Business Intelligence for the Analysis of Industrial Accidents Based on MHIDAS Database

Antonio J. Nakhal A.*[a], Riccardo Patriarca[a], Giulio Di Gravio[a], Giacomo Antonioni[b], Nicola Paltrinieri[c]

[a]Department of Mechanical and Aerospace Engineering, Sapienza University, Rome (ITALY)
[b]Department of Civil, Chemical, Environmental, and Materials Engineering, Alma Mater Studiorum – University of Bologna (ITALY)
[c]Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim (NORWAY)

nakhalakel.1836316@studenti.uniroma1.it

Reducing the frequency and severity of accidents in industrial processes is a continuous open challenge. Learning from previous events represents a crucial instrument to ensure an improved design of industrial plants, especially considering the complexity arising in everyday operations. This article is grounded on a database of industrial accidents involving hazardous substances and materials. The Major Hazard Incident Data Service (MHIDAS) was developed in 1986 by the Health and Safety Executive (HSE) to provide a reliable source of data on major hazard incidents and to learn for the past accidents. The database has more than 9000 accident reports covering the periods from 1950 to the end of the 1990s caused by hazardous substances/materials. This paper aims are to provide an understanding of MHIDAS data through quantitative analyses that can be obtained by exploiting the information collected through appropriate data management tools. Therefore, Information Technology (IT) services such as Business Intelligence (BI) tools have been used in this research. The paper describes the process of creating a BI model for data management on MHIDAS database to generate useful information on previous industrial safety events, allowing a detailed search engine as well through any event stored in MHIDAS.

1. Introduction

Safety management aims to ensure the execution of work activities in an orderly and safe manner, including the necessary organizational structures, accountabilities, policies and procedures (Liu et al. 2020). It covers many areas and concepts, through an interdisciplinary perspective expected to support the execution of operations and the analysis of incidents and accidents to ensure safety of operators and infrastructures (Dekker 2019).
Over recent years, Business Intelligence (BI) has been increasingly adopted in safety management, leading in some cases, to the notion of Safety Intelligence (SI). This latter follows an organizational safety management perspective to transform deconstructed data into information for the business (Rouach and Santi 2001; Sharda, Delen, and Turban 2018). SI is expected to generate usable and actionable safety recommendations gained through safety data and information processing. It can thus influence organizational safety management (Wang and Wu 2019), in line with "safety decision-making" (Huang 2018; Wang et al. 2017). Any BI solution for safety management should ensure that safety information has been processed in such a way that it can be helpful to decision-makers (e.g. setting goals, or defining policies) (Wang 2021). SI supports as well proactive risk management for holistic performance, as shown in the domain of aviation safety (Patriarca et al. 2019), with the potential of being increasingly helpful for senior managers (Fruhen et al. 2014). Organizations today collect data at a finer granularity, which implies a much larger data volume. Such activities require organizations to be agile and to make frequent and quick strategic, tactical, and operational decisions, of different complexity and

sensitivity. Making such decisions may require considerable amounts of relevant data, information, and knowledge. Processing them, in the framework of the needed decisions, must be done quickly, frequently in real-time, and usually requires some computerized support. Businesses are leveraging their data asset aggressively by deploying and experimenting with more sophisticated data analysis techniques to drive business decisions and deliver new functionality such as personalized offers and services to customers (Chaudhuri, Dayal, and Narasayya 2011) BI is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies (Sharda, Delen, and Turban 2018). By analyzing historical and current data, situations, and performances, decision-makers get valuable insights that enable them to make more informed and better decisions.

On these premises, this paper aims to investigate the potential benefits arising from the usage of BI tools for the sake of obtaining useful information for the analysis of industrial accidents. This general aim has been explored employing real safety data, like the ones available from the MHIDAS (Major Hazard Incident Data Service) database.

The remainder of the paper is structured as follows. Section 2 introduces the MHIDAS database structure, detailing the phases to build a BI model, and the subsequent Business Analytics (BA). Section 3 shows some results and examples of a dashboard that could be generated from the analysis. Lastly, section 4 suggests some conclusions over the obtained results.

