

## Article

# Does the Integration of Pre-Coded Information with Narratives Improve in-Hospital Falls' Surveillance?

Giulia Lorenzoni <sup>1</sup>, Roberta Rampazzo <sup>1</sup>, Alessia Buratin <sup>1</sup>, Paola Berchiolla <sup>2</sup> and Dario Gregori <sup>1,\*</sup>

<sup>1</sup> Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova, 35131 Padova, Italy; giulia.lorenzoni@unipd.it (G.L.); roberta.rampazzo.1@studenti.unipd.it (R.R.); alessia.buratin@unipd.it (A.B.)

<sup>2</sup> Department of Clinical and Biological Sciences, University of Torino, 10100 Torino, Italy; paola.berchiolla@unito.it

\* Correspondence: dario.gregori@unipd.it; Tel.: +39-049-827-5384; Fax: +39-02-70-044-5089

**Abstract:** To evaluate the value added by information reported in narratives (extracted through text mining techniques) in enhancing the characterization of falls patterns. Data on falls notified to the Risk Management Service of a Local Health Authority in Italy were considered in the analysis. Each record reported detailed pre-coded information about patient and fall's characteristics, together with a narrative description of the fall. At first, multiple correspondence analysis (MCA) was performed on pre-coded information only. Then, it was re-run on the pre-coded data augmented with a variable representing the output analysis of the narrative records. This second analysis required a pre-processing of the narratives followed by text mining. Finally, a Hierarchical Clustering on the two MCA was carried out to identify distinct fall patterns. The dataset included 202 falls' records. Three clusters corresponding to three distinct profiles of falls were identified through the Hierarchical Clustering performed using only pre-coded information. Hierarchical Clustering with the topic variable provided overlapping results. The present findings showed that the cluster analysis is effective in characterizing fall patterns; however, they do not sustain the hypothesis that the analysis of free-text information improves our understanding of such phenomenon.

**Keywords:** falls; narratives; text mining; cluster analysis



**Citation:** Lorenzoni, G.; Rampazzo, R.; Buratin, A.; Berchiolla, P.; Gregori, D. Does the Integration of Pre-Coded Information with Narratives Improve in-Hospital Falls' Surveillance? *Appl. Sci.* **2021**, *11*, 4406. <https://doi.org/10.3390/app11104406>

Academic Editor: Marco Manfrini

Received: 7 April 2021  
Accepted: 10 May 2021  
Published: 13 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Falls in the hospital setting are related to increased morbidity and mortality and a higher cost of care for the injured patient [1,2]. The prevention of such events is crucial to reduce the public health burden of in-hospital falls, which represent the most frequent adverse event occurring during the hospitalization [3].

Several risk factors have been identified for these events, including both risk factors related to the patient and related to the environment [4,5]. Organizations worldwide involved in the improvement of the quality of care and patients' safety have developed initiatives for falls' risk management [6,7]. The core recommendations are represented by patients' risk assessment using validated tools, and the development and implementation of multidisciplinary personalized fall prevention plans modifying all the potential risk factors identified.

Even though substantial efforts have been put forwards for falls' risk management, this phenomenon continues to be worrying with more than 700,000 patients falling every year in the United States' hospitals [8]. It would be of interest understanding the reasons why falls continue to be so frequent. It might be hypothesized that risk management interventions are not adequately implemented in the in-hospital setting; however, it could be suggested that such interventions might be ineffective since the phenomenon of falls has not yet been fully understood.

Systematic epidemiological surveillance of the phenomenon is crucial to better characterize falls. In Italy, it has been established a data collection system of in-hospital sentinel events, including fatal falls and falls resulting in severe injury. Furthermore, it is recommended [9] to the Local Health Authorities to report all falls (not necessarily resulting in severe injury or in the death of the patient). However, such reporting systems are highly variable among Italian Local Health Authorities. Anyway, generally, they are based on a data collection form that might include both pre-coded information and narratives. Therefore, it is crucial to adopt strategies of analyses allowing for the effective exploitation of information reported in narratives, since the manual revision of such information is unfeasible. The use of text mining would be a promising opportunity to exploit such information especially thank to the wide implementation of electronic health records [10], and it has been suggested to have a good performance [11]; however, applications in the field of injury surveillance are sparse.

