Heliyon 10 (2024) e40330

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Tensor-based analysis of eclectic-reported crowd accidents: An innovative retrospective methodology design study

Dóra Edelmann [a,*], Arnold Őszi [b], Tibor Goda [a,c]

[a] Doctoral School on Safety and Security Sciences, Óbuda University, 1034, Budapest, Hungary
[b] Institute for Security Sciences and Cyber Defense, Banki Donat Faculty of Mechanical and Safety Engineering, Óbuda University, 1034, Budapest, Hungary
[c] Institute for Natural Sciences and Basic Subjects, Banki Donat Faculty of Mechanical and Safety Engineering, Óbuda University, 1034, Budapest, Hungary

ARTICLE INFO

Keywords:
Crowd accident
Tensor-based analysis
Media credibility
Press report
Crowd crush

ABSTRACT

Introduction: Analysis of crowd accidents contributes to accident prevention. Therefore, we employ a tensor-based approach. The innovative tensor-based approach facilitates the streamlining of longitudinal studies, promotes error detection, and enhances the transparency and traceability of data collection. This study focuses on crowd accidents, the direct cause of which is the movement of the crowd (Excluding other external factors: e.g. fire, structural damage.). It aims at investigating the reliability of the records documented in relation to crowd accidents and the type of repetitions that can be found in the events.

Materials and methods: The study employed a web-based retrospective methodology with innovative tensor-based analysis, examining 186 fatal crowd accidents from 1979 to 2023. Data was collected from public sources, including news reports, government reports, and scientific publications. The analysis considered the following variables: event type, place, date, number of victims, cause, environmental characteristics, date and reliability of documented information source origination. Tensor-based method combines the improvement of the quality of the coverage and investigates changes in content over time. The seven-step method, which stores information about crowd accidents in matrices, is presented here in detail. The v_{cr} factor is introduced to evaluate the credibility of sources.

Results: The results show that those news items about crowd accidents are the most reliable which were created 2 years after the events. Crowd accidents are analyzed based on their influencing defining characteristics. We claim that we were able to isolate new risk factors related to the locations of crowd accidents. Globally, we focus on accidents that occurred during donation distributions and when entering buildings.

Conclusion: It can be concluded that the new, seven-step, tensor-based data collection method improves the credibility value of individual information by more than 25 %. The impact of accident factors plays a key role in establishing risk factors and in the prevention of accidents. The tensor-based approach can be directly applied to record databases, enhance data provenance, and capture the temporal evolution of information.

* Corresponding author.
E-mail addresses: edelmann.dora@phd.uni-obuda.hu (D. Edelmann), oszi.arnold@bgk.uni-obuda.hu (A. Őszi), goda.tibor@bgk.uni-obuda.hu (T. Goda).

https://doi.org/10.1016/j.heliyon.2024.e40330
Received 28 April 2024; Received in revised form 3 July 2024; Accepted 10 November 2024
Available online 19 November 2024
2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The number of reports on crowd accidents is steadily increasing. The phenomenon can be explained both by the increase in the number of accidents and the amount of information about accidents. The increase in the number of accidents is caused by urbanization, increased mobility [1], building complexes, mass events [2,3], the increase in the number of organizational and planning errors, the lack of an adequate exchange of information can all provide explanations. However, the increase in the amount of information about accidents can be explained by the increase in the number of information tools, news items [4,5] and the number of information sources. The reliability of information about accidents at mass gatherings varies greatly. Nevertheless, to prevent accidents, a thorough study and comparison of the cases that have already happened is essential, which requires the close cooperation of international organizations, researchers, experts and professionals. The limited access to adequate quality information, the limitations arising from the technical background, and the non-uniform professional concepts make it difficult, however, the increasing availability of public visual contents and the eyewitness accounts on the Internet help research in the field. The availability of high-quality information is essential for research and accident prevention.

Mass events can often strain the planning and response resources of the community or country hosting the event [6], which is most often associated with the number of participants. In its risk classification of mass events, the World Health Organization (WHO) classifies the dangers arising from crowd motions, such as injuries, stampedes, trampling, in an independent category, in addition to epidemiological and environmental factors, in the same category as non-contagious diseases and excessive drug use [7]. Due to their characteristics, the accidents that can be caused by crowd motion differ in many respects from other accidents experienced at mass events. A simultaneous view on research groups related to evacuation, stampede and crowd crush (such as: a) event or case analysis of stampedes, b) computer-based simulations, c) biology and medical-level researches, d) models of pedestrian dynamics [8]) is necessary.

Accidents caused by crowd motion can be traced back to several causes, which are often difficult to separate from each other, and the detailed recognition and analysis of the events is essential in preventing further tragedies. The methods used to investigate the causes and determine individual risk factors investigate the triggering factors in a multidimensional (complex) or a one-dimensional (linear) manner [2,9–16]. Individual studies most often investigate the degree of overcrowding [7], the number of people present [17], the degree of threat [18] and the nature of individual events, seeking for a correlation between those and the accidents. However, the availability of adequate quality and quantity of data required for analysis is often hindered by the different technological background of each area, by economic, political, religious reasons, or even professional deficiencies. When referring to actual cases, the credibility, evaluability, reliability and independence of the information are crucial, yet they are not given enough emphasis in the relevant literature. Conversely, this study places a strong emphasis on the reliability and time-varying nature of information, so global problems can also be recognized.

The amount of information about crowd accidents suitable for the comparison of events is most often collected from several different sources. A typical collection method is the initial use of publications containing already available comparative data, institutional databases, websites collecting events, search engines, and special software. The available information is often textual, and the non-uniform terminology in the field [19] in many cases requires the joint analysis, comparison and expert interpretation of several sources. The lack of appropriate terminology and the need for a common language is well illustrated by the fact that even basic concepts (for example, “crowd” or “pedestrian”) have only been circumscribed and remained recommendations for the time being [20]. The WHO emphasizes the joint participation of large numbers of people and the wide-ranging health effects in connection with mass events [21,22]. In its statement, the National Association of Emergency Medicine Services (EMS) Physicians (NAEMSP) recognized that, when defining mass events, there can also be found examples in the literature of minimum numbers of 1000 and 25,000 people [3]. In contrast to traditional text-based information storage, this study presents a novel approach to storing information in the form of numbers (and colours), enabling efficient data compression and retrieval.

Another characteristic of information sources about crowd accidents is that they often only summarize events that happened in one single period or in one area. The detailed process of data collection is often difficult to trace due to the variety of available sources, which not only complicates the process of collecting work, but also reduces its scientific value. Also, in many cases, non-English language information is excluded from the information that can be used as data sources. Recently, a study has been made on the manner of reporting on crowd accidents and the changes in terms used over time [23]. The change in the content of the information found in individual sources related to accidents caused by crowd motion compared to the time of an accident, however, has not been investigated so far. Due to the repetitive nature of crowd accidents caused by crowd motion, the analysis of their characteristics and the documentation of the records are of fundamental importance [4].

This study focuses on accidents at mass events that can be attributed to crowd motion (stampede, crowd collapse and crush). The analysis based on the tensor-cell model provides an opportunity to evaluate individual factors affecting the accidents that occurred at different times compared to the times of the accidents. It also provides a basis for predicting individual time series and determining the characteristics of individual accidents. Here, the use of matrices created with the seven-step data collection method promotes cooperation and correction between individual information collections (e.g. correction of the number of non-fatal accidents), while at the same time retaining the possibility of thorough analyses, which is essential in accident prevention.

Here, during the analysis, 186 crowd accidents that occurred between 1979 and 2023 are investigated, all of which were caused exclusively by crowd motion (stampede, crowd collapse and crowd crush). Due to the different reliability value of the individual reports (online and printed press, image and video material, scientific publications, expert reports), they were treated as mixed sources, called eclectic sources. Based on these, the recognized risk factors of crowd accidents (e.g. number of personnel, event location) will be investigated, with a level of detail that has not been used until now. This way, risk factors such as the crowd trying to get inside, closed gates, or club venues are also recognized. These are analyzed as independent, one-dimensional, percentage effects. During the analysis, the written and visual (photo, video) sources are handled together, which also gives the opportunity to estimate the number of personnel on site. In this way, the nature of the forms of movement preceding crowd accidents is also identified in many cases. Regarding specific events, similar studies are usually conducted during the Hajj [24–26] and in India [27–29] emphasize accidents caused by crowd motion, due to the large number of deaths and their recurrence. Through the upward trend recognized during the analysis, this study also draws attention to fatal accidents due to donation distributions and access to facilities.

In the light of the above, the aim of our investigation is to facilitate the collection of information about accidents related to crowd motion, and to produce the most authentic information possible, taking into account the available sources and the time of generation of the information. Our goal is to draw attention to new risk factors with this, as well as the correlations of the data available to us, with which we help reduce the number of accidents caused by crowd motions.

This paper is organized as follows. Section 2 presents the data collection and analytical approach. Section 2.3 describes data collection. The type of data collected and the content structure of the matrices containing them are according to Section 2.4. Some information about mass gatherings that have been included in the database is detailed here. The data structure method of the matrices belonging to each event is according to Section 2.5. The 7 steps of data expansion, correction and addition will be described in detail here. (Section 2.5-2.7.6.). The v_{cr} (values related to the reliability of the data source) value presenting the authenticity of the individual information is introduced in Section 2.7.2. In Section 2.8, the tensors built from the individual matrices and the new tensor-cell model procedure, which provides the possibility to enter missing data, are presented. In Section 3, the number of people and the number of fatal accidents (Section 3.1), the type of events (Section 3.2), the age of the deceased (Section 3.3), the physique of the victims (Section 3.4), the environmental characteristics (Section 3.5), the forms of movement (Section 3.6) are analyzed. Section 4 provides a more general discussion based on the results where limitations of this work are also listed. Section 5 concludes this paper summarizing the main findings and providing advice to address future research.

Table 1
The seven-step process, design, theory and usage.

step		design	theory	usage			
nr	ref		step type		detail	outcome matrix	section
1	a	input	data collection/ search	Generation of a database complying with minimum-criteria standards	date type/language: date, geographic data/english keywords: "crowd crush" OR "stampede" OR "crowd collapse"	A	2.3 2.4 2.5
	b	3× filter	1. data filtering	First stage of professional filtering of elements, excluding events that can be attributed to mass movement		B	2.5.1
2	a		2. data filtering	Excluding sources with 100 % identical content (i.e., duplicates) and informational value		B	2.5.2
	b		3. data filtering	Removal of data with a single source of information		C	2.5.2
3		input	new datas	Expansion of the dataset by incorporating additional events identified through content analysis of information		D	2.6
4	a	correction	formal correction	Data cleaning: Removing redundant empty columns to improve database efficiency and organization		E	
	b	input	new datas/ search	Database augmentation with native language sources and detailed event search	date type/language: date, geographic data, official source/english and native keywords: "event date" AND "location" AND "fatal accident" OR "police" OR "forensic" OR "medical" OR "pathological"	E	2.7
	c	3× check	check v_{cr} value	Increasing the v_{cr} value		E	2.7.2
	d		professional check	Database augmentation: Expert-driven analyses, expert identification of participant numbers and movement patterns		E	2.7.4
	e		professional bias check	Expert-driven database error checking based on injury and casualty values from peer-reviewed literature		E	
5		build	build groups	Content grouping based on quantity, into 3 groups		E	2.7.5
6	a	build	build tensors	Tensor construction from grouped matrices		F	2.8
	b	use	use f,g functions	Data processing using traditional mathematical, statistical methods		F	2.9
7		analyze	analyze	Analyzis			2.10

Therefore, we hope to contribute to a better understanding of accidents arising from crowd pressure. This includes the traceability of individual data collection processes, the necessity of source evaluation, and the investigation of temporal changes in individual analyzed data using tensor-based methods.

Glossary and Abbreviations

This study uses the following crowd motion concepts:

- crowd accident, here: an event in a crowd of at least 100 people, involving injury, caused by crowd motion, crowd compression;
- accident caused by a quasi-static crowd: an accident caused by the slower filling of the crowd, high personnel density. As its antecedent, the dynamic movement of a determinable part of the crowd cannot be determined;
- dynamic crowd motion: faster, joint progressive movement of a separable part of the crowd;
- N: number of participants;
- N_D: number of fatalities;
- n: sample number, case number, piece number.

2. Materials and methods

The study employed a web-based retrospective methodology with innovative tensor-based analysis, examining 186 fatal crowd accidents from 1979 to 2023. Review and/or approval by an ethics committee was not needed for this study. Data was collected from public sources, including news reports, government reports, and scientific publications. Quantitative data collection with systematic sampling, expert video analysis, data processing with a uniquely developed tensor-based spreadsheet, data determination with Excel (MSOffice) database extension was used.

Data was manually extracted from the identified sources using a structured data collection protocol. Multiple researchers were involved in the data extraction process to minimize observer bias. The initial data was gathered from the comparative studies listed in Table 10 and from publicly accessible Internet search platforms. The web-based search was conducted between 2022.11 and 2024.04 using the following English keywords on Google, ResearchGate, Web of Science, Elsevier, and PubMed: "crowd crush," "stampede," and "crowd collapse." The minimum criteria for inclusion were: (1) At least one fatality (2) Occurrence between 1979.08.13 and 2023.05.20 (3) Inclusion of the event date (4) Inclusion of the event location. Detailed search was conducted on publicly accessible internet search platforms (Google, ResearchGate, Web of Science, Elsevier, PubMed) between 2022.11 and 2024.04 in English and in the first official language recorded for the event location. The following keywords were used: "event date," "location," AND "fatal accident" OR "police" OR "forensic" OR "medical" OR "pathological". The information described in Chapter 4, points 1–13, was recorded. Events with only one information source and those with other (fire, structural damage) factors as the direct cause of the incident were excluded.

The study tool was developed parallel with the data-collection, summarized at Table 1 in design, theory and usage, as a seven-step process.

It is important to acknowledge that publicly available data may be incomplete or underreported. This could introduce potential bias in the sample. The study is acknowledged to be potentially susceptible to sampling bias due to the manual data collection methods (observer bias, data entry errors, selection bias) and the use of publicly available information. The completeness of publicly available information could introduce non-response bias, as some accidents might not be reported or detailed information might be missing. To minimize errors, only information about events that were corroborated by at least two independent sources was retained.

The analysis considered the following variables: event type, place, date, number of victims, cause, environmental characteristics, date and reliability of documented information source origination. Tensor-based method combines the improvement of the quality of the coverage and investigates changes in content over time. The seven-step method, which stores information about crowd accidents in matrices, is presented here in detail. The v_{cr} factor was introduced to evaluate the credibility of sources.

To analyze accidents that can be traced back to crowd motion and occurred at mass events, and to determine the risk value of individual factors, it is necessary to explore the causes as influencing factors. The location, the date, the estimated number of

Table 2
Crowd Crush Disasters, data analyzed in comparative reviews.

nr.	1	2	3	4	5	6	7	8	9	10	11
ref.	[32]	[28]	[33]	[27]	[34]	[35]	[26]	[7]	[1]	[4]	this study
assessed risk											
location	•	•	•	•	•		•	•	•		•
date	•	•	•	•	•		•	•	•		•
number of participants	•	•		•	•		•				•
number of fatalities	•	•	•	•	•		•	•	•		•
number of injured	•	•	•	•	•			•		•	•
type of accident	•	•	•		•			•			•
event type	•	•	•	religious	•		religious	stadium	•	•	•
part of the day	•	•									
form of movement	•	•			•						•
other	•				•						•

participants, the type of events and the time of day are the most common among the factors that are included in the comparative reviews (Table 2). The duration of the event, the location of the event (outdoor or indoor), the position of the crowd within the venue (seated or mobile), the enclosure of the event (closed or open), the type of event, the atmosphere of the crowd, the availability of alcohol and drugs, crowd density, the geography of the event (terrain/venue), previous history of similar events, and the average age of the crowd are named as factors [30]. Nevertheless, some of these risk sources receive little emphasis in the relevant literature. The database collection method presented in this study provides an opportunity to record sources of risk at work that have been treated with insufficient weight in the literature and can be traced back to crowd motion. During the analysis, among the accidents at mass events that can only be attributed to crowd motion, the correlations regarding age and the direct location of the accidents are also presented. Among the collective crowd motions, the crowd that wants to enter is specifically identified as a new source of risk (Section 2.10.). To thoroughly analyze the data and recognize new sources of risk, the distinction between event types is different in the literature. Here, for the sake of comparability and detail, a large number of 11 event types are distinguished (sport, religious, political, entertainment, donation distribution, concert, club, transport, fireworks, university event, trade and other).