2. Materials and Methods

A Business Model (BM) summarizes the configuration and logic of a business (Baden-Fuller and Mangematin 2015). Three essential BM dimensions have been identified in the literature: value creation, value proposition and value capture (Clauss 2017). The first dimension concerns the resources and capabilities employed in infra- and inter-organizational processes that generate value for the customer (Achtenhagen, Melin, and Naldi 2013). The value proposition dimension defines the range, nature and features of the offered products and services and the conditions at which these are provided (Ciampi et al. 2021). The value capture dimension explains how the business value proposition is converted into profits in a sustainable way (Teece 2010).

2.1 Exploring MHIDAS database

The MHIDAS database was created following the 1970's investigation by the UK Health and Safety Executive (HSE) on operational hazards. The study was carried out by the Safety and Reliability Directorate (SRD) of the UK Atomic Energy Authority (UK AEA), and became the most comprehensive non-nuclear application of risk assessment techniques at the time, revealing several areas where reliable data were not available.

Following the occurrence of several major incidents (e.g. Seveso, 1976, Mexico City and Bhopal, 1984), the HSE commissioned a survey to collect incident-related information, including data on toxic releases and hazardous materials that resulted in or had the potential to produce an off-site impact.. (Harding 1997).

The operating version of MHIDAS was then launched in 1986 by UK AEA, SRD and the UK HSE. It draws on public domain information sources (press cuttings, magazine articles, journals, published reports) to ensure that such information might be widely disseminated, and continuously updated (Harding 1997).

Regarding the information on MHIDAS, the database contains a detailed compilation of all the parameters necessary to know precisely what happened. Table 1 presents an excerpt from some database fields. The most remarkable feature of MHIDAS is that each record includes an entry for each substance that has been involved in the event itself. Therefore, it will be possible to study in-depth the variety of substances playing a role in the historic data (Llopart 2001).

Table 1. Description of main parameters used on MHIDAS database.

CODE	MEANING	DESCRIPTION
AB	Abstract	A summary of the incident, with detailed textual information.
AN	Record number	The registration number on the database.
DA	Date of incident	Date of the incident in the form DD/MM/YY.
DG	Economic Damage	Estimate (in dollars) of the material damage caused by the incident.
GC	General causes	The general cause of the incident.
LO	Location of incident	The geographical location of the incident in three granularity levels: city/region/country.
MH	Material hazard	Field used to associate the most likely risk for each material or situation.
MN	Material name	Substance name involved in the accident.
NP	People affected	Estimation of the number of fatalities, injured or evacuated people due to the incident.

| QY | Quantity of material | Estimation of the amount of material involved in the incident. |
| SC | Specific causes | The specific cause of the incident, such as "overheat", "overload", etc. |

The information that is collected in the database can be obtained from many sources and with the availability of background material for its review, with information that allows the calculation of coefficients or parameters for deterministic analysis and parameters to perform descriptive analysis.

2.2 Constructing the Business Intelligence Model

Decision support queries require operations such as filtering, join, and aggregation. To efficiently support these operations, special data structures have been developed (Chaudhuri, Dayal, and Narasayya 2011). The research was focused on the creation of a BI model in which users can interact with the information displayed in the Data Warehouse (DW) (Sharda, Delen, and Turban 2018). Implementing a BI system requires careful planning to assure that it meets users' expectations, usually following these basic steps (Oracle 2004):

I. **Identify End-User requirements:** It is important to know how the end-users (for the purpose the end-users are research in safety and risk management) will analyze the data. It is possible to identify the questions that the BI system needs as: What information do you have now? What additional information do you need? How do you want the information presented? The answers to these questions refer to the MHIDAS, and industrial incident reports involved with hazardous materials. Each report has parameters that describe what occurred and the parameters for quantifying the extent of the accident.

II. **Identify the Data Source:** The data can be distributed among numerous locations, such as transactional databases and flat files. MHIDAS database is available in a .txt file, which has been historically distributed and developed from the National Institute for Occupational Safety and Health, the UK HSE, and the UK AEA in the form of an Occupational Safety and Health on CD-ROM (OSH-ROM).