Narratives could be a great source of additional information helping to better understand the phenomenon of injuries because it has been shown that integrating narratives with pre-coded information in injury surveillance can result in a better understanding of injury patterns [12]. Such integrated information would serve as input for ad hoc data analysis approaches aimed at detecting and grouping observations with similar characteristics (e.g., cluster analysis) each one potentially identifying a specific pattern of injury i.e., in terms of subjects' characteristics and falls' characteristics. The recognition of such patterns would serve for guiding the development of further primary prevention strategies in the field.

The present study aimed to evaluate the value added by information reported in narratives (extracted through text mining techniques) in enhancing the characterization of falls patterns. To achieve this aim, the data from the incident reporting system of an Italian Local Health Authority were used.

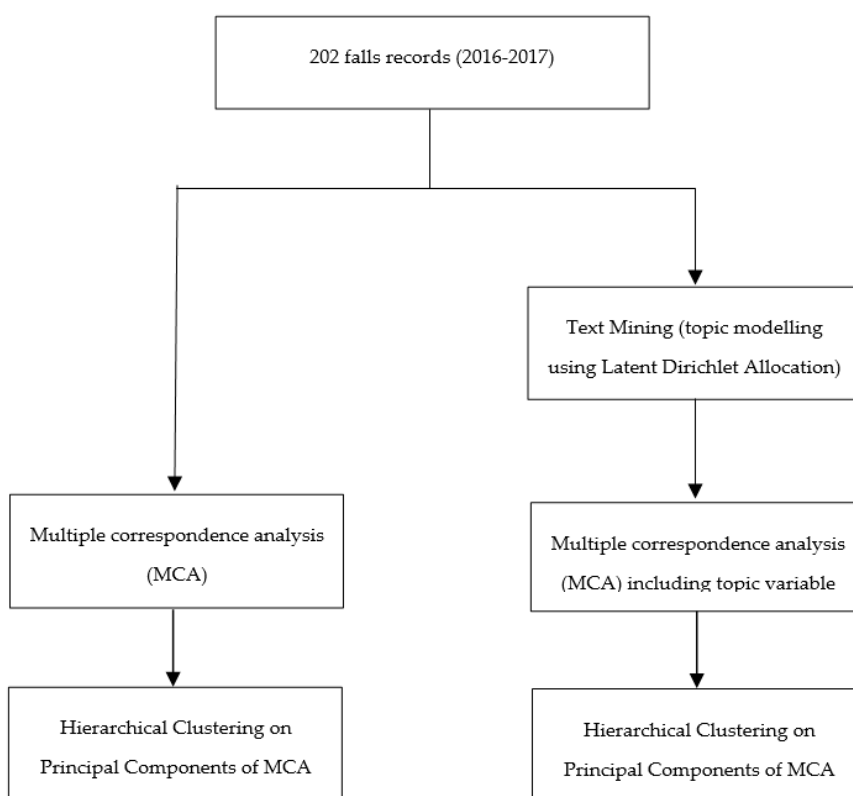
## 2. Materials and Methods

Falls notified to the Risk Management Service of a Local Health Authority in Italy (name not disclosed as per research agreement) were considered in the analysis.

The Local Health Authority included two hospitals and three health districts aimed at outpatients' care. Falls are notified via a data collection form made of two parts. In the first part, information was provided by the health care professional involved in patient's care (e.g., nurse, physical therapist, midwife) and included socio-demographic information, risk factors for falls and strategies adopted for the management of the falls' risk management, fall's characteristics (place and dynamics), potential causes of the fall, and a brief description of the event. The second section of the data collection tool was under the physician's responsibility and reported the potential complications resulting from the fall (injury type, procedures performed for the injury's diagnosis, therapeutic interventions performed) and the drug therapy of the patient at the time of the fall.

All the information was pre-coded except for the event's description that was a narrative. The data analysis workflow is summarized in Figure 1.

At first, multiple correspondence analysis (MCA) was performed on pre-coded information only. Then, it was re-rerun on the pre-coded data augmented with a variable representing the output analysis of the narrative records. This second analysis required a pre-processing of the narratives followed by text mining, which consisted in the implementation of topic modelling, using Latent Dirichlet Allocation (LDA) model, for clustering the narrative records into topics. Finally, a Hierarchical Clustering on the two MCAs was carried out to identify distinct fall patterns.