In this study, great emphasis is placed on the available visual material (video and photographs) when analyzing accidents caused by crowd movement and crowd compression. In this way, not only the forms of movement, but also the number of personnel on site can be better estimated and controlled. At the same time, the available video material sets limits on the time of the events that can be investigated, so the accidents between 1979 and 2023 are analyzed. The internal risks of mass gatherings are focused on (intrinsic risk, such as overcrowding, crushing, pushing, stampeding), the external risks (extrinsic risks, such as the location, the venue, environmental conditions, or the actions of third parties, terrorism and violence, and cascading risks, where a minor incident causes a disproportionate reaction) [31] are screened. Accordingly, accidents resulting from structural damage, fires, or natural disasters will not be investigated. When analyzing the root causes and exploring individual forms of movement, it is extremely important to know the motivation of emergency movements and the timing of the events as thoroughly as possible. Among the forms of movement recognized in the crowd, evacuation and self-excited crowd motion processes can be distinguished. The information required for more thorough research of the events is also reviewed in the native language of the location during the comprehensive analysis of the 186 events (Sections 3.1-3.5).

2.1. The credibility value of information

In this study the v_{cr}, (values related to the reliability of the data source) introduced. This provides an opportunity to evaluate the perceived credibility and volume of individual sources of information about accidents at mass events, by grouping them. Since the change of information over time is also investigated here, it is essential to evaluate the content and quality of individual sources instead of filtering out duplicates. For example, multi-source information fusion plays a significant role in decision-making, sample classification, fault diagnosis, quantum analysis, medical services, information volume processing and in the fight against deception [36]. At the same time, with the spread of Internet social media platforms, more and more checklist methods for evaluating information have spread, e.g. CARS (Credibility, Accuracy, Reasonableness, Support), RADAR (Relevance, Authority, Date, Appearance, Reason for Writing), CRAAP (Currency, Relevance, Authority, Accuracy, Purpose), CCOW (Credentials, Claims, Objectives and Worldview [37]. With the v_{cr} factor introduced in this study, the au value will not be weighted, but their change over time will be evaluated on a linear scale.

Table 3
Crowd Crush Disasters, comparative reviews. (Studies according to Table 2).

					sources used					
nr	ref	period	events	area	press	scientific publications	media	other	language	intrinsic risk
1	[32]	1980–2007	215	Global	•				English	n.a.
2	[28]	2001–2010	40	Global				D1	English	E1
3	[33]	1971–2011	21	Global	•	•				
4	[27]	1954–2012	34	Global	•	•				n.a.
5	[34]	1711–2017	116	Global		•			n.a.	–
6	[35]	1902–2019	243	Global	•					–
7	[26]	1994–2006	6	Saud Arabi	•	•		D2	English, Arabic	n.a.
8	[7]	1901–2021	40	Global				D3		E3
9	[1]	1990–2022	45	Global	•	•			English	+
10	[4]	1900–2019	281	Global	•		•			+
11	this study	1979–2023	186	Global	•	•	•	•	English, Origin	+

D1: LexisNexis Academic and manual search (Ngai); Indian journals (Roy).
D2: Medline and OVID.
D3: Emergency Events Database (EM-DATS).
E1: excluded: events related to fire and terrorism.
E2: excluded: extreme weather conditions.
E3: stadium accidents (collapse, fire, other).

2.2. The sources

In addition to the sources used in similar studies (Table 3) (press, eyewitness accounts, videos, scientific publications, police, judicial, and medical expert reports), the previously little-used native language sources, judicial reports, and pathological expert opinions are also included. The available sources of information about accidents at mass events are the press [4,23], eyewitness accounts [38], videos [39–43], scientific publications, police, forensic and medical experts, pathology reports [44–49]. Social media often increase the number of eyewitness accounts, in many cases providing valuable visual material or additional information about the emotional state of eyewitnesses [50].

2.3. Data collection

Existing studies use at least two steps to collect information about crowd accidents. The first data collection step usually means starting from a larger set (e.g. existing studies, searching with keywords in specific databases). In further steps, targeted searches for individual events, filtering of duplicates, or the use of additional databases are performed. Unlike the previous one, the method included here does not focus on filtering out data collection duplicates: To monitor the reliability of individual sources about crowd accidents and the change over time in the risk content affecting accidents, it investigates information from several different sources.

During the data collection, 186 items, events between 13 08 1979–20 05 2023 were investigated, of which those after 29 07 2014 were the subject of investigation. The systematization of each reported data was made according to the definitions published in Alhadhira's study, by supplementing them [34]. Different reliability values were determined for each source (press, eyewitness accounts, videos, scientific publications, police, judicial, medical expert reports). The time of appearance and generation of information from accidents were also recorded. In addition to the English-language data sources, the official language sources of the given event locations were also used. The information extracted from the data sources belonging to the given event (date, geographical data, events that caused the accident, number of participants, number of dead, number of injured, type of movement that caused the accident, type of event) together with the time of their appearance were recorded.

2.4. Structure of data matrices

The information extracted from sources describing events resulting in fatal accidents caused by crowd motion were entered as numbers into a matrix (Fig. 1) corresponding to the number of events. Columns of matrices:

1: date {year};
2: date {month};
3: date {day};

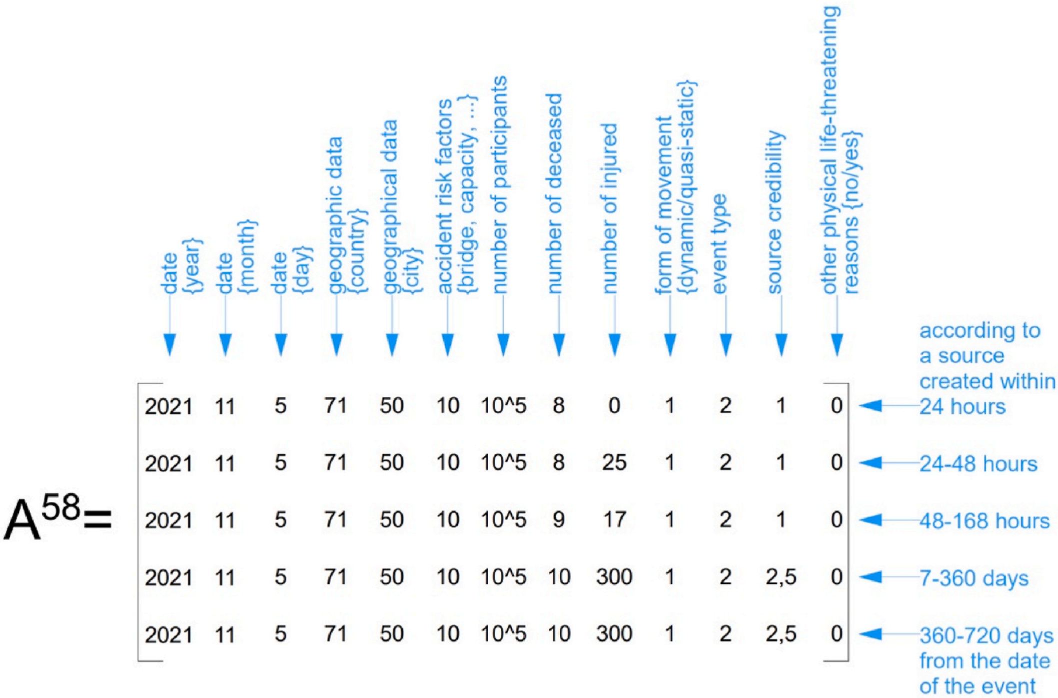

Fig. 1. Sample matrix of the event: November 5, 2021, Astroworld Festival crowd crush.

4: geographic data {country}, where {country} is the number assigned to each country, value: (1, 2, 3, …79) (see Table 4);

5: geographic data {city}, where {city} is the number assigned to the settlements of each location, its value is: (1, 2, 3, …147) (see Table 4);

6: accident-causing event {bridge, number of people, crowd trying to get in, … }, where {bridge, number of people, crowd trying to get in, … } is the number assigned to each accident-causing event, its value: (0, 1, 2, …15). Factors that can be identified from the descriptions as direct or indirect causes of individual accidents and that are included in the description of at least two independent events are referred to as accident-hazardous influences. (see Table 5);

7: number of participants {the number of participants in each event};

8: number of deaths {person};

9: number of injured {person};

10: the type of motion form that caused the accident {static/mostly static/dynamic}, where {quasi-static/mostly quasi-static, cannot be determined/dynamic} is the number assigned to each motion form, value: (1, 2, 3)(quasi-static, dynamic form of crowd motion: see Glossary);

11: the type of event {sport, concert, religious, political, … }, where {sport, concert, religious, political, … }, the number assigned

Table 4
Values of the variables assigned to the textual data of the matrices containing the data of fatal accidents, column 4–5.

Location					
Country (values of column 4)		City (values of column 5)			
Afghanistan (1)	Liberia (41)	Abidjan (1)	Cotonou (37)	Lilongwe (74)	Pasuruan (110)
Algeria (2)	Líbia (42)	Abuja (2)	Dhakka (38)	Lisbon (75)	Pekalongan (111)
Angola (3)	Madagascar (43)	Accra (3)	Dar es Salaam (39)	Lomé (76)	Beijing (Beijing (Peking)) (112)
Austria (4)	Hungary (44)	Abidjan (4)	Diffa (40)	Los Olivos (77)	Phnom Pen (113)
Bangladesh (5)	Malaysia (45)	Akashi (5)	Duisburg (41)	Luanda (78)	Piraeus (114)
Benin (6)	Malawi (46)	Alexandria (6)	Durban (42)	Lubumbashi (79)	Port Harcourt (115)
Brazil (7)	Mali (47)	Algiers (7)	Freetown (43)	Lucknow (80)	Pretoria (116)
Bulgaria (8)	Marocco (48)	Allahabad (8)	Fuxin (44)	Lusaka (81)	Quetzaltenango (117)
Burkina Faso (9)	Mexico (49)	Anambra church (9)	Gopalganj (45)	Madhya Pradesh (82)	Rabat (118)
Chile (10)	Germany (50)	Andhra Pradesh (10)	Harare (46)	Madrid (83)	Rajasthan (119)
Columbia (11)	Nepal (51)	Antananarivo (11)	Haridwar (47)	Malang (84)	Rochester, New York (120)
Dania (12)	Niger (52)	Aracaju (12)	Himachal Pradesh (48)	Malang (85)	Roskilde (121)
South Africa (13)	Nigeria (53)	Asansol, West Bengal (13)	Hong Kong (49)	Maligawatta (86)	San Salvador (122)
South Korea (14)	Italy (54)	Bagdad (14)	Houston (50)	Manila (87)	Sanaa (123)
Egypt (15)	Russia (55)	Balu Khali (15)	Innsbruck (51)	Medvode (88)	Sanghai (124)
Côte d'Ivoire (16)	Pakistan (56)	Bamako (16)	Jakarta (52)	Mecca (89)	Sangju (125)
Northern Ireland (17)	Peru (57)	Bandung (17)	Jalalabad (53)	Mexico City (90)	Santiago (126)
Ethiopia (18)	Portugal (58)	Banjarmasin (18)	Jeddah (54)	Mina (91)	Sao Paulo (127)
Belarus (19)	Salvador (59)	Basra (19)	Jharkhand (55)	Minsk (92)	Sheffield (128)
Philippines (20)	Sierra Leone (60)	Bepanda (20)	Jodhpur, Rajasthan (56)	Monrovia (93)	Sidi Boulaalam (129)
Ghana (21)	Spain (61)	Bihar (21)	Johannesburg (57)	Moshi (94)	Simaria (130)
Greece (22)	Sri Lanka (62)	Birmingham (22)	Kairo (58)	Moscow (95)	Srinagar (131)
Guatemala (23)	Sweden (63)	Brazzaville (23)	Kampala (59)	Mount Meron (96)	Soffia (132)
Guinea (24)	Saudi Arabia (64)	Brixton, London (24)	Kandukur (60)	Multan (97)	Seoul (133)
Honduras (25)	Slovenia (65)	Budapest (25)	Karachi (61)	Mumbai (98)	Tegucigalpa (134)
Hong Kong (26)	Sudan (66)	Bydgoszcz (26)	Karbala (62)	Nagpur (99)	Tehran (135)
India (27)	Tanzania (67)	Cali (27)	Katmandu (63)	Nairobi (100)	Timbuktu (136)
Indonesia (28)	Togo (68)	Caracas (28)	Kerala (64)	New York (101)	Torino (137)
Iraq (29)	Uganda (69)	Chennai (29)	Kerman (65)	Ningxia (102)	Tripoli (138)
Iran (30)	UK (70)	Chicago (30)	Khaled Al-Mahdi (66)	Northampton (103)	Uíge (139)
Israel (31)	USA (71)	Chittagong (31)	Kikwit (67)	Obio/Akpor (104)	New Delhi (140)
Japan (32)	Venezuela (72)	Cincinnati (32)	Kinshasa (68)	Odisha (105)	Ullhasanager (141)
Yemen (33)	Zambia (73)	Colombo (33)	Kuala Lumpur (69)	Omdurman (106)	Varanasi (142)
Cambodia (34)	Zimbabwe (74)	Conakry (34)	Kunda (70)	Oromia (107)	Wai (143)
Kameron (35)	Slovenia (75)	Cookstown (35)	Kwekwe (71)	Ouagadougou (108)	West Bengal, Kalimpong (144)
Kenya (36)	Sudan (76)	Corinaldo (36)	Lagos (72)	Pakpattan (109)	Yaounde (145)
China (37)	Tanzania (77)		Lampung (73)		Zunjubar (146)
Congo (38)	Togo (78)				
Democratic Republic of the Congo (39)	Uganda (79)				
Poland (40)	Haiti (80)				
	Senegal (81)				

Table 5
Values of the variables assigned to the textual data of the matrices containing the data of fatal accidents, columns 6, 11.

Incidence	
Accident risk factors (column 6)	Event type (column 11)
unknown (0)	unknown (0)
bridge (1)	club (1)
capacity (2)	concert (2)
crowd wanting to get in (3)	donation (3)
crowd wanting to get out (4)	fireworks (4)
dark (5)	mall (5)
locked gates (6)	other (6)
mass panic (7)	political (7)
muddy, slippery ground (8)	religious (8)
other obstacles (e.g. luggage) (9)	sports (9)
police measure (10)	transport (10)
returning crowd (11)	university (11)
slippery stairs, ramp (12)	
stairs, slope (13)	
tear gas, pepper spray (14)	
tunnel (15)	

to each type of event, its value: (0,1, …11) (see Table 5);

12: data on the authenticity of the source, its value: (1, 1.5 … 2.5, 3) (see Table 6**).** As follows:

$a_{i,12}$, $b_{i,12}$, $c_{i,12}$, $d_{i,12}$, $e_{i,12}$, $f_{i,12}$, $g_{i,12} \leq 3$, and

$a_{i,12}$, $b_{i,12}$, $c_{i,12}$, $d_{i,12}$, $e_{i,12}$, $f_{i,12}$, $g_{i,12} = \max(v_{cr})$, where: $a_{i,j}$ is $\boldsymbol{A}$, $b_{i,j}$ is $\boldsymbol{B}$, …$g_{i,j}$ is the element of the i-th row and j-th column of the $\boldsymbol{G}$ matrix.

13: the existence of an external other physical life-threatening cause that can be inferred from the description of the event {no/yes}, value: (0, 1).

The rows of the matrices:

1: value, generated within 24 h of the date of the event.

2: value, generated within 24–48 h from the date of the event.

3: value, generated within 48–168 h from the date of the event.

4: value, generated within 7–360 days from the date of the event.

5: value, according to the source generated within 360–720 days from the date of the event.

2.5. Method, construction of database matrices

2.5.1. Structure and content of matrices A and B, first event data filtering

The initial $\boldsymbol{A}$ matrices (Figs. 1 and 2A) contained the data of the events that, according to the sources, ended in fatal accidents that occurred in the dense crowd, as a result of crowd motion or excessive personnel density in the period between 29 07 2014–20 05 2023. The sources of the events had at least the following information values: date (year, month, day), geographic data (country). The first four columns of mxn matrices related to a single event contain at least one value for which,

$$a_{i,j} \neq 0, if\ j = 1, 2, \ldots 4$$

where:

i: is the row index of the element;

j: is the column index of the element.

Table 6
Values related to the reliability of the data source (v_{cr}), column 12.