III. **Design the data model**: The data model firstly defines dimensions, measures, and so forth. Afterward, it can map the metadata objects to the physical data sources. For relational tools, it defines items, calculations, joins, etc. using an existing relational data source. The research was designed with the Extraction-Transformation-Loading (ETL) process explained in the next section:

    a. <u>Create the Data Store</u>: It must deploy the data model as physical objects in the database and load the data from its sources. The data store is an analytic workspace. By extracting the data from the source and to import them into the software to create the workspace.

    b. <u>Generate the Summary Data</u>: BI data is essentially hierarchical so that data can be summarized at various levels. In analytic workspaces, summary data is stored in the same analytic workspace objects as the base-level data. By creating and managing the queries (the tables where the data model takes the data).

    c. <u>Prepare the data for client access and grant access to end-users</u>: The client tools query the metadata to find out what data is available, where to get it, and how to present it. By managing the relationship between the queries inside the architecture model, it guarantees the right information for the users. Users must have database access rights so that they can view and manipulate the data. For research purposes, the client's tools are not present in this study since access to the information is not monitored or controlled but is freely accessible through a public openly accessible link.

IV. **Create and Distribute Reports:** At this step, it is possible to develop reports and share them with the user community. The reports created for this research show the parameters to describe what occurred in the incidents.

2.3 Developing the Business Analytics model

BI is employed for monitoring the performance of business processes through accurate presentation and analysis of multidimensional data, taken from distributed transaction processing systems across the enterprise (Al-Aqrabi et al. 2015).

The BA suggests three independent steps, tightly interacting and partly overlapping. The first step is a **Descriptive Analysis** where the response to some fundamental questions (e.g. "What happened?" or "What is happening?") is provided through enablers such as Business reporting, Dashboards, etc. The second step is called **Predictive Analysis**. It uses as a source the descriptive analysis and allows it to know "What will happen?" or "Why will it happen?" via enablers such as Data Mining, Text Mining, etc. The last step is **Prescriptive Analysis** which allows answering questions such as "What should I do?" or "Why should I do...?" using Optimization techniques, Decision Modelling and Simulation (Sharda, Delen, and Turban 2018).

Through the BA process, it is possible to transform the data source into actions: the process starts by defining a well-established business process, and identifying the opportunities for projections on future events and outcomes, which could be used for selecting the more appropriate business decisions and actions (Sharda 2020). The data from MHIDAS database was managed as a snowflake BI model, one of the most famous data

warehousing architectures. The attention is focused on building a scalable and maintainable infrastructure (often developed in an iterative way, subject area by subject area) that includes a centralized data warehouse and several dependent Data Marts (each for an organizational unit), which could be a subset of a data warehouse, typically consisting of a single subject area. The developed solution allows dealing with more than 9000 industrial incident reports worldwide. Each report has maximum 21 parameters (either textual, categorical, or numerical) used to describe the respective event in a structured and systematic way. Then, applying the BA, a set of dashboards was created (e.g. a search engine by type or class; a descriptive analysis of the accidents by specific and/or general causes).

3. Results and discussion

This data model highlights the large number of interactions needed to relate each information. The core of the model has a table (Facts table) where the parameters have a relationship One-to-One with the accident identification (ID). In other words, the parameters in the Facts Table have the same numerosity (data, location, economic damage, people affected, record number, abstract, people density, contributor, material, major event and quantity of material). In the branches, the data model has several relationships Many-to-One, where the parameters have higher numerosity compared to accident ID (hazardous, origin, causes, incident type, ignition source, specific causes, keywords, general causes, material code). With the description, it understands how the BI model was created to be a snowflake architecture (Facts Table in the core and the parameters with a higher number of hierarchies being in the branches).

Figure 1 proposes an excerpt of a dashboard, which represents a search engine by type of accidents. The key features of the dashboard can be listed as follows:

There are two sliders where the first one reports the years, allowing single or range filter. The second one shows buttons that report the type of severity to classify the accident (i.e., A, B, C, D, E, F, G, H, I, No Class).

There are four text filters where each one represents a searcher (i.e., country, state, city and material). In each filter, it is possible to write the name of the desired parameter to force the respective dashboard update.

A stacked column chart reporting the number of incidents collected in the database per year.

A geographical map showing the number of incidents per country. The size of the bubble describes the number of fatalities by accident. It is intended to show a worldwide distribution of criticality.

Two additional cards that allow respectively to (i) show the total number of incidents, (ii) the amount of economic damage (in dollars) involved in the incident.