**Figure 1.** Flow-chart presenting the summary of the study procedures.

### 2.1. Pre-Processing of Narratives

The pre-processing of narratives consisted in the tokenization, stop words removal, and stemming. The first step of the pre-processing was the tokenization, by which the text was split into words. Then, a stop words list was identified including the most common words in the text of the documents (mainly articles, prepositions, and pronouns). Stop words were removed from the text along with terms with less than four characters. Finally, the Martin Porter's stemming procedure [13] implemented in the SnowballC R package was used to identify the root/stem of each word.

### 2.2. Text Mining: Latent Dirichlet Allocation Model

LDA [14] model is the most popular probabilistic topic modelling. It assumes that each document contains a mixture of topics, and each topic is a probability distribution over words. LDA model was used to detect topics inside the pre-processed narratives.

To carry out the LDA model, a Dirichlet prior was specified on the per-topic word distributions and on the per-document topic distribution. According to Griffiths and Styvers [15], the parameters were set equal to 0.1 and to 50 divided by the number of topics ( $50/k$ ), respectively.

The input data for the LDA was the document-term matrix built over the pre-processed narratives records, each row corresponding to a record and each column to a term. In addition, to reduce the sparsity of the document-term matrix, only terms that occurred in a minimum number of 5 documents were retained. Finally, we used the mean term frequency inverse document frequency (TF-IDF) over the documents to reduce the sparsity of document-term matrix. TF-IDF allows for removing both less frequent and very common terms [16].

Different models with a number of topics ranging from 2 to 50 were fitted and estimated using Gibbs sampling [17] with a burn-in of 2000 iterations, retaining 1 every 100 iterations for the subsequent 1000 iterations. Model selection with respect to the num-

ber of topics was done by splitting the data into training and test sets and using 10-fold cross-validation.

The perplexity measure, a standard way of evaluating language models, i.e., the probability distribution constructed over the text, was used to assess the topic model on hold out data and then to identify the best number of topics [18]. Perplexity is the geometric mean per word-likelihood and that can be interpreted how well a probability model predicts a sample.

### 2.3. Multiple Correspondence Analysis

As usual when analyzing questionnaire data, multiple correspondence analysis (MCA) was used as a pre-process step before clustering [19]. MCA allows for coding the categorical variables into a set of principal components.

MCA was performed both on pre-coded information only and on the pre-coded information together with the information extracted from narratives. To include the information from the text mining of the events, a supplementary variable (topic variable) that identified each narrative with the most represented topic was added.

### 2.4. Hierarchical Clustering on Principal Components

The hierarchical clustering of the principal components (HCPC) obtained through MCA was used for the clustering process [20]. Similarity measures were generated computing the Euclidean distance on the five first components by Ward's method [21]. The number of clusters was set equal at that one suggested by the algorithms, which was based on the ratio of the variance within-clusters and the total variance.

For each category of categorical variables, the V test was computed. The V-test was used to assess which categories are over-represented in the clusters. The V-test statistics can be considered as a standardized deviation between the mean of those individuals with the specific category and the general average [21]. The categories that are over-represented in the clusters have a V-test statistics greater than 1.96 ( $p$ -value < 0.05). The higher the statistics, the more strongly the category is over-represented in the cluster.

### 2.5. Model Validity

To compare the performance of the analyses with and without the topic variable (corresponding to the information retrieved from narratives), the within-cluster sum of square (WSS) was computed [22]. WSS is a standard measure of the goodness of a clustering structure without considering external information and indicates how closely related are the individuals in a cluster. For our purpose, WSS was not used to determine the number of clusters, but just to formalize the within-groups sum of squares distances.

### 2.6. Statistical Analysis

Descriptive statistics were reported as a percentage and frequency for the categorical variables. Analyses were performed using R System ver 3.5.1 [23]. The R packages FactoMiner [24] and MissMDA [25] were used to perform MCA and HCPC analyses. The R packages tm [26] and topic models [27] were used to perform the pre-processing of the text and the topic modelling.

## 3. Results

Two hundred and two falls were considered in the analysis. Patients' median age was 77 years. Almost all falls occurred when the patients were hospitalized (95%) while they were alone (48%) in his hospital bedroom (70.1%) or in the toilet of his hospital bedroom (21.4%) during the afternoon (39.6%). About two-thirds of patients were at risk of fall (66.2%) according to the STRATIFY tool [28], but only in a small proportion of patients, a personalized fall prevention plan (27.5%) was developed (Table 1).