Data source	
Type	Value (v_{cr})
press	1
eyewitness reports	2
analyzable videos	2
scientific publication	3
police report	3
judicial report	3
medical report	3
correction factor:	+0,5
large amount (number of pieces>5) photos and videos	

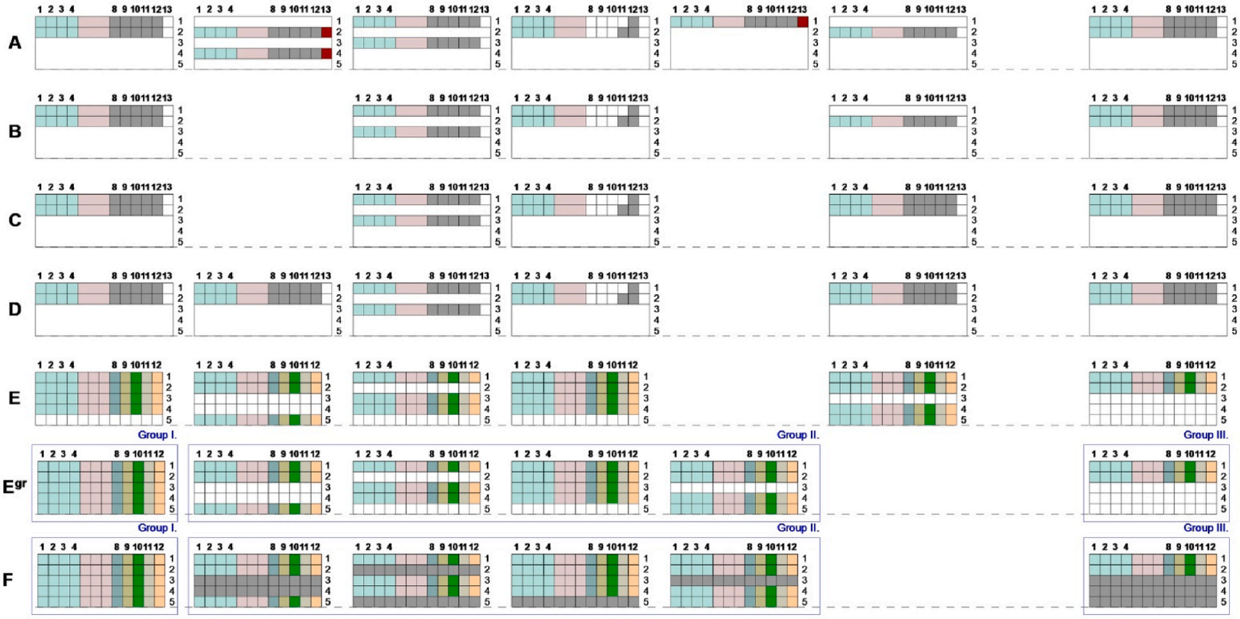

Fig. 2. The key process of the methodological flow.

After subtracting the duplicates, the number of these events from the sources is 79. Thus, the shape of a total of 79 different $\boldsymbol{A}$ matrices is:

$$\boldsymbol{A}_{(k=1-79)} \in \mathbb{R}^{5\times 13} \tag{1}$$

where:

$\mathbb{R}$: is the set of rational numbers;

m: is the number of matrix rows;

n: is the number of matrix columns;

k: is the serial number of the matrix.

Each matrix $\boldsymbol{A}$ is without exception a sparse matrix.

Among the event information recorded in the $\boldsymbol{A}$ matrices, there are also those whose available documents show that the accidents were not solely caused by crowd motion without an external physical cause. (In column 13, the value: 1 is included.) In addition to the incidents involving accidents that can be attributed purely to crowd motion, the number of cases describing other life-threatening conditions (e.g. fire, structural damage, smoke) was 5. These events were deleted from the initial data group, so the number of $\boldsymbol{B}$ matrices (Fig. 2B) obtained as a result is 74, and its shape is:

$$\boldsymbol{B}_{(k=1-74)} \in \mathbb{R}^{5\times 13} \tag{2}$$

2.5.2. Structure of *C* matrices, second event-data filtering

In the second step, the events with only one source and the events with multiple sources but with the same information value were deleted, the number of which was 3. Thus, the number of $\boldsymbol{C}$ matrices (Fig. 2C) that are the subject of further investigations is 71, and their shape is:

$$\boldsymbol{C}_{(k=1-71)} \in \mathbb{R}^{5\times 13} \tag{3}$$

2.6. The structure of D matrices, first data extension

In the third step, new events were added to the data list based on references to similar cases in the description of the events. There were 4 of them, so the number of $\boldsymbol{D}$ matrices (Fig. 2D) obtained as a result is 75, and their structure is:

$$\boldsymbol{D}_{(k=1-75)} \in \mathbb{R}^{5\times 13} \tag{4}$$

2.7. Structure of the *E* matrix, second data extension

In the fourth step, the 13th column of each matrix was deleted, due to the filtering detailed in 2.5. The thus came about

$$\boldsymbol{E}_{(k=1-75)} \in \mathbb{R}^{5\times 12} \tag{5}$$

shaped data matrices containing information on individual accidents (Fig. 2E) were expanded with additional data content (Fig. 3) as follows:

2.7.1. *Filling in missing lines, reducing rarity*

The sparseness of the matrices belonging to each event was reduced after a targeted search for information from descriptions from the time interval corresponding to the row. In addition to English-language press sources, the search also included the exploration of information prepared in the official language of the given location.

2.7.2. *Increasing information credibility values (column 12)*

Credibility values were increased by exploring additional public source documents found on the Internet for each event. The targeted search was conducted using the following keywords, in English and in the official language of the given location:

{Date (year/month/day)} & {city/name of accident site facility} & fatal accident/deceased & police/judicial/forensic/medical/pathological.

The found police information, descriptions of legal proceedings, medical opinions and scientific publications were marked as the most authentic ($v_{cr} = 3$) sources. Details of additional values are summarized in Table 6. Sources with $v_{cr} \geq 2.5$ are detailed in Table 7 Based on the above, the information value (v_{cr}) for all events increased by 25.3 %:

$$\sum_{k=1}^{75} \left(v_{cr(A)}\right)_k = 1.000 \tag{6}$$

$$\sum_{k=1}^{79} \left(v_{cr(E)}\right)_k = 1.253 \tag{7}$$

where:

$\left(v_{cr(A)}\right)_k$: the largest v_{cr} data value related to the credibility of the source belonging to each matrix $\boldsymbol{A}$; $\left(v_{cr(E)}\right)_k$: the largest v_{cr} data value for each matrix $\boldsymbol{E}$ related to the credibility of the source.

2.7.3. *Increasing information credibility values (column 12) with a v_{cr} correction factor*

The exploration of the video and photo documentation material belonging to each event also provides an opportunity to further increase the credibility values. The relevant information, such as the accident risk factors (column 6), the number of participants (column 7) and the form of movement (column 10) calculated after analyzing photographs with the appropriate perspective, were entered and specified accordingly.

2.7.4. *Correction of the number of injured people (column 9)*

The number of non-fatal injuries at each event is crucial both for the subsequent understanding of what exactly happened and for the preliminary preparation for each event. The purpose of this method is to compare events involving fatal accidents that can only be traced back to crowd motion and crowd density. Accordingly, the number of non-fatal injuries for each event is corrected by the number of accidents that occur in other cases based on the type of event and the number of participants.

The number of patient care provided at events, broken down by type of event: 4.85 PPTT at baseball matches, 6.75 PPTT at football matches, 30 PPTT at rock concerts (PPTT, patients per 10,000) [56]. The need for medical care at music festivals can be estimated from <0.01 % to 2 % of the number of participants, however, specialist resuscitative care is quite rare, required in 0.01%–0.04 % of cases [57]. Regarding international cases, the documented value of medical care (patient presentation) in relation to the number of participants is 0.01%–0.19 % [58]. The individual correction factors applied are summarized in Table 8. The authenticity of the

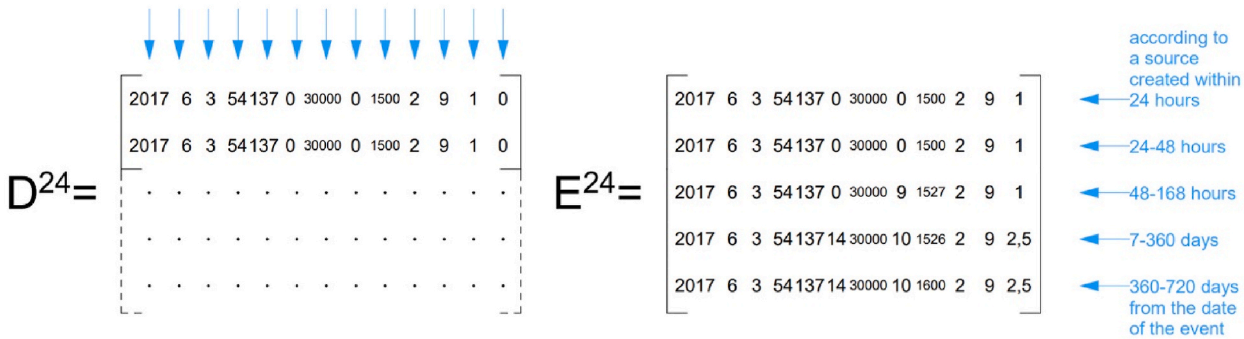

Fig. 3. Sample, Sparse D-matrix into E-matrix, Event: Juny 3, 2017, Turin, Italy, crowd accident.

Table 7
Sources with $v_{cr} \geq 2{,}5$.

date/city	source
2014.12.31., Shanghai	[51]
2015.10.14., Bydgoszcz	[52]
2017.06.03., Torino	judicial reports
2020.08.22., Los Olivos	security camera recording
2021.04.30., Mount Meron	[53]
2021.11.05., Houston	[54]
2022.10.29., Itaewon	[55]

Table 8
Correction of the number of non-fatal accidents (column 11).

event type (column 11)	correction factor
unknown (0)	−0.1 %
club (1)	−0.1 %
concert (2)	−0.3 %
donation (3)	−0.1 %
fireworks (4)	−0.1 %
mall (5)	−0.1 %
other (6)	−0.1 %
political (7)	−0.1 %
religious (8)	−0.1 %
sports (9)	−0.05 %
transport (10)	−0.1 %
university (11)	−0.1 %

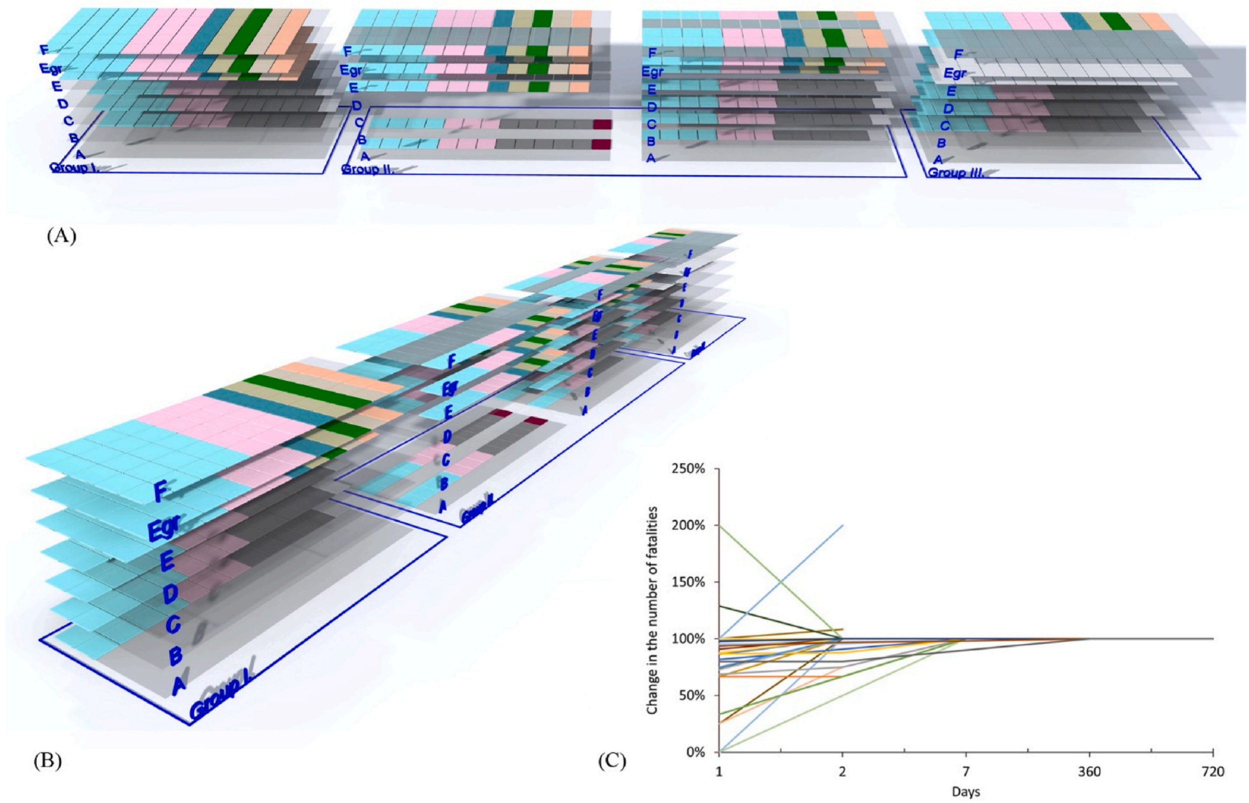

Fig. 4. (A) Tensor constructions with the "F" process of the methodological flow. (B) Tensor construction, incomplete tensors. (C) Change in the number of fatalities compared to the values of the 720th day.

information can be further expanded as needed. Analysis of video or photo documentation of individual events can provide accurate data on the number of participants. The available archival on-site weather information can reveal any missing elements of the history of the accident (e.g. muddy ground, rain). An example of the difference between the matrix $\boldsymbol{D}$ described in 2.6 and the matrix $\boldsymbol{E}$ obtained after the steps detailed in 2.7 belonging to the same event is shown in Fig. 3.

2.7.5. Grouping of matrices filled with information values

In the fifth step, the incomplete 2 matrices, the matrices with 3–4 rows and the complete matrices with 5 rows are placed in separate groups (Fig. 2E^{gr}). The groups are given the following names:

$$\text{Group I} : \boldsymbol{E}^{gr,I} \in \mathbb{R}^{n\times 12}, n = 5;$$

$$\text{Group II} : \boldsymbol{E}^{gr,II} \in \mathbb{R}^{n\times 12}, 3 \le n \le 4;$$

$$\text{Group III} : \boldsymbol{E}^{gr,III} \in \mathbb{R}^{n\times 12}, n = 2. \quad (8)$$

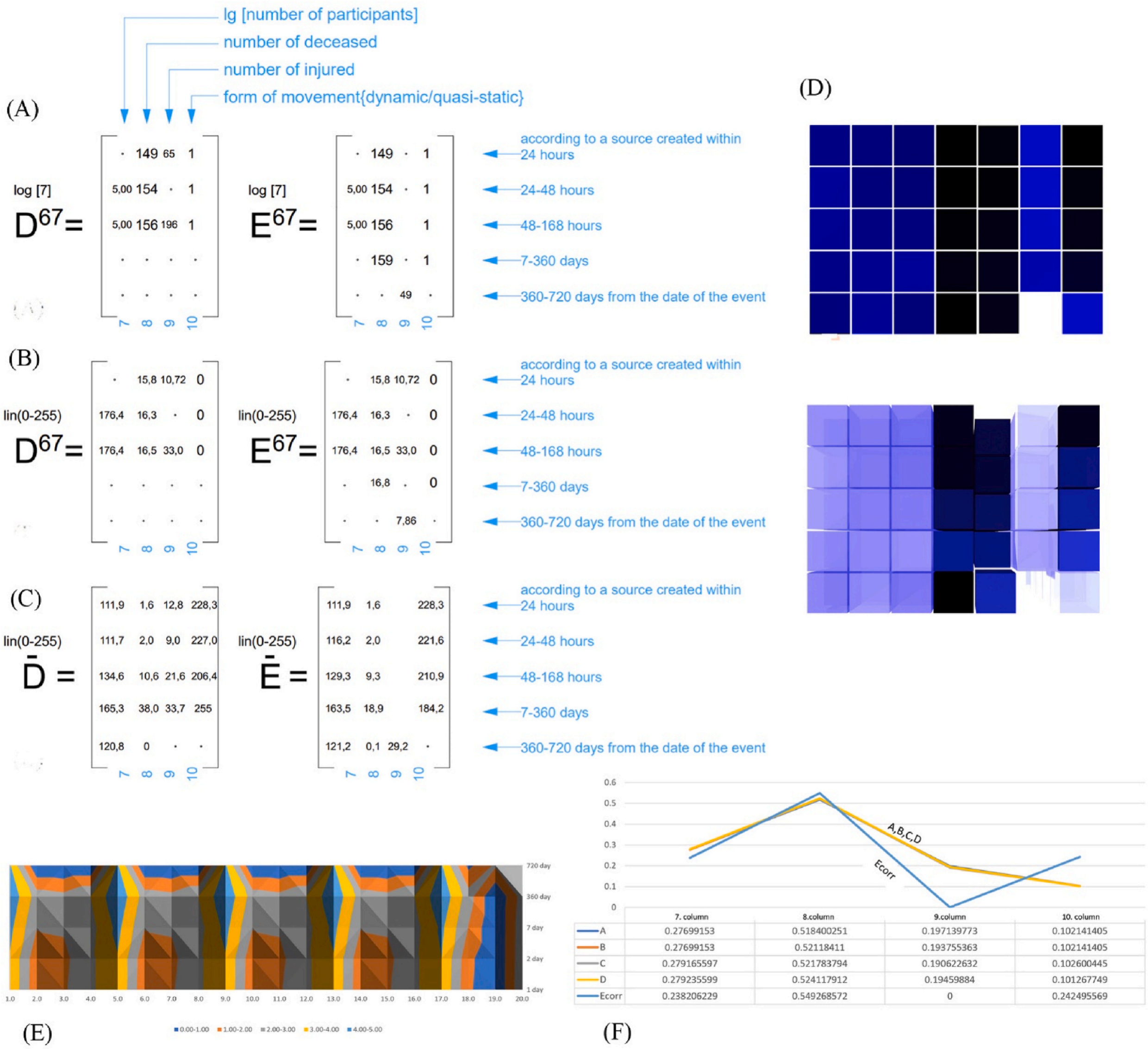

Fig. 5. (A)–(C) Sample, D-matrix and E-matrix, Event: Juny 3, 2017, Turin, Italy, crowd accident. (D) Containing data stored according to RGB colours as tensors, where R = 0, G = 0 and 0 ≤ B ≤ 255. (E) Contour distribution surface diagram of the matrices. (F) The change in v_{cr} value between the group of matrices. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.7.6. Structure of *F* matrices

In the sixth step, the missing data of the matrices grouped according to saturation can be used as a basis for calculations with traditional mathematical methods that are suitable for the accuracy of the additional results required, even different for each group.