A dynamic matrix that reports the parameters to present synthetically main parameters on the incidents. In the dynamic matrix of the dashboard, it is possible to visualize the accidents classified by the number of fatalities, it observes the famous accident of Bhopal represented twice, this is an example of the cross-join multiplicity of the parameters since the accident was classified by two types of hazard (fire and toxic).
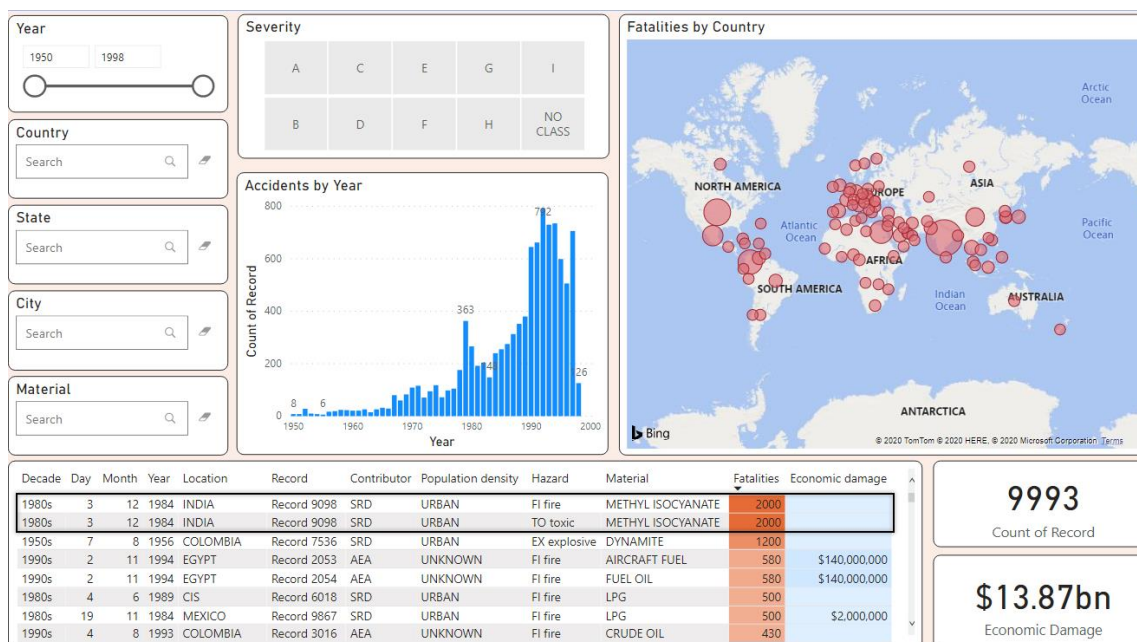

Decade	Day	Month	Year	Location	Record	Contributor	Population density	Hazard	Material	Fatalities	Economic damage
1980s	3	12	1984	INDIA	Record 9098	SRD	URBAN	FI fire	METHYL ISOCYANATE	2000	
1980s	3	12	1984	INDIA	Record 9098	SRD	URBAN	TO toxic	METHYL ISOCYANATE	2000	
1950s	7	8	1956	COLOMBIA	Record 7536	SRD	URBAN	EX explosive	DYNAMITE	1200	
1990s	2	11	1994	EGYPT	Record 2053	AEA	UNKNOWN	FI fire	AIRCRAFT FUEL	580	$140,000,000
1990s	2	11	1994	EGYPT	Record 2054	AEA	UNKNOWN	FI fire	FUEL OIL	580	$140,000,000
1980s	4	6	1989	CIS	Record 6018	SRD	URBAN	FI fire	LPG	500	
1980s	19	11	1984	MEXICO	Record 9867	SRD	URBAN	FI fire	LPG	500	$2,000,000
1990s	4	8	1993	COLOMBIA	Record 3016	AEA	UNKNOWN	FI fire	CRUDE OIL	430	

Figure 1. Search engine by type of incident, and aggregated data.