**Table 1.** Falls characteristics (2016–2017). Data are absolute numbers (percentage).

	N = 202
Setting: Inpatient	191 (95.0)
People witnessing the fall: None (the patient was alone)	97 (48.0)
The patient was able to provide information about the fall: Yes	152 (76.0)
The patient was at fall risk: Yes	129 (66.2)
Fall prevention plan	
Personalized fall prevention plan	53 (27.5)
Standard fall prevention plan	84 (43.5)
None	56 (29.0)
Previous falls (during the same hospitalization)	
No	183 (91.5)
Yes, once	13 (6.5)
Yes, twice	4 (2.0)
Time of the day	
Morning	52 (25.7)
Afternoon	80 (39.6)
Evening	35 (17.3)
Night	35 (17.3)
Location	
Outpatient clinic	8 (4.0)
Hospital corridor	6 (3.0)
Outside	3 (1.5)
Toilet	43 (21.4)
Hospital bedroom	141 (70.1)
Fall out of	
the stretchers	2 (1.0)
the wheelchair	6 (3.0)
the bed	36 (18.0)
the bed (raised bed rails)	23 (11.5)
chair	36 (18.0)
The patient was upright	97 (48.5)
Mechanism	
Stumbling	13 (6.5)
Losing consciousness	11 (5.5)
Off balance	58 (28.9)
Slipping	95 (47.3)
Not available	24 (11.9)
Action at time of fall	
Standing up/Sitting down	25 (12.8)
Lying on a stretcher	1 (0.5)
Walking	47 (24.1)
Sleeping	5 (2.6)
Reaching for something	14 (7.2)
Getting out from bed	1 (0.5)
Getting on bed	53 (27.2)
Using toilette	36 (18.5)
Not available	13 (6.7)
Extrinsic cause (patient)	
Clothes	10 (8.0)
Slippers (no shoe style)	48 (38.4)
Medical devices (e.g., surgical drains)	13 (10.4)
Walking Aids	8 (6.4)
Walking barefoot	46 (36.8)
Extrinsic cause (environment)	
Bed not adjustable	7 (9.0)
Poor lighting	8 (10.3)
No handrail in the toilet	4 (5.1)
Wet floor	9 (11.5)
Confined space	11 (14.1)
Lowered bed rails	22 (28.2)
Raised bed rails	17 (21.8)
Intrinsic causes (number)	
One	73 (50.3)
Two	49 (33.8)
Three	14 (9.7)
Four	9 (6.2)
Concomitant medications	
1–2 concomitant medications	44 (29.1)
3 concomitant medications	5 (3.3)
≥4 concomitant medications	102 (67.5)
Fall-related injury: No	119 (61.3)
If any, injury type	
Contusion	46 (51.7)
Wound	41 (46.1)
Fracture	2 (2.2)
Number of diagnostic tests	
None	141 (75.0)
One	39 (20.7)
Two	6 (3.2)
Three	2 (1.1)
Injury treatment	
Dressing	24 (13.0)
Suture	4 (2.2)
None	157 (84.9)

### 3.1. Pre-Processing Procedures and Text Mining

The pre-processing led to the identification of 212 vocabulary terms and 2310 tokens in the corpus. Using the cross-validation procedure, it was found out that ten topics were enough for a good sorting of documents, with a perplexity value of 443.5 (Table 2).

**Table 2.** Top terms of the topic analysis.

Top Terms
urine, parrot, stumble
pram, episode, dizziness
during, water, walk
stretcher, found, forward
erected, revive, damage
lift, bedside table, outspread
accidental, get up, bedside table
urin, tried, wear
bruise, take, corridor
slipper, stumble, perception

Such procedures led to the identification of a topic variable that was incorporated in the further analyses.

### 3.2. MCA and HCPC Analyses

The MCA was run both with and without the topic variable. In the MCA model without the topic variable, the first three dimensions explained the 19.72% of the overall data space, while with the topic variable the first three dimensions were found to explain the 17.4% of the overall data space.

For what concerns the results of the HCPC, three clusters were identified using only pre-coded information (Table 3).

**Table 3.** Results of the HCPC analysis without topic variable. V-test represents the influence of variables in the cluster composition. Cluster refers to the three distinct profiles of falls obtained from the HCPC model.

