA datasets.

Finally, based on overall comparison results, it could be noted that be said that accidents in the construction sector in the USA and Italy have similar characteristics. In detail, in both cases the specialized construction sector is the most exposed to accidents, while the civil engineering sector is the least exposed. In addition, the most frequent types of accidents are also similar, both with respect to the entire construction sector and within subsectors. It emerges that, both in the USA and in Italy, falls, contact with objects and physical effort are the most frequent type of fatal accidents. In the case of the civil engineering subsector, accidents involving means of transport are also among the most frequent ones.

PREDICTIVE ANALYTICS TROUGH SAFETY DATA MANAGEMENT

This last part of the study is related to applying the safety data analysis to practical incident/injury prevention solutions through the intensive use fo digital technologies. The analysis focused on evaluating the effectiveness of a prototype risk profiling cloud-based tool called PRESTO (PREdictive Safety Tool) developed by ADAM AI solutions. The tool, using past safety data integrated with real time information collected by wearables (e.g. smart watch) and other IoT systems applied to workers as well as in the working area, allows to estimate the probability of incidents under certain circumstances (predictive analytics) and, consequently, recommends actions (prescriptive analytics) especially to prevent them e.g. by interrupting high hazardous

human activities. The system has been tested in a pilot project at a construction site for building railway in Northern Italy. The application of the tool aimed to support the "zero accident" strategy adopted by all companies involved in the construction site.

In August 2021, following preliminary talks on the application of digital technologies for the prevention of accidents in the workplace and in particular in construction sites for the construction of railway infrastructure, the companies Italferr S.p.A. - an Italian company in railway construction sector - in the (technical subject on behalf of the client Trenitalia S.p.A.), Salc S.p.A. (contractor) Elettri-Fer s.r.l. (executing company) and ADAM AI Solutions signed an agreement for the testing of PRESTO at the Italferr shipyard of 'Executive Design of the IMC Milano Martesana Enhancement - Construction of the New ETR 1000 Shed. The experimental activity started by acquiring all safety documents and analyses regarding risk assessment of the specific construction site, was developed for 6 months at the construction site and involved 7 workers of different companies working in the construction site (i.e. one contractor and one executor). In detail, each worker was equipped with mobile phones for dynamic interaction with the PRESTO platform and smart watches for the collection of biometric data that contributed to the definition of real time risk indices. The collection and recording of personal data were carried out completely in compliance with privacy regulations. From one hand, a training period was necessary for communicate all features and capability of the prototype platform to workers; on the other hand, the experimental period was also useful to understand the usability of the tool and redesign the prototype in order to be more user friendly to workers.

Several relevant results could be outlined at the end of the experimentation period. First of all, an evaluation of the overall level of "perceived usefulness" together with the acceptance level of workers for digital solutions aimed to improve safety at their workplace. The issue is becoming increasingly critical as digital technologies are wide spreading in all processes of a company, especially for improving effectiveness of health and safety management. One example is how artificial intelligence and *machine learning* tools are perceived, and therefore used at best for specific prevention purposes. After an ex-post analysis carried out on workers during the experimentation phase, some interesting remarks could be outlined. Outcomes reported a good level of appreciation of the tool and an ease of use. Some criticalities referred to the use of such technological tools - i.e. the adopted smart phone seemed to be excessively bulky - and reporting precursor events (near misses, unsafe conditions). Positive feedbacks were focused on the increase in awareness and perception of risk by the workers, which is one of the primary target of the proposed tool. The experimentation activity, even with limitations due to time and size of the pilot application, has provided important indications not only for the improvement of the tool, but also for evaluating the impact of digital technologies at workplace, thus confirming the importance of using these solutions to support the prevention of accidents at the workplace.

CONCLUSION

The study has proposed an analysis of how safety data – at different level of aggregation – could support additional knowledge to prevent in a more efficient way accidents at workplace, especially in the construction sector. The first statistical analysis has outlined outline that workers more exposed to risks are the pivotal worker roles of the sector, such as bricklayers, manual workers, plumbers and electricians. The analysis has also provided information about most frequent dynamic of accident, as almost a third of accidents occur due to contact with objects or machinery, furthermore, data outline also critical activities as accidents occur during manual activities or in the material transportation. An interesting result is that "interaction with other objects" does not represent a high critical activity as data analysed highlighted that injuries occur frequently during movement activities, mainly due to incorrect actions of the worker - such as slips, missteps - whether or not they result in falls. From this point of view, the analysis of serious and fatal accidents has shown that in 40% of cases the accident is determined by the activity of the worker himself, greater than the influence of the specific working environment as well as the adequacy of personal protective equipment (PPE). This is a very interesting result as it outlines where priorities of intervention must be direct by companies. Next, the comparison with US and Italian data has confirmed several results just outlined by the cross analysis developed with Italian data: one example is the results outlined for incorrect movements as the one of major causes of accidents in both countries. It has to be noted that this does not indicate that workplaces are always suitable and safe for workers: the remaining 60% of serious or fatal accidents are in fact caused by factors not strictly related to the worker. More deeper data analysis outlines that the words ladder/staircase and roof have similar frequencies in the accident description in the analysed datasets. Main other causes refer to the absence of adequate PPE or its incorrect use: in Italy this group represents 15% of the total serious or fatal accidents. In addition to objects and environmental aspects, vehicles and transportation equipment can also pose an element of risk in the construction sector, especially during driving activities or during a loss of vehicle control. The proposed cross analysis has also outlined that although one of the main cause of the accident may lie not only in the worker's own activity but other contributing factors must be evaluated for assessing actual root causes. Considering the centrality of these behaviours in the determination of accidents, it may be useful, from a preventive point of view, to investigate these aspects.