For the construction of *F* matrices, Section 2.8. Presents a new method based on the tensor construction principle, specially designed for this database, which can also serve as a basis for solving several similar problems, especially in cases where the combined effect of two different time series prevails.

2.8. Tensor construction

Third-order tensors are built from the matrices corresponding to each event, in increasing order in time corresponding to the database collection time of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$, $\boldsymbol{E^{gr}}$, $\boldsymbol{F}$ (Figs. 2, 4A and 4B). Tensors also provide an opportunity to determine the similarities between individual events, the correlations between the information recorded in the data collection, and the missing values of the *F* matrices containing the most reliable information. Distinct patterns are evident in certain data points (e.g., number of fatalities) when considered in relation to the publication dates of the sources (Fig. 4C).

2.9. Tensor cell model procedure

The tensor-cell model procedure is explained on tensors of size 5 × 4 × 5 obtained from the data.

In the first step, 4 adjacent columns are selected, at least two of which are related to each other in terms of information content. These are, for example, the columns marked with numbers 7, 8, 9, and 10 (according to Section 2.5).

In the second step, due to the correction of the magnitudes, the logarithm of the values in columns 7 is determined (Fig. 5A):

$$a_{i,j,k}(\log[7]) = \log\left[a_{i,j,k}\right], if\ j = 7\ . \tag{9}$$

In the third step, each tensor data value is linearly interpolated with a target value between 0 and 255, (Fig. 5B) and then a tensor separated according to the data collection phases is created from the average of the values included in separate tensors belonging to each event, keeping the location of the data (Fig. 5C).

For the values of the tensor data produced in the previous way, it is always true that:

$$0 \le a_{i,j,k} \le 255 \tag{10}$$

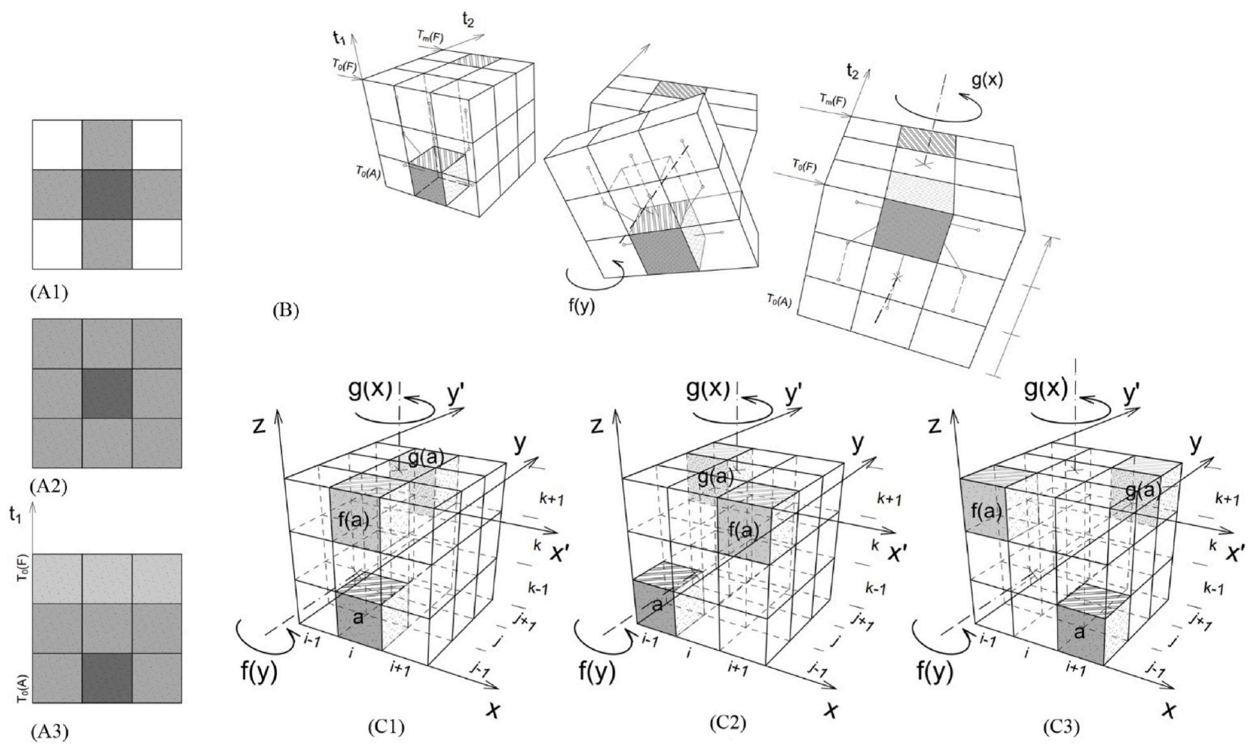

Fig. 6. The tensor model method presented in this study. (A) Neighbourhood relations of cell models (A1) Neumann (A2) Moore (A3) this study. (B) Tensor-cell model method neighbourhood relations and basic principle. (C) Tensor cell model method with function notation.

Accordingly, the data values can be directly used as an example for storing data in colours corresponding to the RGB colour scale (Fig. 5D).

In the fourth step, a new logarithm of the values in columns 8 and 9 of the matrices is determined to determine the relationship between the individual data. The values included in the tensors A, B, C, D and E are shown in a Fig. 5E.

It can be concluded that the matrices (***A,B,C,D***) from the earliest phases of data collection are similar in terms of their data values. A greater discrepancy can be seen between the group of matrices ***A, B, C, D*** and the individual $\boldsymbol{E_{corr}}$ data tables (Fig. 5F).

According to the above, in the fifth step, the matrices ***B, C, D*** are removed from the tensor constructions, and then tensors with data size 3 × 3 × 3 containing three adjacent columns and rows are selected. According to this method, the incomplete matrices ***F*** are constructed using a unique cell model (Fig. 6A).

According to this, the first collected data stored in the ***A*** matrix and their neighbouring relations form a relationship with the data stored in the own matrix, the ***E*** matrix and the ***F*** matrix. In order to maintain the adjacency conditions, the F level of the tensor is created by rotating each specific, nine-element zone of the 3 × 3 × 3 cube, as well as with the function placed on the data in them (Fig. 6B), according to the following geometric correspondence:

Let t_1 axis containing $T_0(\mathbf{A})$; $T_0(\mathbf{E_{gr}})$, $T_0(\mathbf{F})$: $t_1 \equiv z$; the axis marking the columns is x, $T_0(\mathbf{F})$; $t_2 \equiv y'$ containing $T_m(\mathbf{F})$, $T_0(\mathbf{A})$; and y containing $T_m(\mathbf{A})$, Fig. 6C according to his nominations. Denote a cell by $a_{x,y,z}$, where:

$$(i-1) \leq x \leq (i+1);$$

$$(j-1) \leq y \leq (j+1);$$

$$(k-1) \leq z \leq (k+1). \tag{11}$$

The function f(y) is the prediction function interpreted in the z,x plane, and the g(x) function is the prediction function interpreted in the y',x plane. Based on the method:

If $a_{i,(j-1),(k-1)}$ is not an empty cell (Fig. 6C1), then:

$$a_{i,(j-1),(k+1)} = f\left(a_{i,(j-1),(k-1)}\right). \tag{12}$$

If they are full cells, $f(a_{(i-1),(j-1),(k-1)})$ and $f(a_{(i+1),(j-1),(k-1)})$ symmetry are to be examined relations (f_2), as well as

$$f\left(a_{x,(j-1),(k-1)}\right) \ll \left(a_{x,(j),(k+1)}\right), or$$

$$f\left(a_{x,(j-1),(k-1)}\right) \gg \left(a_{x,(j),(k+1)}\right)$$

in case function smoothing is performed (f_3), thus:

$$a_{i,(j-1),(k+1)} = f_3 f_2 f\left(a_{i,(j-1),(k-1)}\right). \tag{13}$$

Then according to the function g(x) in the zy' plane:

$$a_{i,(j+1),(k+1)} = g\left(f_3 f_2 f\left(a_{i,(j-1),(k-1)}\right)\right). \tag{14}$$

Accordingly, if $a_{(i-1),(j-1),(k-1)}$ is not an empty cell (Fig. 6C2), then:

$$a_{(i+1),(j-1),(k+1)} = f\left(a_{(i-1),(j-1),(k+1)}\right); \tag{15}$$

$$a_{(i+1),(j-1),(k+1)} = f_3 f_2 f\left(a_{(i-1),(j-1),(k+1)}\right); \tag{16}$$

$$a_{(i-1),(j+1),(k+1)} = g\left(f_3 f_2 f\left(a_{(i-1),(j-1),(k+1)}\right)\right). \tag{17}$$

If $a_{(i+1),(j-1),(k-1)}$ is not an empty cell (Fig. 6C3), then:

$$a_{(i-1),(j-1),(k+1)} = f\left(a_{(i+1),(j-1),(k-1)}\right); \tag{18}$$

$$a_{(i-1),(j-1),(k+1)} = f_3 f_2 f\left(a_{(i+1),(j-1),(k-1)}\right); \tag{19}$$

$$a_{(i+1),(j+1),(k+1)} = g\left(f_3 f_2 f\left(a_{(i+1),(j-1),(k-1)}\right)\right). \tag{20}$$

2.10. Analysis

Several comprehensive studies have already been conducted to investigate the characteristics of crowd accidents attributable to crowd motion [2,59]. These summarize the experiences of cases that have already happened, and their importance in determining the accident risks and causal factors of individual mass events is outstanding. Both the exploration of the factors influencing the occurrence of accidents and the search for their complex relationship are ongoing. Lu et al. called the analyzes that investigate the impact of a single factor (e.g. accident location, accident type, environmental factors, crowd behavior) on the accident as one-dimensional analysis. An independent investigation of one factor does not give an adequate picture of its relationship with other factors, and thus of the complex process of the formation and course of crowd accidents. At the same time, the ratio-statistical analysis of these factors contributes to establishing their extent in the sample of accident victims [9].

3. Results

Overall, this analysis reveals a disproportionately high number of fatal stampede-related accidents in Saudi Arabia and India (Fig. 7). As per the analysis of accident events by year, the highest number of fatalities was consistently observed during large-scale religious gatherings. However, the relation between the number of participants and the number of fatalities remains unclear is not linear, as will be demonstrated later in this study (Fig. 8). However, a rigorous investigation is warranted.

Table 3 describes the comparative studies summarizing and analyzing the accidents that occurred at mass events and were caused by the movement of the crowd. Since the events can be analyzed more thoroughly, as well as in accordance with the recommendations of international organizations [30], in this section we describe in detail the results that supplement the research carried out so far. Accordingly, in connection with accidents at mass events that can only be traced back to crowd motion, in Section 3.1 the number of people, in Section 3.2 the types of events, in Section 3.3 the age of the deceased, in Section 3.4 the physique of the victims, in Section 3.5 the characteristics of the environment, in Section 3.6 certain characteristics resulting from crowd motion recognized effects are described.

3.1. *Correlation between the number of participants and the number of fatalities*

Hsieh et al. investigating 213 accidents caused by human stampedes between 1980 and 2007 [32], he was able to estimate the number of participants in 133 cases. According to the results, the median fatality rates of the events were 32.5/100,000, with a range of 0–44,000/100,000 (IQR 6–315.0/100,000). According to his results, 33 (23.3 %) accidents were not fatal, 47 (35.3 %) had a lower fatality rate, and 55 (41.4 %) had a fatality rate greater than or equal to 100/100,000. Ngai et al. When analyzing 40 accidents caused by crowd compression in India between 2001 and 2010 using the Ngai method [28], the Median Fatalities (IQR) was 5.5 (2–17), and when analyzed using the Roy method, it was 5 (2–39). The Median Injuries (IQR) value was 13.5 (5–31) using the Ngai method and 17 (10–45.5) using the Roy method (Table 9).

The correlation between the number of participants at individual events and the number of casualties (injured and fatalities) is often determined on a logarithmic scale. Feliciani et al. factually traces this back to the particularities of estimating the number of people in the crowd. According to his results, crowd accidents and death and injury rates show an obvious trend, namely that the larger the crowd, the less likely the accident occurred. He mentions this on the one hand as a “logical consequence” of the numbers (according to his description, the number of victims must always be smaller than the size of the crowd), on the other hand as the “safety-in-numbers” recognized in cycling and pedestrian traffic [60,61] relates it to a characteristic [4]. Feliciani et al. and from the comparison of the results investigated in this study, while the trend of the results is similar, the mortality rate is different. The results suggest that in the case of dense crowds, the “safety-in-numbers” relationship probably applies differently below and above 200,000 people (Fig. 9).

The average mortality rate for the entire case study was 567/100,000 in the study by Feliciani et al. [4] analyzing 195 cases, while

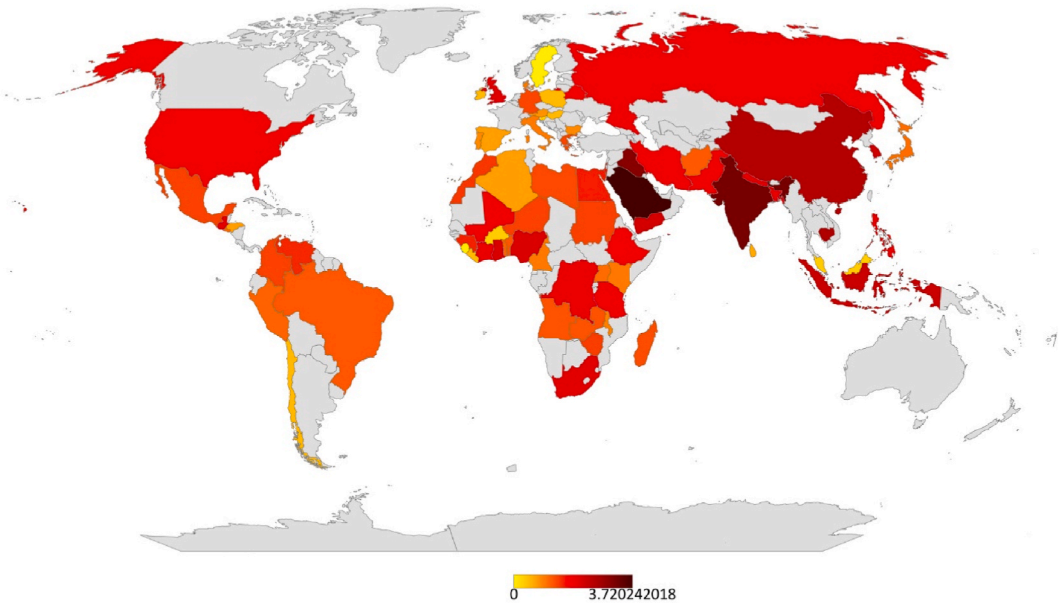

Fig. 7. Fatal crowd accident location, number of fatal victims due to crowd motion disasters, 1979–2023, numbers at logarithmic scale.