	V-Test	Cluster
Location: <i>Toilet</i>	12.03	1
Action at time of fall: <i>Using toilet</i>	10.47	1
People witnessing the fall: <i>None (the patient was alone)</i>	4.72	1
Fall out off: <i>the chair</i>	4.29	1
Extrinsic cause (patient): <i>Slippers (no shoe style)</i>	4.28	1
Location: <i>Hospital bedroom</i>	10.94	2
Action at time of fall: <i>Getting out from bed</i>	5.04	2
Fall out off: <i>the bed</i>	4.51	2
Fall out off: <i>the bed (raised bed rails)</i>	3.89	2
Extrinsic cause (environment): <i>Lowered bed rails</i>	3.79	2
Location: <i>Outpatient clinic</i>	6.83	3
Setting: <i>Outpatient</i>	6.27	3
Injury treatment: <i>Suture</i>	3.42	3
Fall prevention plan: <i>None</i>	3.4	3
The patient was at fall risk: <i>No</i>	2.99	3

The first one corresponded to elimination-related falls, accounting for 51 fall cases (27%). The second one grouped 139 (64%) falls occurred while the patient was in his hospital bedroom and was trying to get out from his hospital bed. The last cluster was the less represented (12, 8% of falls) corresponding to outpatients fell in the outpatient clinics without any fall risk with a resulting injury requiring a suture. The HCPC with the topic variable provided similar results (Table 4).

**Table 4.** Results of the HCPC analysis with topic variable. V-test represents the influence of variables in the cluster composition. Cluster refers to the three distinct profiles of falls obtained from HCPC model.

	V-Test	Cluster
Location: <i>Toilet</i>	10.92	1
Action at time of fall: <i>Using toilet</i>	9.15	1
Extrinsic cause (patient): <i>Slippers (no shoe style)</i>	4.66	1
People witnessing the fall: <i>None (the patient was alone)</i>	3.64	1
Mechanism: <i>Off balance</i>	3.51	1
Location: <i>Hospital bedroom</i>	9.23	2
Fall out off: <i>the bed</i>	5.03	2
Action at time of fall: <i>Getting out from bed</i>	4.88	2
Extrinsic cause (environment): <i>Lowered bed rails</i>	4	2
Fall out off: <i>the bed (raised bed rails)</i>	3.74	2
Location: <i>Outpatient clinic</i>	6.28	3
Setting: <i>Outpatient</i>	5.95	3
Injury treatment: <i>Suture</i>	3.39	3
Fall prevention plan: <i>None</i>	3.19	3
Fall out off: <i>the stretchers</i>	2.92	3

The first cluster identified elimination-related falls, accounting for 53 falls; 113 falls were grouped in the cluster 2, which included falls in the hospital bedroom while the patient was trying to get out from the hospital bed. Even in this case, the cluster 3 was the less represented, with 11 falls involving outpatients.

The fact that results provided by the HCPC with only pre-coded information and with pre-codes information together with the topic variable was confirmed by the analysis of the WSS, which was 9.22 for the HCPC without the topic variable and 8.71 with the addition of the topic variable to the HCPC.

#### 4. Discussion

The present study aimed to explore the value added by information reported in narratives in characterizing patterns of falls. Present findings showed that adding data from narratives did not improve the understanding of such phenomenon, even though the performance of the cluster analysis was good.

To our knowledge, no previous studies have assessed if the integration of information from narratives improves the understanding of falls in the hospital setting; however, similar studies have been conducted in other settings with promising results [12,29]. Results of the present study did not demonstrate the hypothesis that the integration of the information improves the understanding of in-hospital falls since the cluster analysis did not improve by adding information derived from narratives. Such findings suggest that a validated, well-conceived, data collection tool including only pre-coded information could be enough to characterize this phenomenon and guaranteeing its surveillance. Thus, avoiding the collection of narratives, which is time-consuming for healthcare professionals, and the application of sophisticated techniques to derive information from free text, which requires specific expertise in the field of text mining, not always readily available in the hospital services. In addition to that, the results of the MCA analysis showed that the proportion of variability explained by the pre-coded variables of the data collection tool was very low (less than 20%). This means that, probably, pre-coded information collected was not especially useful to differentiate falls. Such findings highlight the need for validating data collection tools before their use, avoiding the collection of redundant information that is not useful for the differentiation of falls patterns.

However, the present findings do not mean that free text information is not useful at all since we cannot rule out that such type of information would be useful in a different study setting using a different data collection form. We might hypothesize several reasons to understand the reason why in this specific study free-text information did not provide

any additional value to the characterization of falls patterns. The main one is that the type of data collected are redundant and the data collection form needs to be improved through a validation process. For this reason, we might hypothesize to develop ad hoc strategies to overcome such issues, including the use of validated data collection forms and the improvement of narratives quality by involving health care professionals, e.g., by providing a training in the use of the data collection form.