Any dashboard has been conceived dynamically, allowing also drill-through functionalities. A user can navigate through a page in the report that focuses on the details of a specific entity drilling-through designated fields. This functionality allows exploring in the same environment data from multiple pages, but automatically restricted by the active filters; and to use the cross-report drill through to connect two or more reports. For example, the user can take the report identification number and navigate through a detailed description of the respective event on a dedicated page (Figure 2).
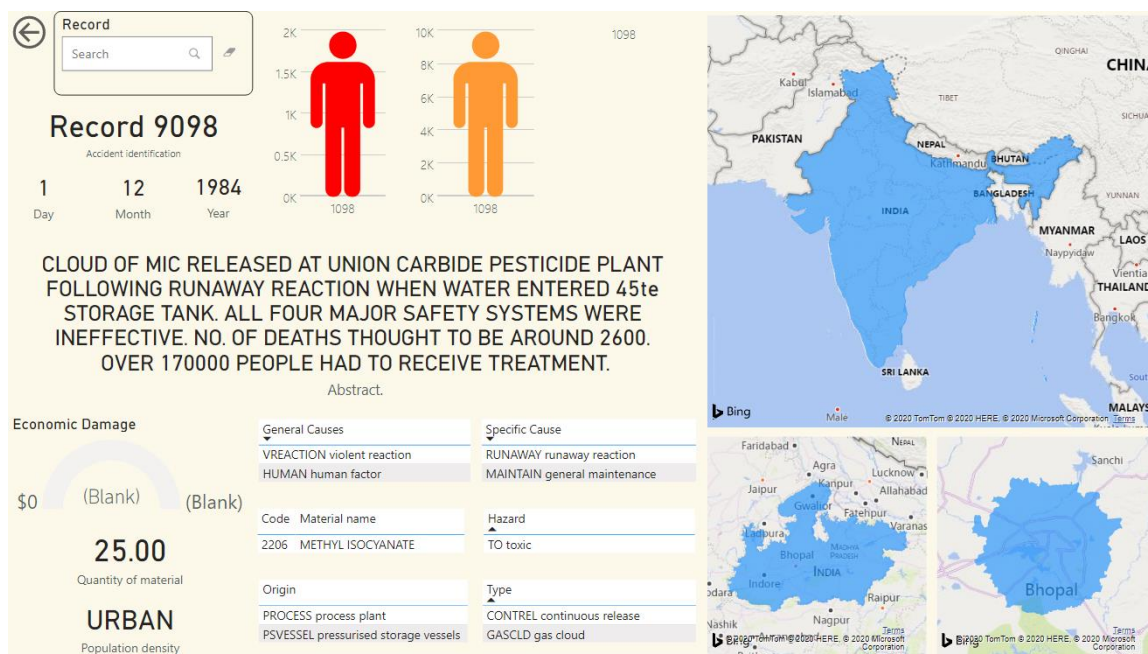

Figure 2. Example of a drill-through function related to Bhopal event, listing the information available in MHIDAS.

4. Conclusion

The contribution of this research illustrates the results of the application of the BI tools in an incident reporting system, related to the chemical industry. It shows that with the help of BI tools, a set of dashboards can be obtained to allow a visual-descriptive analysis of extensive data information reported in a database. The developed analyses can provide useful information for different users, and they are expected to support decision-making. For instance, the damage cost of past accidents may justify the cost of additional safety measures in a system (Paltrinieri et al. 2012). Through a snowflake BI architecture model, fast and efficient answers can be obtained to create specific data analytics. This structured analysis also allows a progressive enhancement of meta-knowledge for improving the quality of the investigations and data gathering. These results are only the firsts step into more complex IT applications for safety management, but they indicate the way forward for a wider risk learning process. In this regard, they also constitute the basis for other optimizations, e.g. through Machine Learning (ML) algorithms, as for promising research in this area (Paltrinieri et al. 2020). For example, it could be possible, in future research to perform dedicated text mining on the narrative available in MHIDAS, eliciting knowledge that could have not been reported in the other structured fields. With these studies, new parameters can be included in the reports, contributing to predictive and prescriptive analyses. The present study provides an early example of these techniques in the domain of risk and safety management, showing the real potential for their adoption at a larger scale, in any industrial system.

References

Achtenhagen, Leona, Leif Melin, and Lucia Naldi. 2013. "Dynamics of Business Models – Strategizing, Critical Capabilities and Activities for Sustained Value Creation." *Long Range Planning* 46 (6): 427–42.

Al-Aqrabi, Hussain, Lu Liu, Richard Hill, and Nick Antonopoulos. 2015. "Cloud BI: Future of Business Intelligence in the Cloud." *Journal of Computer and System Sciences* 81 (1): 85–96.