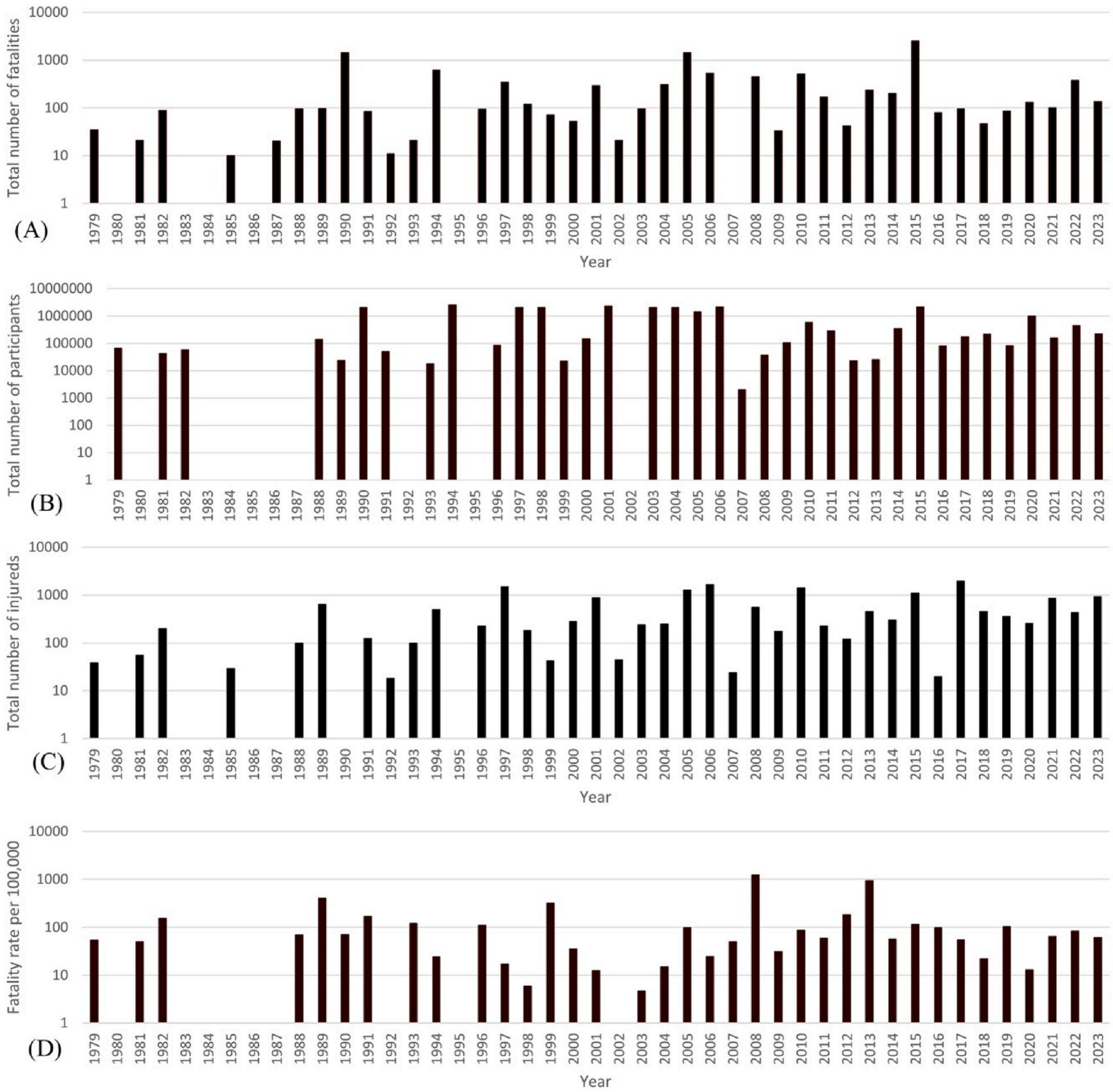

Fig. 8. Number of injuries caused by crowd motion, according to 1-year periods (1979–2023) (A) number of fatal accidents (B) number of participants in each event ending in fatal accident (C) number of non-fatal injuries (D) number of fatalities in proportion to the number of participants, average value.

Table 9
The number of fatal and non-fatal victims in accidents caused by crowd compression, crowd motion and crowd collapse depending on the number of participants (IQR: interquartile range).

source	area	event number	death toll	Median Fatalities (IQR)	Median Injuries (IQR)
Hsieh et al. [32]	Global	133	32.5/100,000 (0–44,000)/100,000)	32.5 6–315.0/100,000	
Ngai et al. Ngai Method [28]	India	34		5.5 (2–17)	13.5 (5–31)
Ngai et al. Roy Method [28]	India	27		5 (2–39)	17 (10–45.5)

Table 10
Fatal and Non-Fatal Casualties in Crowd Crush, Crowd Movement, and Stampede Incidents Based on Participant Count (IQR: interquartile range). MF: Median Fatalities, IQR-F: Interquartile Range-Fatalities, pers.: person, nr.: number of incidents analyzed. æ

number of participants	nr.	IQR-F (min-max) [pers.]	MF (IQR) [pers.]	IQR-F (min-max) [percent]	MF (IQR) [percent]	fatality rate/100,000
1-1000	19	10 (1–31)	7	1.89 (0.25–15.50)	1.88	3421 (250–15,500)
1000–10,000	24	9.25 (1–85)	4	0.77 (0.02–8.50)	0.21	785 (17–8500)
10,000–100,000	58	30 (1–224)	30	0.14 (0.00–2.24)	0.04	163 (2-2240)
100,000–1,000,000	15	43.50 (1–375)	11	0.04 (0.00–0.16)	0.01	32 (1–159)
1,000,000 ≤	11	585.50 (14–2411)	325	0.04 (0.00–0.12)	0.01	33 (1–121)
all (1–2,500,000)	127	31.50 (1–2411)	11	0.39 (0.00–15.50)	0.07	741 (1–15,500)

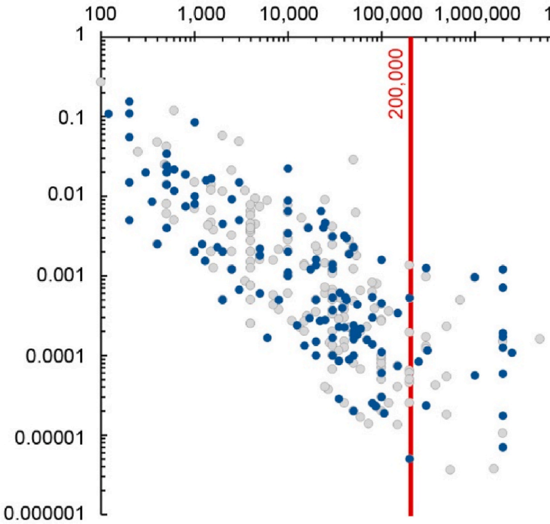

Fig. 9. The correlation between the number of fatalities and the size of the crowd, logarithmic scale, Feliciani et al. [4] and this study. Legend: gray circle: Results according Feliciani et al. [4], sample number: 194, blue circle: Results according this study, sample number: 127. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in this study, based on 127 cases, it was 741/100,000 (Table 10). The following provide logical explanations for the difference (31 %): (a) The data collection method used in this study follows the number of deaths for years after the event. Accordingly, the number of victims who were injured as a result of the accidents but died months after the time of their occurrence was also recorded. This data is often not found in press sources, but only in subsequent forensic and medical expert reports. (b) This study only focused on accidents caused by crowd movement, while the comparative analysis also included accidents caused by other external factors. This draws attention to the severity of crowd-related accidents and the need for separate investigation of each cause. (c) The difference in the total number of participants reported in various sources, even in the most reliable sources, is considered difficult to estimate.

Considering the above, it is necessary to study the correlation between the number of participants and the number of deaths in more detail for a more thorough analysis. Based on the data of this study, 185 events, taking place between 13 08 1979–20 05 2023 the

Table 11
Equations of the maximum curve enveloping the data series of fatal accidents caused by the movement of the crowd at mass events, according to 5 personnel groups, in log decomposition, with linear and polynomial maximum curves, N: number of participants, N_D: number of fatal accidents.

number of participants (N)	number of fatal accidents (N_D) (linear)	number of fatal accidents (N_D) (polynomial)
≥1,000,000	$N_D = + 0.0014*N-481$	$N_D =$ -2E-15*N^2 + 5E-09*N – 0.0025
100,000 ≤ N < 1,000,000	$N_D = + 0.0011*N+51$	$N_D =$ 9E-14*N^2 - 4E-08*N + 0.0044
10,000 ≤ N < 100,000	$N_D = - 0.0023*N+224.45$	$N_D =$ 6E-12*N^2 - 9E-07*N + 0.0306
1000 ≤ N < 10,000	$N_D = - 0.0109*N+88.359$	$N_D =$ 3E-09*N^2 - 4E-05*N + 0.1209
<1000	$N_D = - 0.0267*N+36.333$	$N_D =$ 6E-07*N^2 – 0.0008*N + 0.2944

Depending on the number of participants, the events were analyzed in 5 groups. Accordingly.
(I) Number of participants N < 1000.
(II) Number of participants 1000 < N < 10,000.
(III) Number of participants 10,000 ≤ N < 100,000.
(IV) Number of participants 100,000 ≤ N < 1,000,000.
(V) Number of participants 1,000,000 < N.

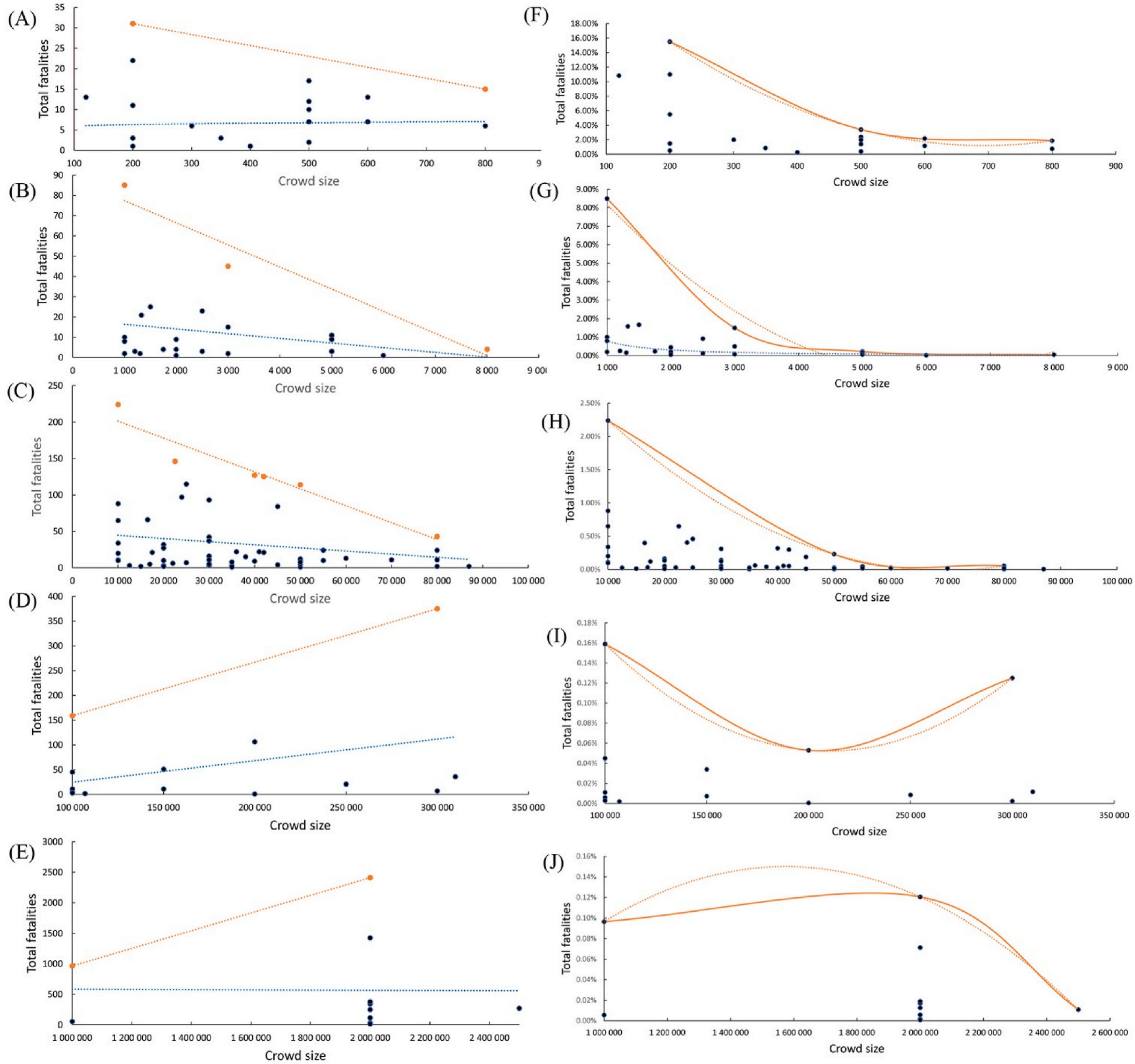

Fig. 10. The correlation between the number of fatalities and the size of the crowd, (A) N < 1000 number of participants (Group I). (B) The correlation between the number of fatalities and the size of the crowd, 1000 < N < 10,000 number of participants (Group II). (C) The correlation between the number of fatalities and the size of the crowd, 10,000 < N < 100,000 number of participants (Group III). (D) The correlation between the number of fatalities and the size of the crowd, 100,000 < N < 1,000,000 number of participants (Group IV). (E) The correlation between the number of fatalities and the size of the crowd, 1,000,000 ≤ N number of participants (Group V). (F) The correlation between the number of fatalities (percentage) and the size of the crowd, 1000 ≤ N number of participants (Group I). (G) The correlation between the number of fatalities (percentage) and the size of the crowd, 1000 < N < 10,000 number of participants (Group II). (H) The correlation between the number of fatalities (percentage) and the size of the crowd, 10,000 < N < 100,000 number of participants (Group III). (I) The correlation between the number of fatalities (percentage) and the size of the crowd, 100,000 < N < 1,000,000 number of participants (Group IV). (J) The correlation between the number of fatalities (percentage) and the size of the crowd, 1,000,000 < N number of participants (Group V).

number of participants could be estimated in 127 cases. The development of the number of participants and the number of fatalities can be seen in Fig. 9, Table 11 and Fig. 10 shows.

According to the detailed graphs for each group (Fig. 10), it can be summarized that the number of participants (N) and the number of fatal accidents (N_D) according to the 5 personnel groups have linear and polynomial maximum values. They can be characterized according to Table 11. In each case, the maximum values were recorded in a way that could be adjusted to the values of the largest number of victims for each crowd size, using linear regression. In the analysis according to group breakdowns, the linear relationship between the number of participants and the number of victims based on maximum values (Fig. 10) shows an increasing trend in the case of larger events (100,000 ≤ N) (Fig. 10D–E, Fig. 10I–J) and a decreasing trend in the case of smaller events (Fig. 10A–C, Fig. 10F–H). It can be concluded that the average number of participants of the investigated events resulting from crowd motion and ending in fatal accidents was 198,423 people, and the average death rate was 0.74 %. At the 19 events with the smallest number of participants, the number of participants was below 1000 people, of which 9 took place in clubs/indoor entertainment venues, and 6 took place during donation distribution. The smallest number of participants was 120 people.

3.2. *Type of events*

When analyzing the characteristics of accidents at mass events, the type of each event is decisive. The type of events is decisive not only in the investigation of the events of the accidents that have occurred, but also in the determination of the risk factors of the planned events and in the investigation of the movement phenomena. According to the mass gathering event assessment characteristics, the WHO [6] recommends distinguishing at least sports, religious, cultural and political gatherings when investigating the effects on accident risks. At the same time, in connection with crowd accidents, the event types are grouped according to the purpose of each study or other event types based on the available sources (see Table 3). In this study, based on the repetition of the type of accident events and locations, a detailed type of classification containing 11 groups is applied. The more detailed grouping also enables the analyzation of the occurrences of individual events and the environmental characteristics.

Due to the high number of accidents caused by crowd motion during sports and religious events, separate studies have been published to investigate these events [7,27,29]. Among the number of accidents that occurred at all mass events, resulting exclusively from crowd motion, the high number of sports and religious events is confirmed by this study. Based on the classification according to Table 8, among the events without an external physical cause, which can only be traced back to the movement of the crowd, fatal accidents occurred at sports (23 %) and religious (23 %) events in most cases (Fig. 11A). Religious events claimed the largest number of victims (69 %) (Fig. 11B). Compared to the number of participants, most fatalities were in clubs (Fig. 12). In the investigation of the 11 types of events according to 5-year periods, based on both the ratio of fatal accidents (Fig. 13A) and the ratio of the number of events (Fig. 13B), there was a significantly increasing trend in the number of crowd accidents that occurred at donation distributions (Fig. 14A and B).

3.3. *Age of the deceased*

In connection with the age of the injured and deceased in accidents at mass events, summary studies were usually prepared in the field of emergency medical care and accident care. These often describe the age of the accident victims in relation to the type of event. (e.g. average age of injured persons at mass events by type of event: Baseball: 29.1, Football: 33.3, Rock Concerts: 20.1 [56]). Or, at the mass events of a specific event or region, based on the number of individual patient services, injury types, and environmental characteristics (e.g. time and weather) [24,62]. In accidents that can only be traced back to crowd motion, the relationship between the age of the victims and the accidents has so far been little researched. Among the studies so far, however, in some cases the young age of the victims [63] was mentioned as one of the possible risk factors for accidents related to crowd motion. Of the events analyzed in this study, data on the age of the deceased was available in 32 cases, which represented a total of 154 people. According to them, the average age of the victims was 25.60 years (Fig. 15). At the event with the lowest average age of victims, the average age of the deceased was 14.69. For the oldest, it is 81.50 years.