#### 4.1. Limitations

A limitation of this work is the small sample size. Text mining techniques implemented depend on whether the sample size is big enough to be representative of the heterogeneity of the corpus vocabulary. However, the small sample size is balanced by the nature of the narratives, which are usually written adopting a plain and standard vocabulary, without substantial variation in the choice of the words.

#### 4.2. Implications for Clinical Practice

From the clinical point of view, the characteristics of falls are consistent with those reported in the international literature [30–32]. They occurred most often in the elderly patients at fall risk in the hospital bedroom while they were trying to get out from bed (or getting on the bed) or they were using the toilet. The clustering analysis succeeded in the identification of three well-defined specific clusters of falls. The first two clusters identified two patterns of in-hospital falls that are already well-known in the literature [30–32]. Patients with elimination-related falls and those fell off the hospital bed identified respectively the first and the second clusters. Conversely, the third cluster is represented by outpatients fell in the outpatient clinic with a fall-related injury requiring a suture. There is already a substantial body of literature on falls occurring among community-dwelling individuals, particularly the elderly [33]. Even though a low number of events occurred in outpatients, such results further highlight the need for implementing risk management strategies also for outpatients attending the outpatient clinics. The risk management of the outpatients is much more difficult of that of the inpatients (often patients attend the outpatient clinic only once, not allowing for a comprehensive risk assessment). Considering such a framework, it could be useful to manage the risk of falls also in the outpatient setting [34,35].

## 5. Conclusions

Even though several efforts have been made to improve the risk management of in-hospital falls, they continue to represent a severe health care burden. The present findings showed that the cluster analysis is effective in characterizing fall patterns; however, they do not sustain the hypothesis that the analysis of free-text information improves our understanding of such phenomenon.

**Author Contributions:** Conceptualization, D.G.; Formal analysis, A.B.; Investigation, R.R.; Methodology, G.L.; Writing—original draft, G.L.; Writing—review and editing, P.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Review and approval were waived for this study, due to the retrospective nature of the study based on anonymized data.