Baden-Fuller, Charles, and Vincent Mangematin, eds. 2015. "Introduction: Business Models and Modelling Business Models." In *Advances in Strategic Management*, 33:xi–xxii. Emerald Group Publishing Limited.

Chaudhuri, Surajit, Umeshwar Dayal, and Vivek Narasayya. 2011. "An Overview of Business Intelligence Technology." *Communications of the ACM* 54 (8): 88–98.

Ciampi, Francesco, Stefano Demi, Alessandro Magrini, Giacomo Marzi, and Armando Papa. 2021. "Exploring the Impact of Big Data Analytics Capabilities on Business Model Innovation: The Mediating Role of Entrepreneurial Orientation." *Journal of Business Research* 123 (February): 1–13.

Clauss, Thomas. 2017. "Measuring Business Model Innovation: Conceptualization, Scale Development, and Proof of Performance: Measuring Business Model Innovation." *R&D Management* 47 (3): 385–403.

Dekker, Sidney. 2019. *Foundations of Safety Science: A Century of Understanding Accidents and Disasters, Chapter 2: The 1920s and Onward: Accident Prone*. Boca Raton: Taylor and Francis, CRC Press.

Fruhen, L.S., K.J. Mearns, R. Flin, and B. Kirwan. 2014. "Safety Intelligence: An Exploration of Senior Managers' Characteristics." *Applied Ergonomics* 45 (4): 967–75.

Harding, A B. 1997. "MHIDAS: The First Ten Years." *141*, 12.

Huang, Lang. 2018. "Big-Data-Driven Safety Decision-Making_ A Conceptual Framework and Its Influencing Factors." *Safety Science*, 11.

Liu, Zimei, Kefan Xie, Ling Li, and Yong Chen. 2020. "A Paradigm of Safety Management in Industry 4.0." *Systems Research and Behavioral Science* 37 (4): 632–45.

Llopart, Sergio Carol. 2001. "BASES DE DATOS sobre accidentes industriales." *MAPFRE* N. 155: 47–56.

Oracle. 2004. *Oracle Business Intelligence. Concept Guide*. Vol. Second. Oracle.

Paltrinieri, Nicola, Sarah Bonvicini, Gigliola Spadoni, and Valerio Cozzani. 2012. "Cost-Benefit Analysis of Passive Fire Protections in Road LPG Transportation: Cost-Benefit Analysis of Passive Fire Protections." *Risk Analysis* 32 (2): 200–219.

Paltrinieri, Nicola, Riccardo Patriarca, Michael Pacevicius, and Pierluigi Salvo Rossi. 2020. "Lessons from Past Hazardous Events: Data Analytics for Severity Prediction," 8.

Patriarca, R., G. Di Gravio, R. Cioponea, and A. Licu. 2019. "Safety Intelligence: Incremental Proactive Risk Management for Holistic Aviation Safety Performance." *Safety Science* 118 (October): 551–67.

Rouach, Daniel, and Patrice Santi. 2001. "Competitive Intelligence Adds Value:" *European Management Journal* 19 (5): 552–59.

Sharda, Ramesh. 2020. *Analytics, Data Science, & Artificial Intelligence*. Eleventh edition. Hoboken, NJ: Pearson.

Sharda, Ramesh, Dursun Delen, and Efraim Turban. 2018. *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. Fourth edition. New York, NY: Pearson.

Teece, David J. 2010. "Business Models, Business Strategy and Innovation." *Long Range Planning* 43 (2–3): 172–94.

Wang, Bing. 2021. "Safety Intelligence as an Essential Perspective for Safety Management in the Era of Safety 4.0: From a Theoretical to a Practical Framework." *Process Safety and Environmental Protection* 148 (April): 189–99.

Wang, Bing, and Chao Wu. 2019. "Demystifying Safety-Related Intelligence in Safety Management: Some Key Questions Answered from a Theoretical Perspective." *Safety Science* 120 (December): 932–40.

Wang, Bing, Chao Wu, Bo Shi, and Lang Huang. 2017. "Evidence-Based Safety (EBS) Management: A New Approach to Teaching the Practice of Safety Management (SM)." *Journal of Safety Research* 63 (December): 21–28.