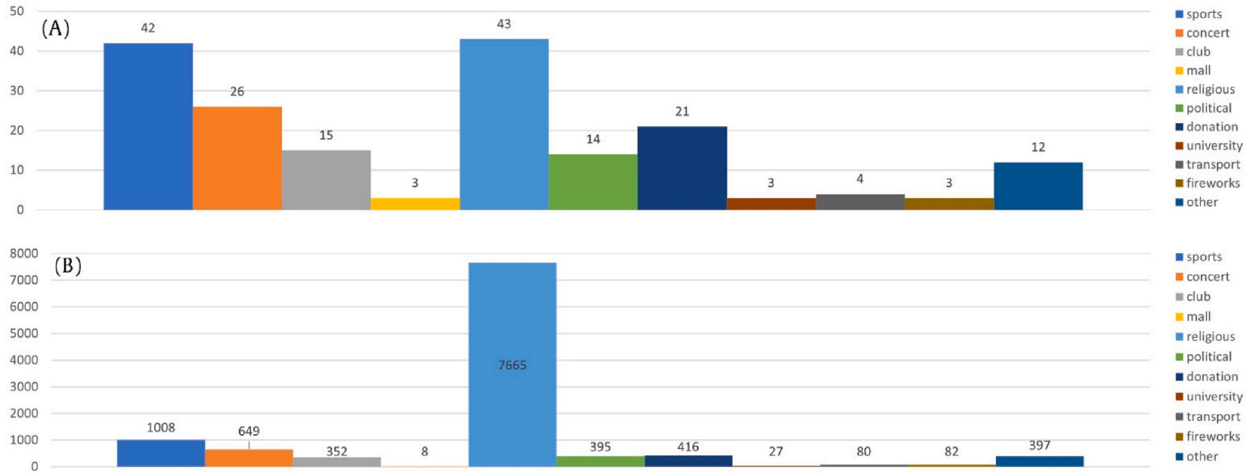

Fig. 11. Type of events (A) Number of events with a fatal outcome due to crowd motion processes, number of events, 1979–2023 (B) Number of events with a fatal outcome due to crowd motion processes, number of victims, 1979–2023.

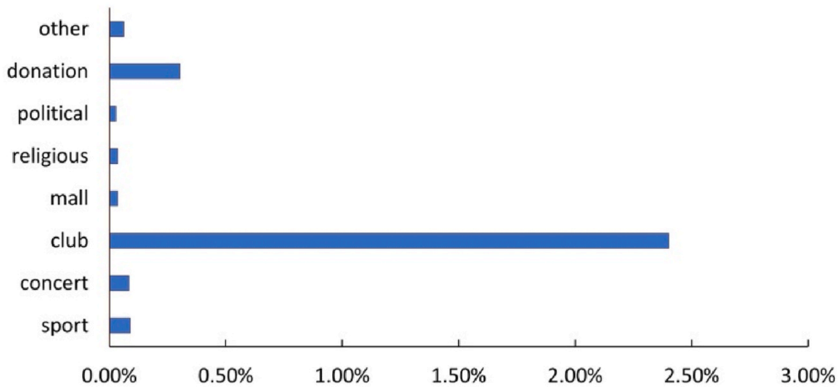

Fig. 12. Number of fatalities caused by crowd crush and the ratio of the number of participants, %, 1979–2023.

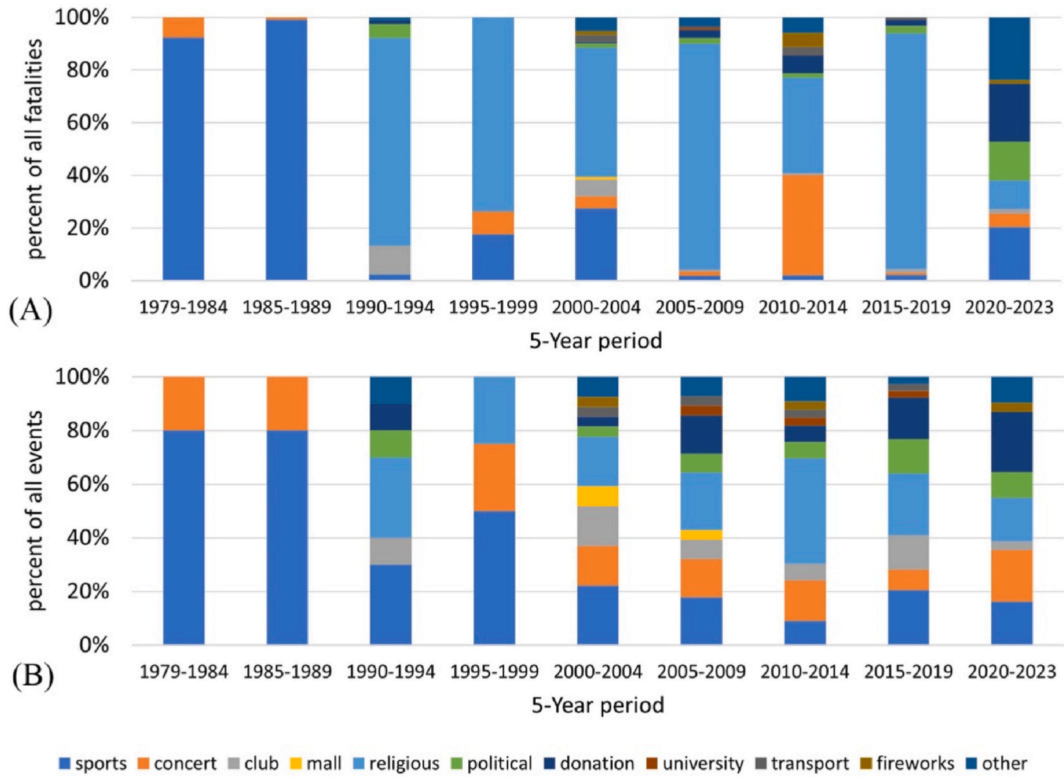

Fig. 13. Number of fatal accidents caused by crowd motion, by type of event, broken down over a 5-year period. (1979–2023) (A) proportion of fatal accidents (B) event number.

3.4. The physical characteristics of the victims

Information on the victims' physical characteristics was available for a total of 2 incidents [7,44]. Due to the small number of available data, relevance cannot be established, the BMI values of the deceased belonged to 5 different body types (Fig. 16).

3.5. Environmental characteristics

A thorough analysis of accidents provides an opportunity to recognize the repetition of certain environmental characteristics. In this study, their combined effect was not investigated. Thus, they are presented as a one-dimensional, percentage effect. During the research, the factors resulting from the characteristics of the immediate location of the accidents were the subject of the investigation, which have already been recognized in research on pedestrian traffic and evacuation - usually have a speed-modifying effect. Their

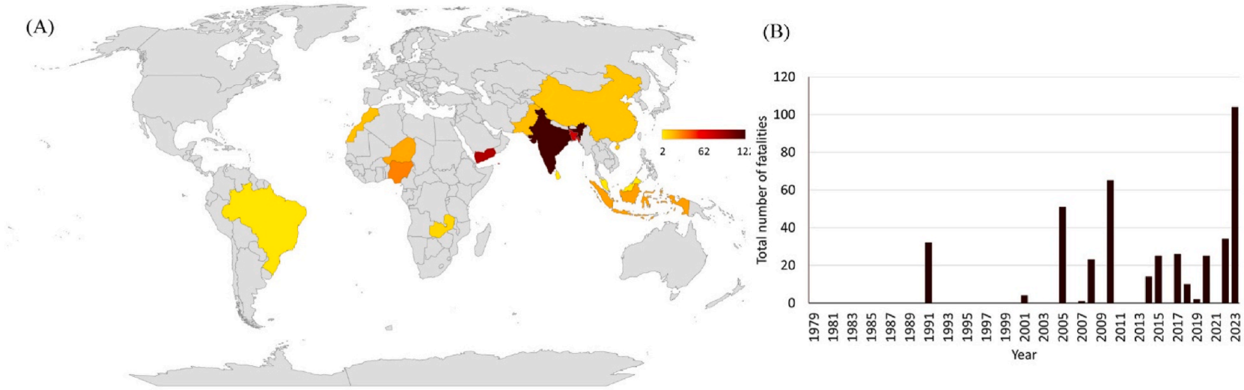

Fig. 14. Fatal accidents caused by crowd movement during donation distribution (1979–2023).

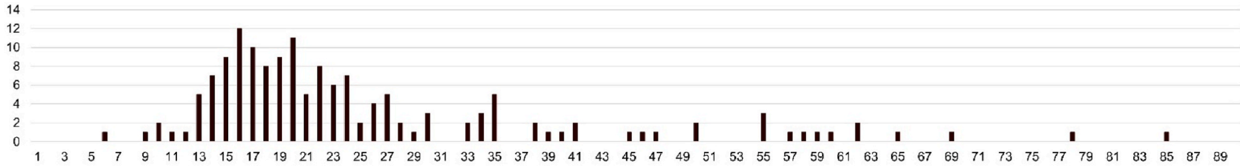

Fig. 15. Age of fatal victims, caused by crowd crush, 1979–2023.

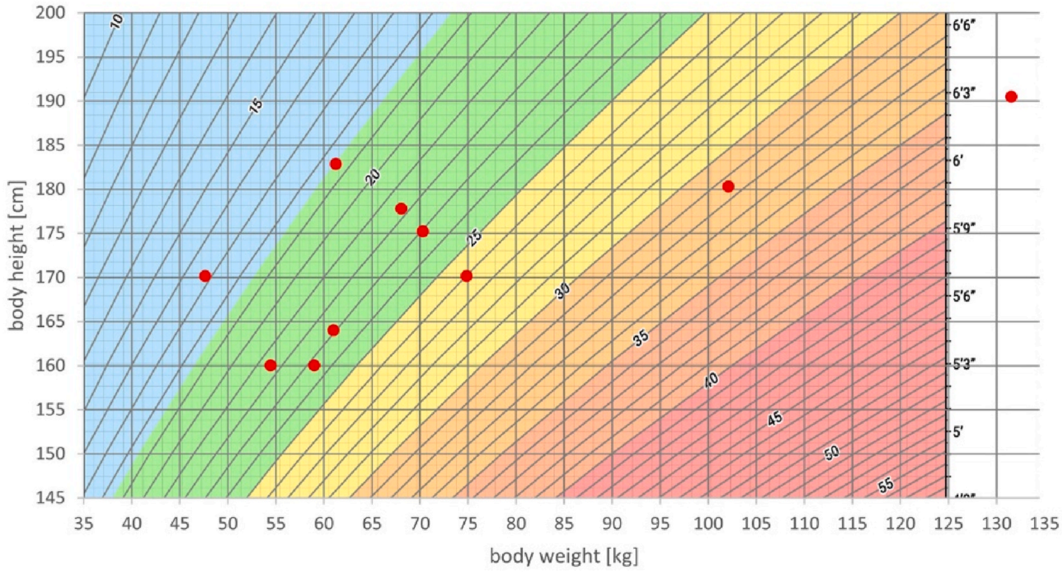

Fig. 16. BMI values of 10 victims, graph wallpaper source: nagualdesign (CC BY-SA 4.0), Colours indicate BMI categories defined by the World Health Organization; underweight, normal weight, overweight, moderately obese, severely obese and very severely obese. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

occurrence was recorded in relation to the number of all accidents, as well as the number of deaths and injuries. At the same time, the available sources did not fully record the events in all accidents, so these findings should be treated as minimum values. In this study, factors that can be traced back to the physical - primarily architectural - characteristics of the accident site, event organization errors, crowd management and crowd behaviour were separately analyzed. Accordingly, in the 186 accident descriptions examined, the factors that could be named as factors that caused the accident or increased its severity in at least three cases were analyzed. Fifteen varied factors could be identified (Fig. 17). However, in each case of an accident, multiple sources of risk could be observed.

Among the risk factors arising from the characteristics of the space, the effect of the presence of gates, ground conditions, vertical traffic bridges (stairs, ramps), tunnels and bridges was analyzed (Table 12). Based on these, the effects of gates and doors were

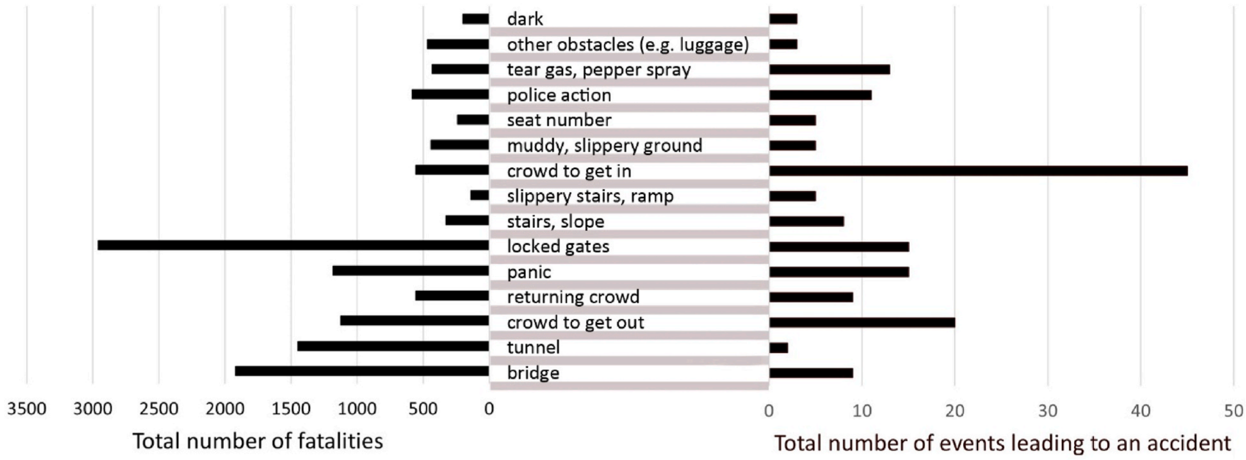

Fig. 17. Cases of events (15 cases type) with a fatal outcome due to crowd motion processes, Total number of fatalities/Total number of events leading to an accident (1979–2023).

Table 12
The impact of event location, physical spatial characteristics, based on the data of 186 accidents attributable to crowd motion (1979–2023).

spatial characteristics	event number [percent (nr/all)]	number of deaths [percent (nr/all)]	number of injured [percent (nr/all)]
doors, gates	43.01 (80/186)	41.83 (4634/11079)	31.59 (5643/17865)
slippery ground	2.69 (5/186)	3.97 (440/11079)	3.40 (608/17865)
stairs, ramp	6.99 (13/186)	4.19 (464/11079)	5.07 (905/17865)
lack of lighting	1.61 (3/186)	1.80 (199/11079)	1.23 (220/17865)
tunnel	1.08 (2/186)	13.06 (1447/11079)	3.21 (575/17865)
bridge	4.84 (9/186)	17.31 (1918/11079)	10.23 (1827/17865)

Table 13
The effect of event location, doors and gates, based on the data of 186 accidents attributable to crowd motion (1979–2023).

crowd progress/motivation	event number [percent (nr/all)]	number of deaths [percent (nr/all)]	number of injured [percent (nr/all)]
exit	10.75 (20/186)	10.14 (1123/11079)	16.64 (2973/17865)
entrance	24.19 (45/186)	5.00 (554/11079)	8.51 (1521/17865)
closed gates	8.06 (15/186)	26.69 (2957/11079)	6.43 (1149/17865)

recognizable in most cases (Table 13). In 43.01 % of fatal accidents caused by crowd motion, doors and gates were a factor influencing the occurrence of accidents. And 31.59 % (5643) of the non-fatal injuries (17,865) of all investigated accidents can be traced back to this factor. The effect of doors and gates on the movement of dense crowds is a continuously active area of research. Several motion model proposals, empirical observations, and computer simulations have been prepared to describe how a large community of people behaves when passing through narrow exits and in its surroundings. These studies usually analyze the relationships between the number of people in motion, the size of the door, the density of personnel, the spatial and temporal process of the movement, or the motivation of the participants [64–68]. Passing through a door or corridor from a space of defined width (so-called “bottleneck effect”) have been investigated by countless studies [69–71]. However, they analyze pedestrian behaviours in normal situations or during evacuation. At the same time, the analysis of 186 fatal crowd accidents that can be traced back to crowd motion in the present study showed that the impact of doors and gates should be considered not only in the planning of normal travel and emergency evacuation processes.

When analyzing the accidents, the impact of traffic openings (doors and gates) could be determined as three additional, clearly distinguishable factors. Thus, the crowd trying to get out, the crowd trying to get in, and the effect of closed gates on accidents. Therefore, they are studied separately (Table 13). It can be concluded that in 24.19 % of all accidents (186), the movement and motivation of the crowd trying to get in was among the factors influencing the outcome of the tragedies (Fig. 18). It is necessary that future research investigates the locations of mass events not only from the point of view of the movement outside, but also from the point of view of the crowd motivated by the desire to get inside.

The effect of the closed gates during this study on the so-called It was recognized as a negative consequence of the phenomenon of “searching for a familiar path”. 8.06 % of all (186) accidents analyzed, and 26.69 % of all fatal accidents, were the result of closed gates. The phenomenon can be explained to get to each location as a search for a familiar way in an emergency. The effect of pre-planned or on-site closure of the inbound route of event participants requires further research.