Finally, the description of the pilot project regarding the adoption in a real test case of several digital technologies adopted for both acquiring data about workers health, workplace conditions and analysing information for providing real time feedbacks has pointed out positive potentialities and criticalities of managing safety data in real time. One main obtained result refers not only on the effectiveness of the specific digital tool, but also on the evaluation of the perception of workers in adopting these technologies in their working activity. It has to be noted that at the centre of any experimentation and application of these technologies is the worker and his interaction with

the environment that surrounds him, with other workers, with machinery and tools present in the workplace. It is therefore essential that the worker understands the functioning of these technologies and, also through the satisfaction he expresses, uses them to his advantage (and others) for prevention.

Further developments will be oriented from one side to amplify the datasets for improving the reliability of the comparison USA/Italy also involving different industrial sectors aiming to bring out the intrinsic risks of the specific sector and to identify any more specific patterns characterizing each single country and/or both countries. On the other side, further development will be oriented in evaluating large scale application of the proposed tool in order to validate models and methods in larger application. The use of digital technologies for prevention is the new "frontier" in terms of further improving safety performance. While new technologies do not replace processes, management systems, training activities, they integrate with what already exists at the organizational and site level and leverage the availability of technological tools such as wearable tools, sensors, proximity devices that increase the potential for risk perception.

ACKNOWLEDGMENT

This paper presents some of the results of the project CONDIVIDO, BRIC ID 01/2019 funded by INAIL.

REFERENCES

Baker, H., Hallowell, M. R., & Tixier, A. J. P. (2020). Automatically learning construction injury precursors from text. Automation in Construction, 118, 103145.

Bureau of Labor Statistics (BLS). Dataset: Occupational Injuries and Illnesses and Fatal Injuries Profiles. Parameters: year from 2015 to 2019; industry NAICS 23*; states all U. S; table Case and Demographic Incidence Rates. Data available at: https://data.bls.gov/gqt/

Campo G, Cegolon L, De Merich D, Fedeli U, Pellicci M, Heymann W C, Pavanello S, Guglielmi A, Mastrangelo G. The Italian National Surveillance System for Occupational Injuries: Conceptual Framework and Fatal Outcomes, 2002–2016. International Journal Environmental Research Public Health, 2020, 17(20), 7631; https://doi.org/10.3390/ijerph17207631 (registering DOI).

De Merich, D., Gnoni, M. G., Malorgio, B., Micheli, G. J. L., Piga, G., Sala, G., Tornese, F., 2020. A Cloud-Based Tool for Integrating Occupational Risk Assessment Within Management Systems for SMEs. Safety 6, 47. https://doi.org/10.3390/safety6040047

De Merich D, Gnoni MG, Guglielmi A, Micheli GJ, Sala G, Tornese F, Vitrano G. Designing national systems to support the analysis and prevention of occupational fatal injuries: Evidence from Italy. Safety Science, article 105615, vol. 147/2022. DOI: https://doi.org/10.1016/j.ssci.2021.105615

Occupational Safety and Health Administration (OSHA), 2021. Dataset: Severe Injury Reports Parameters: industry NAICS 23*; states all U. S. Data available at: https://www.osha.gov/severeinjury

Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (Inail). Dataset Inail Flussi e Informo, Parametri: anni 2015-2019; Settori Codice Ateco 2007 42*. In Italian.

Istituto Nazionale di Statistica (Istat). Dataset: Rilevazione forze lavoro. Parametri: Italia; anni 2017, 2020; Settori Codice Ateco 2007 42*. In Italian.

Liu, G., Boyd, M., Yu, M., Halim, S. Z., & Quddus, N. (2021). Identifying causality and contributory factors of pipeline accidents by employing natural language processing and text mining techniques. Process Safety and Environmental Protection, 152, 37–46.

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. Automation in Construction, 62, 45–56.