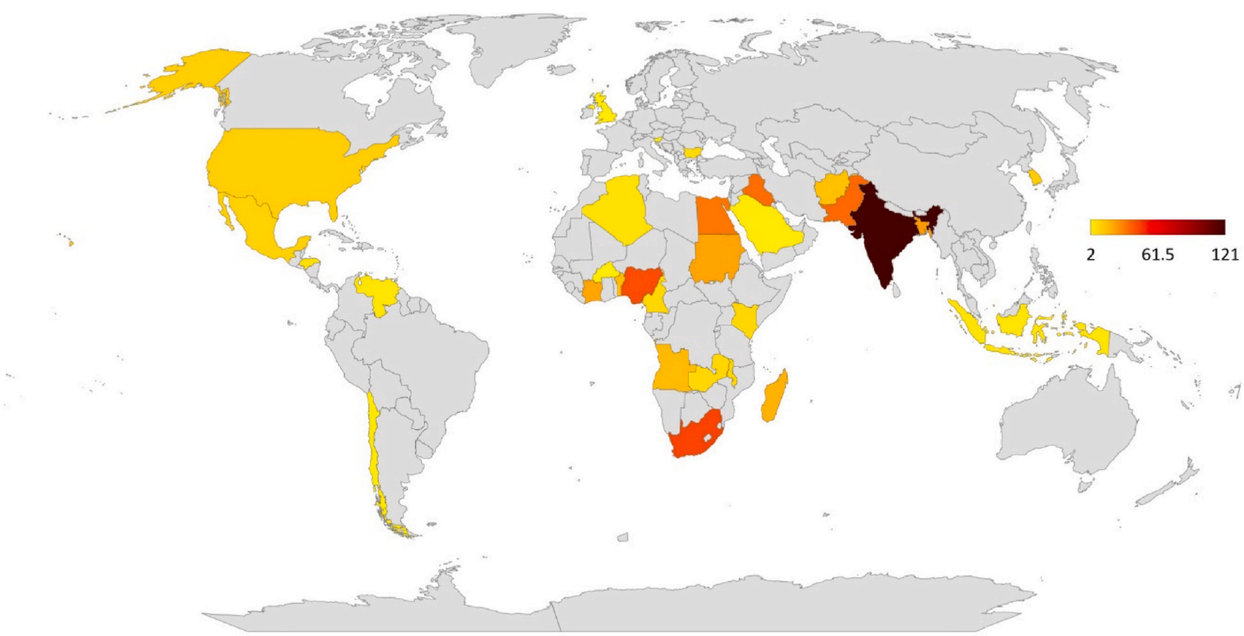

Fig. 18. Fatal accidents, due to the movement of the crowd motivated to enter facility (1979–2023).

The crowd waiting for at least 24 h was documented in at least 4.8 % of all fatal accidents (n = 186, 1979–2023), causing 125 fatal accidents. A logical explanation for the phenomenon can be the connection with the direct health effects of prolonged waiting. At the same time, compared to the estimated number of participants, the death rate of cases in which the crowd waited for at least 24 h was 57.98/100,000 (n = 9, 1979–2023). Due to the small number of samples, no conclusion can be drawn. It can also be assumed that the effect of the long waiting crowd resulted in greater and faster on-site preparedness of healthcare organizations. The effect of prolonged waiting on crowd accidents resulting from crowd motion is an area that has not been researched in detail until now.

3.6. Forms of movement and accidents

In terms of the characteristics of human crowd motions, pedestrian movements in normal situations and emergency crowd motions can be distinguished [72]. The general characteristic of emergency crowd motions is the main motivation of the movement, according to which the individuals want to be in a safe position from some perceived or real source of danger, or they are the victims of the created danger situation. When searching for the cause-and-effect relationships of individual processes, the timing of the events is crucial. According to this, the movement can follow the emergency situation, it can develop almost simultaneously with it, or it can even be the cause of it. Based on the relationship between the movement and the emergency situation, escape, sudden reaction movement occurring in the crowd and events without a prior external physical emergency can also be distinguished. In comparative studies, forms of movement are usually divided into three groups. Hsieh et al. and Ngai et al. in his research, the forms of movement experienced during crowd accidents were classified based on their mechanism into one-way, turbulent and non-determinable categories [28,32]. Alhadhira et al. accidents resulting from crowd density and crowd motion were analyzed based on events documented with photographs and/or video recordings, primarily to determine their physical variables. In their research, they also analyzed the temporality of the events. The movements were analyzed based on the coordinates on the X-, Y- and Z-axis (Z-axis displacement pathology (fall, jump, pushed up, or fall off)) [34].

Different forms of movement observed in dense crowds can result in different accidents. The most frequently observed death in a dense, static crowd is compression asphyxia, which occurs as a result of the force exerted on the chest. Suffocation in a crowd is assumed to occur before the victims fall, in a standing position, so it can often be associated with overcrowding [46,47,73]. Due to their characteristics and the size of the crowd, accidents in large crowds are also called “silent MCI” (silent mass casualty incidents) and can result in a slower response of the emergency medical services (EMS) due to access to the injured [53]. In this study, based on the above, the individual forms of movement that cause accidents are organized based on their basic mechanical nature. Thus, they were divided into quasi-static, mostly quasi-static/not determinable and dynamic groups. Accidents that can be traced back to the high density of the crowd and the forms of movement preceding them are called quasi-static forms of movement. As a dynamic form of crowd motion, phenomena experienced at those events where the progressive movement of the crowd could be determined were marked (Section 2.4, Matrix column 10.).

During our analysis, in the case of 165 fatal crowd accidents, it was possible to infer the nature of the type of movement that caused the tragedy. According to this, 12 % (n_1 = 19, 1979–2023) of the fatal accidents attributable to crowd motion were due to the increased crowd density (in the interpretation of this study: a quasi-static form of crowd motion), and 88 % (n_2 = 146, 1979–2023) were dynamic can be traced back to a movement process. At the same time, quasi-static cases attributable to density (crowd crush) claimed twice as many fatalities (N_{D1} = 2,467, 30 %, 1979–2023) as those attributable to dynamic crowd motion (N_{D2} = 5,843, 70 %, 1979–2023).

4. Discussion

This section summarizes and discusses the most important findings of the study. It also covers the limitations that affect individual results.

The following key findings emerged from our analysis:

(1) (about data sources) In this study, with the introduction of the v_{cr} (values related to the reliability of the data source) value, the reliability value of the information increased by at least 25 % to a quantifiable level, taking into account the changes in the information elements affecting the risk of crowd accidents and the timing of the events. During the analysis, a change in the content of the information on the individual accidents attributable to the movement of the dense mass of people was observed. Monitoring changes in information over time is necessary. At the same time, by using tensors and choosing the appropriate mathematical method of prediction, the data from the very first reports, even from press sources, are the benchmark in the evaluation of the cases. The tensor-cell model analysis also provides the opportunity for the joint analysis of two successive processes in time (here, for example: time elapsed since the accident and time of creation of information about the accident). The reliability of information about fatal accidents occurring in dense crowds, caused by crowd motion and crowd compression, can be increased. Evaluating the credibility of available sources (press, eyewitness accounts, videos, scientific publications, police, judicial, medical experts) provides an opportunity for this. It can be concluded that this information changes over time compared to the time of the accidents. The most reliable information can be found in sources created at least 2 years after the accident occurred. In the data collection methods, it is necessary to record the time of creation of the sources and to use local native language sources of the events. Recording the time of creation of individual sources also facilitates the traceability of information.

(2) (on the number of fatal victims) In the analysis according to logarithmic group breakdowns derived from the published data of fatal accidents attributable to crowd motion, a correlation can be recognized based on the maximum value curves between the number of participants and the number of victims, which in the case of larger events (100,000 $\leq$ N) shows an increasing trend, and in the case of smaller events, a decreasing trend. The results suggest that in the case of dense crowds, the "safety-in-numbers" relationship applies differently below and above 200,000 people.

(3) (on crowd motion) The death rate of accidents (n = 186, 1979–2023) that can only be caused by crowd motion (internal risk cause: overcrowding, crushing, pushing, stampeding) is higher (741/100,000) than that of crowd accidents where other life-threatening factors (external risk causes) cannot be excluded (567/100,000). For the comparative analysis of crowd accidents attributable to crowd motion, it is necessary to exclude other life-threatening conditions (e.g. fire, structural damage, smoke) based on the available information. Isolation of individual forms of movement, recording of information on staff numbers, the environment and the event as accurately as possible. At the same time, this requires a proper evaluation of the information contained in the individual sources, due to the different terminology characteristic of text documents, and due to the proper recognition of the movement processes. During our analysis, in the case of 165 fatal crowd accidents, it was possible to infer the nature of the type of movement that caused the tragedy. According to this, 12 % (n_1 = 19, 1979–2023) of the fatal accidents attributable to crowd motion were due to the increased crowd density (in the interpretation of this study: a quasi-static form of crowd motion), and 88 % (n_2 = 146, 1979–2023) were dynamic can be traced back to a movement process. At the same time, quasi-static cases attributable to density (crowd crush) claimed twice as many fatalities (N_{D1} = 2,467, 30 %, 1979–2023) as those attributable to dynamic crowd motion (N_{D2} = 5,843, 70 %, 1979–2023).

(4) (about motivation) It can be concluded that in 24.19 % of all accidents (n = 186, 1979–2023), the movement and motivation of the crowd trying to get in was among the factors influencing the outcome of the tragedies. The effect of the crowd waiting for at least 24 h on crowd accidents can also be observed. The data on the entry into the buildings are surprising, because it could have been inferred that they fled outside.

(5) (on environmental characteristics) Fifteen environmental features could be identified. Doors and gates were identified as an environmental influencing factor in the largest number of fatal crowd motion accidents (43.01 %, n = 186, 1979–2023). 26.69 % of all fatal accidents analyzed were the result of locked gates.

(6) (about event locations) Due to the recurrence of crowd accidents, 11 distinct types of event locations could be distinguished. From the point of view of crowd accidents caused by crowd motion, based on the number of accidents in relation to the number of participants, clubs were the most dangerous (2.40 %, n = 186, $N_{participant}$ = 25,199,745, 1979–2023). it can be explained by its recognized factors (e.g. frequent deficiencies in event organizers, the use of mind-altering drugs), as well as the architectural and spatial organization characteristics of the function. Religious events claimed the largest number of victims (69 %). The largest number of crowd accidents resulting from crowd motion occurred at sports (23 %) and religious (23 %) events. (n = 186, 1979–2023).

(7) (about injured persons) The number of non-fatal injured persons during the study of each accident had high standard of deviation based on the data provided for the individual events. This can be explained by the different health and emergency response and care of each location, the participants' trust in health care, and the health status of the participants.

During the research, we used mixed sources (press, media, scientific publications, expert reports). Estimated values were available for the number of participants in individual mass casualty events, and the number of all events is unknown.

Our objective in this study was to develop a data collection methodology that can be implemented using simple tools. This methodology provides researchers with adequate flexibility and emphasizes the importance of expert analysis. We illustrate this tensor-based approach through the analysis of crowd accidents arising from mass movements, a globally prevalent issue. While most similar studies assign nearly equal weight to all information sources, we introduce methods to assess the evaluability and enhance the reliability of information. Additionally, we identify risk factors and elements that have not been considered in previous studies (Table 2).

Strengths and limitations: The findings derived from the analysis of crowd crush incidents presented in this study can be applied to the planning and execution of safe mass gatherings. The tensor-cell-based approach can be independently employed for database construction, particularly when the following criteria are met: (A) Joint analysis of information from multiple sources (B) Traceability of recorded information (C) Numerical representation of recorded information (D) Colour-coded representation of recorded information.

Our study has several limitations. Firstly, our database does not encompass all crowd accidents stemming from mass movements. The data is contingent on the reporting systems of individual countries and the formats and content of news reporting. Secondly, manual data collection and expert analysis, despite meticulous efforts to mitigate human error, remain susceptible to inherent biases and limitations of human cognition. However, this process enables flexible research development and the identification of previously unknown associations. Thirdly, the analysis employed a single-factor approach, precluding the examination of the interplay of multiple factors.

5. Conclusions

This study dealt with data collection and analysis of crowd accidents caused by crowd motion. The data collection highlights that the credibility of individual information sources is different. It was also recognized that the content of mass casualty reports often changes over time. The value v_{cr} (values related to the reliability of the data source) was introduced to evaluate the different credibility of the sources. It can be concluded that the new, seven-step, tensor-based data collection method improves the credibility value of individual information by more than 25 %. It was recognized that the construction of tensors from data matrices is suitable for the joint analysis of two consecutive time processes. During the analysis of reports on crowd accidents, it became clear that the death rate is different in the case of an accident caused by crowd motion above a certain number of people. During the analysis, the one-factor, percentage effect of the event, the number of participants, the movement of the crowd, as well as other characteristics known about the participants and the environment, was investigated. It has been recognized that doors and gates have a particularly high-risk effect during fatal accidents caused by crowd motion. The effects of exits, entrances and closed doors were analyzed separately. Based on our results, it can be concluded that the inward movement of the motivated crowd plays an extremely significant role in the development of crowd accidents.

The results suggest that a more thorough study of crowd accidents caused by crowd motion and a more thorough investigation of the risk factors is necessary. During the analyzes combining different time intervals (Figs. 8 and 13), it was recognized that the annual reports on crowd accidents are more recommendable than the analyzes combining 5 or 10-year time periods due to the occurrence of relatively rare events that claim extremely high casualties.

It can also be concluded that adequate information about accidents is necessary. The most accurate determination of the number of participants, the crowd density, and the most reliable and detailed information from the locations are necessary for further research. According to previous research, there is a correlation between the nature of individual mass gatherings and accidents.

However, it can be noted that in many cases the definition of the nature is not clear. The classification of political and religious, religious and donation distribution, political and musical events by nature is often difficult. Qualitative sample collection does not provide a clear answer to the problem either, but at the same time, the other recorded data analyzed (e.g. location, age, form of movement) can help in determining the nature. It is therefore necessary to record as much causal information as possible in databases about crowd accidents, to search for correlations between individual data types. New analyzes are necessary to identify additional risk factors and to verify those identified so far. To evaluate the data on the age and physique of the victims in each of the cases that have occurred, additional information is needed, both for the crowd present and for the individual accidents. To prevent accidents resulting from crowd motion, we consider the close cooperation of individual organizations, researchers, experts and the profession to be absolutely necessary.

To increase crowd safety, we have proposed a new, seven-step data collection method for the analysis of crowd accidents. In this way, the data collection is detailed, yet uniform to the extent necessary, traceable, the credibility value of the sources can be increased, and their changes over time can be considered. By analyzing crowd accidents that have occurred, we founded risk factors in their percentage effects. We identified new risk factors influencing the occurrence of crowd accidents.

The authors further expand the study of crowd accidents that can be traced back to crowd motion, focusing on the environmental characteristics and the physical investigation of individual movement phenomena.

CRediT authorship contribution statement

Dóra Edelmann: Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arnold Őszi:** Validation, Supervision, Project administration, Investigation, Conceptualization. **Tibor Goda:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation,

Conceptualization.

Data availability statement

Data is available upon request.

Funding sources

This research did not receive any specific funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Sharma, B. McCloskey, D.S. Hui, A. Rambia, A. Zumla, T. Traore, S. Shafi, S.A. El-Kafrawy, E.I. Azhar, A. Zumla, A.J. Rodriguez-Morales, Global mass gathering events and deaths due to crowd surge, stampedes, crush and physical injuries – lessons from the Seoul Halloween and other disasters, Trav. Med. Infect. Dis. 52 (2023), https://doi.org/10.1016/j.tmaid.2022.102524.
[2] W. Weng, J. Wang, L. Shen, Y. Song, Review of analyses on crowd-gathering risk and its evaluation methods, Journal of Safety Science and Resilience 4 (2023) 93–107, https://doi.org/10.1016/j.jnlssr.2022.10.004.
[3] A.M. Margolis, A.K. Leung, M.S. Friedman, S.P. McMullen, F.X. Guyette, N. Woltman, Position statement: mass gathering medical care, Prehosp. Emerg. Care 25 (2021), https://doi.org/10.1080/10903127.2021.1903632.
[4] C. Feliciani, A. Corbetta, M. Haghani, K. Nishinari, Trends in crowd accidents based on an analysis of press reports, Saf. Sci. 164 (2023), https://doi.org/10.1016/j.ssci.2023.106174.
[5] M.M. De Almeida, J. Von Schreeb, Human stampedes: an updated review of current literature, Prehospital Disaster Med. 34 (2019), https://doi.org/10.1017/S1049023X18001073.
[6] World Health Organization, Public Health for Mass Gatherings: Key Considerations, World Health Organization, 2015.
[7] D. Tin, R. Hata, G.R. Ciottone, The anatomy of stadium disasters: causes, consequences and safeguarding the future from a medical perspective, Publ. Health 224 (2023), https://doi.org/10.1016/j.puhe.2023.08.014.
[8] P. Lu, Z. Zhang, M. Li, D. Chen, H. Yang, Agent-based modeling and simulations of terrorist attacks combined with stampedes, Knowl. Base Syst. 205 (2020), https://doi.org/10.1016/j.knosys.2020.106291.
[9] Y. Lu, L. Qiu, X. Lyu, X. Jiang, Human stampede causative factors and cluster risk: a multi-dimensional analysis based on ISODATA and Fuzzy Theory, Int. J. Disaster Risk Reduc. 66 (2021), https://doi.org/10.1016/j.ijdrr.2021.102581.
[10] Y. Lu, X. Shi, X. Jiang, J. Tang, Analyzing dynamic risk of stampede in stadium: a quantitative method considering the various status of risk factors in whole process, Int. J. Disaster Risk Reduc. 82 (2022), https://doi.org/10.1016/j.ijdrr.2022.103339.
[11] X. Hu, H. Zhao, Y. Bai, J. Wu, Risk analysis of stampede in sporting venues based on catastrophe theory and Bayesian network, Int. J. Disaster Risk Reduc. 78 (2022), https://doi.org/10.1016/j.ijdrr.2022.103111.
[12] J. Fruin, The causes and prevention of crowd disasters. First International Conference on Engineering for Crowd Safety, 1993.
[13] J. Santos-Reyes, S. Olmos-Peña, Analysis of the 'news divine' stampede disaster, Saf. Sci. 91 (2017), https://doi.org/10.1016/j.ssci.2016.07.014.
[14] B. Yogameena, C. Nagananthini, Computer vision based crowd disaster avoidance system: a survey, Int. J. Disaster Risk Reduc. 22 (2017), https://doi.org/10.1016/j.ijdrr.2017.02.021.
[15] I. Gursel Dino, E. Kalfaoglu, O.K. Iseri, B. Erdogan, S. Kalkan, A.A. Alatan, Vision-based estimation of the number of occupants using video cameras, Adv. Eng. Inf. 53 (2022), https://doi.org/10.1016/j.aei.2022.101662.
[16] A. Mahdavi, F. Tahmasebi, On the quality evaluation of behavioural models for building performance applications, J Build Perform Simul 10 (2017), https://doi.org/10.1080/19401493.2016.1230148.
[17] F. Huo, Y. Li, C. Li, Y. Ma, An extended model describing pedestrian evacuation considering pedestrian crowding and stampede behavior, Phys. Stat. Mech. Appl. 604 (2022), https://doi.org/10.1016/j.physa.2022.127907.
[18] D. Barr, J. Drury, T. Butler, S. Choudhury, F. Neville, Beyond 'stampedes': towards a new psychology of crowd crush disasters, Br. J. Soc. Psychol. (2023), https://doi.org/10.1111/bjso.12666.
[19] H. Lügering, D. Tepeli, A. Sieben, It's (not) just a matter of terminology: everyday understanding of "mass panic" and alternative terms, Saf. Sci. 163 (2023) 106–123, https://doi.org/10.1016/j.ssci.2023.106123.
[20] J. Adrian, N. Bode, M. Amos, M. Baratchi, M. Beermann, M. Boltes, A. Corbetta, G. Dezecache, J. Drury, Z. Fu, R. Geraerts, S. Gwynne, G. Hofinger, A. Hunt, T. Kanters, A. Kneidl, K. Konya, G. Köster, M. Küpper, G. Michalareas, F. Neville, E. Ntontis, S. Reicher, E. Ronchi, A. Schadschneider, A. Seyfried, A. Shipman, A. Sieben, M. Spearpoint, G.B. Sullivan, A. Templeton, F. Toschi, Z. Yücel, F. Zanlungo, I. Zuriguel, N. Van der Wal, F. van Schadewijk, C. von Krüchten, N. Wijermans, A glossary for research on human crowd dynamics, Collect Dyn 4 (2019) 1–13, https://doi.org/10.17815/cd.2019.19.
[21] P.H.E. and partners World Health Organization. Health Emergency and Disaster Risk Management MASS GATHERINGS, 2017.
[22] V.T. Hoang, P. Gautret, Infectious diseases and mass gatherings, Curr. Infect. Dis. Rep. 20 (2018), https://doi.org/10.1007/s11908-018-0650-9.
[23] C. Feliciani, A. Corbetta, M. Haghani, K. Nishinari, How crowd accidents are reported in the news media: lexical and sentiment analysis, Saf. Sci. 172 (2024) 106423, https://doi.org/10.1016/j.ssci.2024.106423.
[24] M.M. Al-Hayani, S. Kamel, A.M. Al-Hayani, E.A. Al-Hazmi, M.S. Al-Shanbari, N.S. Al-Otaibi, A.S. Almeshal, A.M. Assiri, Trauma and injuries pattern during Hajj, 1443 (2022): a cross-sectional study, Cureus (2023), https://doi.org/10.7759/cureus.41751.
[25] Q.A. Ahmed, Y.M. Arabi, Z.A. Memish, Health risks at the Hajj, Lancet 367 (2006), https://doi.org/10.1016/S0140-6736(06)68429-8.
[26] Y.A. Alaska, A.D. Aldawas, N.A. Aljerian, Z.A. Memish, S. Suner, The impact of crowd control measures on the occurrence of stampedes during Mass Gatherings: the Hajj experience, Trav. Med. Infect. Dis. 15 (2017), https://doi.org/10.1016/j.tmaid.2016.09.002.
[27] F.T. Illiyas, S.K. Mani, A.P. Pradeepkumar, K. Mohan, Human stampedes during religious festivals: a comparative review of mass gathering emergencies in India, Int. J. Disaster Risk Reduc. 5 (2013), https://doi.org/10.1016/j.ijdrr.2013.09.003.
[28] K.M. Ngai, W.Y. Lee, A. Madan, S. Sanyal, N. Roy, F.M. Burkle, E.B. Hsu, Comparing two epidemiologic surveillance methods to assess underestimation of human stampedes in India, PLoS Curr (2013), https://doi.org/10.1371/currents.dis.ab7f298c89854015b74856232c70b62c.
[29] G. Harihara Subramanian, A. Verma, Crowd risk prediction in a spiritually motivated crowd, Saf. Sci. 155 (2022), https://doi.org/10.1016/j.ssci.2022.105877.
[30] World Health Organization, Public Health for Mass Gatherings (2015), 2015.
[31] M. Haghani, Empirical methods in pedestrian, crowd and evacuation dynamics: Part II. Field methods and controversial topics, Saf. Sci. 129 (2020) 104760, https://doi.org/10.1016/j.ssci.2020.104760.

[32] Y.H. Hsieh, K.M. Ngai, F.M. Burkle, E.B. Hsu, Epidemiological characteristics of human stampedes, Disaster Med. Public Health Prep. 3 (2009), https://doi.org/10.1097/DMP.0b013e3181c5b4ba.
[33] L. Soomaroo, V. Murray, Disasters at mass gatherings: lessons from history, PLoS Curr (2012), https://doi.org/10.1371/currents.RRN1301.
[34] A. Alhadhira, M.S. Molloy, M. Casasola, R.R. Sarin, M. Massey, A. Voskanyan, G.R. Ciottone, Use of dimensional analysis in the X-, Y-, and Z-Axis to predict occurrence of injury in human stampede, Disaster Med. Public Health Prep. 14 (2019), https://doi.org/10.1017/dmp.2019.47.
[35] G.K. Still, Introduction to Crowd Science, first ed., CRC Press, 2014.
[36] M. Zhang, H. Cui, X. Tian, B. Kang, L. Huang, Evaluate the reliability of information sources using the non-parametric plausibility ReliefF algorithm for multi-source information fusion, Appl. Soft Comput. 148 (2023), https://doi.org/10.1016/j.asoc.2023.110871.
[37] R.C. Chyne, J. Khongtim, T. Wann, Evaluation of social media information among college students: an information literacy approach using CCOW, J. Acad. Librarian 49 (2023), https://doi.org/10.1016/j.acalib.2023.102771.
[38] A. Sieben, A. Seyfried, Inside a life-threatening crowd: analysis of the Love Parade disaster from the perspective of eyewitnesses, Saf. Sci. 166 (2023), https://doi.org/10.1016/j.ssci.2023.106229.
[39] B. Krausz, C. Bauckhage, Loveparade 2010: automatic video analysis of a crowd disaster, Comput. Vis. Image Understand. 116 (2012) 307–319, https://doi.org/10.1016/j.cviu.2011.08.006.
[40] D. Helbing, P. Mukerji, Crowd disasters as systemic failures: analysis of the love parade disaster, EPJ Data Sci 1 (2012), https://doi.org/10.1140/epjds7.
[41] U. Chattaraj, A. Seyfried, P. Chakroborty, Comparison of pedestrian fundamental diagram across cultures, Adv. Complex Syst. 12 (2009), https://doi.org/10.1142/S0219525909002209.
[42] L. Huang, T. Chen, Y. Wang, H. Yuan, Congestion detection of pedestrians using the velocity entropy: a case study of Love Parade 2010 disaster, Phys. Stat. Mech. Appl. 440 (2015) 200–209, https://doi.org/10.1016/j.physa.2015.08.013.
[43] J. Ma, W.G. Song, S.M. Lo, Z.M. Fang, New insights into turbulent pedestrian movement pattern in crowd-quakes, J. Stat. Mech. Theor. Exp. 2013 (2013) 1–22, https://doi.org/10.1088/1742-5468/2013/02/P02028.
[44] J.R. Gill, K. Landi, Traumatic asphyxial deaths due to an uncontrolled crowd, Am. J. Forensic Med. Pathol 25 (2004), https://doi.org/10.1097/01.paf.0000147316.62883.8b.
[45] N.K. Tumram, V.N. Ambade, N. Biyabani, Compression asphyxia from a human pyramid, Med. Leg. J. 83 (2015), https://doi.org/10.1177/0025817215588884.
[46] D. DeAngeles, M. Schurr, M. Birnbaum, B. Harms, Traumatic asphyxia following stadium crowd surge: stadium factors affecting outcome, Wis. Med. J. 97 (1998).
[47] J.P. Nolan, J. Soar, N. Cary, N. Cooper, J. Crane, A. Fegan-Earl, W. Lawler, P. Lumb, G. Rutty, Compression asphyxia and other clinicopathological findings from the Hillsborough Stadium disaster, Emerg. Med. J. 38 (2021) 798–802, https://doi.org/10.1136/emermed-2020-209627.
[48] T. Motomura, H. Matsumoto, H. Yokota, M. Suzuki, T. Nishimoto, S. Ujihashi, Thoracoabdominal compression model of traumatic asphyxia to identify mechanisms of respiratory failure in fatal crowd accidents, J. Nippon Med. Sch. 86 (2019), https://doi.org/10.1272/jnms.JNMS.2019_86-607.
[49] B. Colville-Ebeling, M. Freeman, J. Banner, N. Lynnerup, Autopsy practice in forensic pathology - evidence-based or experience-based? A review of autopsies performed on victims of traumatic asphyxia in a mass disaster, J Forensic Leg Med 22 (2014), https://doi.org/10.1016/j.jflm.2013.11.006.
[50] N. Fijačko, J.P. Nolan, G. Štiglic, P. Kocbek, R. Greif, Content analysis of Twitter users' responses to the crowd crush tragedy in Seoul, South Korea in October 2022, Resuscitation 182 (2023), https://doi.org/10.1016/j.resuscitation.2022.11.024.
[51] Y.H. Dong, F. Liu, Y.M. Liu, X.R. Jiang, Z.X. Zhao, Emergency Preparedness for Mass Gatherings: Lessons of "12.31" Stampede, twenty zero ed., Shanghai Bund, Chinese Journal of Traumatology - English, 2017 https://doi.org/10.1016/j.cjtee.2016.08.005.
[52] P. Paciorek, K. Wisniewski, F. Oleszak, B. Zawada, An a analysis of the tragic events which occurred at an inauguration party organized by the student government of the University of Science and Technology in Bydgoszcz during the 2015/2016 academic year, Disaster and Emergency Medicine Journal 2 (2017), https://doi.org/10.5603/demj.2017.0016.
[53] D.T.G. Daniel, E.A. Alpert, E. Jaffe, The crowd crush at mount meron: emergency medical services response to a silent mass casualty incident, Disaster Med. Public Health Prep. 16 (2022), https://doi.org/10.1017/dmp.2022.162.
[54] K. Lawrence, Events Management and Audience Safety: What Went Wrong at Astroworld? (2022), https://doi.org/10.4135/9781529600506.
[55] J. Mao, A study on emergency management policy triggered by the crowd crush in itaewon, South Korea, journal of education, Humanit. Soc. Sci. 15 (2023), https://doi.org/10.54097/ehss.v15i.9109.
[56] A.M. Milsten, K.G. Seaman, P. Liu, R.A. Bissell, B.J. Maguire, Variables influencing medical usage rates, injury patterns, and levels of care for mass gatherings, Prehospital Disaster Med. (2003), https://doi.org/10.1017/S1049023X00001291.
[57] C. McQueen, C. Davies, Health care in a unique setting: applying emergency medicine at music festivals, Open Access Emerg. Med. 4 (2012), https://doi.org/10.2147/OAEM.S25587.
[58] E. Meites, J.F. Brown, Ambulance need at mass gatherings, Prehospital Disaster Med. 25 (2010), https://doi.org/10.1017/S1049023X00008682.
[59] J. Yin, X. min Zheng, R.C. Tsaur, Occurrence mechanism and coping paths of accidents of highly aggregated tourist crowds based on system dynamics, PLoS One 14 (2019), https://doi.org/10.1371/journal.pone.0222389.
[60] R. Elvik, T. Bjørnskau, Safety-in-numbers: a systematic review and meta-analysis of evidence, Saf. Sci. 92 (2017), https://doi.org/10.1016/j.ssci.2015.07.017.
[61] R. Elvik, R. Goel, Safety-in-numbers: an updated meta-analysis of estimates, Accid. Anal. Prev. 129 (2019), https://doi.org/10.1016/j.aap.2019.05.019.
[62] O. Anikeeva, P. Arbon, K. Zeitz, M. Bottema, A. Lund, S. Turris, M. Steenkamp, Patient presentation trends at 15 mass-gathering events in South Australia, Prehospital Disaster Med. 33 (2018), https://doi.org/10.1017/S1049023X1800050X.
[63] F.M. Burkle, E.B. Hsu, Ram Janki temple: understanding human stampedes, Lancet 377 (2011), https://doi.org/10.1016/S0140-6736(10)60442-4.
[64] A. Nicolas, S. Bouzat, M.N. Kuperman, Pedestrian flows through a narrow doorway: effect of individual behaviours on the global flow and microscopic dynamics, Transp. Res. Part B Methodol. 99 (2017), https://doi.org/10.1016/j.trb.2017.01.008.
[65] I. Zuriguel, I. Echeverría, D. Maza, R.C. Hidalgo, C. Martín-Gómez, A. Garcimartín, Contact forces and dynamics of pedestrians evacuating a room: the column effect, Saf. Sci. 121 (2020), https://doi.org/10.1016/j.ssci.2019.09.014.
[66] W. Daamena, S. Hoogendoorn, Capacity of doors during evacuation conditions, Procedia Eng. (2010), https://doi.org/10.1016/j.proeng.2010.07.007.
[67] I.M. Sticco, G.A. Frank, C.O. Dorso, Faster and safer evacuations induced by closed vestibules, Simul Model Pract Theory 128 (2023), https://doi.org/10.1016/j.simpat.2023.102818.
[68] M. Imanishi, A. Jo, T. Sano, Effects of pedestrian motivation and opening shape on pedestrian flow rate at an opening, Fire Saf. J. 120 (2021), https://doi.org/10.1016/j.firesaf.2020.103056.
[69] H. Li, J. Zhang, W. Song, A comparative experimental study on the influence of bottleneck width on evacuation characteristics of pedestrian flow in funnel shape bottleneck and normal bottleneck, Phys. Stat. Mech. Appl. 624 (2023), https://doi.org/10.1016/j.physa.2023.128929.
[70] J. Liddle, A. Seyfried, B. Steffen, W. Klingsch, T. Rupprecht, A. Winkens, M. Boltes, Microscopic insights into pedestrian motion through a bottleneck, resolving spatial and temporal variations, Collect Dyn 7 (2022) 1–23, https://doi.org/10.17815/CD.2022.139.
[71] X. Shi, Z. Ye, N. Shiwakoti, D. Tang, J. Lin, Examining effect of architectural adjustment on pedestrian crowd flow at bottleneck, Phys. Stat. Mech. Appl. 522 (2019), https://doi.org/10.1016/j.physa.2019.01.086.
[72] J. Liu, Y. Chen, Y. Chen, Emergency and disaster management-crowd evacuation research, J Ind Inf Integr 21 (2021) 1–9, https://doi.org/10.1016/j.jii.2020.100191.
[73] N.K. Tumram, V.N. Ambade, P.G. Dixit, Compression asphyxia in upright suspended position, Am. J. Forensic Med. Pathol 35 (2014), https://doi.org/10.1097/PAF.0000000000000052.