**Informed Consent Statement:** Patient consent was waived since only anonymized data was collected.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. Morello, R.T.; Barker, A.L.; Watts, J.J.; Haines, T.; Zavarsek, S.S.; Hill, K.D.; Brand, C.; Sherrington, C.; Wolfe, R.; Bohensky, M.A. The Extra Resource Burden of In-Hospital Falls: A Cost of Falls Study. *Med. J. Aust.* **2015**, *203*, 367. [CrossRef]
2. Dunne, T.J.; Gaboury, I.; Ashe, M.C. Falls in Hospital Increase Length of Stay Regardless of Degree of Harm. *J. Eval. Clin. Pract.* **2014**, *20*, 396–400. [CrossRef]
3. Rubenstein, L.Z.; Josephson, K.R. The Epidemiology of Falls and Syncope. *Clin. Geriatr. Med.* **2002**, *18*, 141–158. [CrossRef]
4. Sillner, A.Y.; Holle, C.L.; Rudolph, J.L. The Overlap between Falls and Delirium in Hospitalized Older Adults: A Systematic Review. *Clin. Geriatr. Med.* **2019**, *35*, 221–236. [CrossRef]
5. Lan, X.; Li, H.; Wang, Z.; Chen, Y. Frailty as a Predictor of Future Falls in Hospitalized Patients: A Systematic Review and Meta-Analysis. *Geriatr. Nurs.* **2020**, *41*, 69–74. [CrossRef]
6. Ganz, D.A.; Huang, C.; Saliba, D.; Miake-Lye, I.M.; Hempel, S.; Ganz, D.A.; Ensrud, K.E. Preventing Falls in Hospitals: A Toolkit for Improving Quality of Care. *Ann. Intern. Med.* **2013**, *158*, 390–396.
7. Commission, J. Preventing Falls and Fall-Related Injuries in Health Care Facilities. *Jt. Comm.* **2015**, *55*, 1–55.
8. Hughes, R. *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*; Agency for Healthcare Research and Quality: Rockville, MD, USA, 2008; Volume 3.
9. Minsitero Della Salute. Dipartimento Della Programmazione e Dell'ordinamento Del Servizio Sanitario Nazionale (SSN) Direzione Generale Della Programmazione, Ex Ufficio III. Raccomandazione per La Prevenzione e La Gestione Della Caduta Del Paziente Nelle Strutture Sanitarie, Raccomandazione n. 13, Novembre 2011 (Aggiornata Al 1 Dicembre 2011). 2007. Available online: [https://www.salute.gov.it/imgs/C\\_17\\_pubblicazioni\\_1639\\_allegato.pdf](https://www.salute.gov.it/imgs/C_17_pubblicazioni_1639_allegato.pdf) (accessed on 6 April 2021).
10. Nguyen, L.; Bellucci, E.; Nguyen, L.T. Electronic Health Records Implementation: An Evaluation of Information System Impact and Contingency Factors. *Int. J. Med. Inform.* **2014**, *83*, 779–796. [CrossRef]
11. Lucero, R.J.; Lindberg, D.S.; Fehlberg, E.A.; Bjarnadottir, R.I.; Li, Y.; Cimiotti, J.P.; Crane, M.; Prosperi, M. A Data-Driven and Practice-Based Approach to Identify Risk Factors Associated with Hospital-Acquired Falls: Applying Manual and Semi- and Fully-Automated Methods. *Int. J. Med. Inform.* **2019**, *122*, 63–69. [CrossRef]
12. Vallmuur, K.; Marucci-Wellman, H.R.; Taylor, J.A.; Lehto, M.; Corns, H.L.; Smith, G.S. Harnessing Information from Injury Narratives in the 'Big Data' Era: Understanding and Applying Machine Learning for Injury Surveillance. *Inj. Prev.* **2016**, *22*, i34–i42. [CrossRef]
13. Jones, K.S. *Readings in Information Retrieval*; Morgan Kaufmann: San Francisco, CA, USA, 1997; ISBN 1-55860-454-5.
14. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
15. Griffiths, T.L.; Steyvers, M. Finding Scientific Topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. [CrossRef]
16. Leskovec, J.; Rajaraman, A.; Ullman, J.D. *Mining of Massive Datasets*; Cambridge University Press: Cambridge, UK, 2014; ISBN 1-107-07723-0.
17. Gilks, W.R.; Richardson, S.; Spiegelhalter, D. *Markov Chain Monte Carlo in Practice*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1995; ISBN 1-4822-1497-0.
18. Ueberla, J.P. More Efficient Clustering of N-Grams for Statistical Language Modeling. In Proceedings of the Fourth European Conference on Speech Communication and Technology, Madrid, Spain, 18–21 September 1995.
19. Greenacre, M.; Blasius, J. *Multiple Correspondence Analysis and Related Methods*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006; ISBN 1-4200-1131-6.
20. Husson, F.; Josse, J.; Pages, J. Principal Component Methods-Hierarchical Clustering-Partitional Clustering: Why Would We Need to Choose for Visualizing Data. *Appl. Math. Dep.* **2010**. Available online: [http://factominer.free.fr/more/HPCPC\\_husson\\_josse.pdf](http://factominer.free.fr/more/HPCPC_husson_josse.pdf) (accessed on 6 April 2021).
21. Husson, F.; Lê, S.; Pagès, J. *Exploratory Multivariate Analysis by Example Using R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017; ISBN 1-315-30187-3.
22. Celebi, M.E.; Kingravi, H.A.; Vela, P.A. A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm. *Expert Syst. Appl.* **2013**, *40*, 200–210. [CrossRef]
23. Team, R.C. R: *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
24. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [CrossRef]
25. Josse, J.; Husson, F. MissMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *J. Stat. Softw.* **2016**, *70*, 1–31. [CrossRef]
26. Feinerer, I.; Hornik, K. *Tm: Text Mining Package*, R package version 0.5-7.1; R Foundation for Statistical Computing: Vienna, Austria, 2012; Volume 1.
27. Hornik, K.; Grün, B. Topicmodels: An R Package for Fitting Topic Models. *J. Stat. Softw.* **2011**, *40*, 1–30.
28. Oliver, D.; Britton, M.; Seed, P.; Martin, F.C.; Hopper, A.H. A 6-Point Risk Score Predicted Which Elderly Patients Would Fall in Hospital. *BMJ* **1998**, *315*, 1049–1053. [CrossRef]
29. Altuncu, M.T.; Mayer, E.; Yaliraki, S.N.; Barahona, M. From Free Text to Clusters of Content in Health Records: An Unsupervised Graph Partitioning Approach. *Appl. Netw. Sci.* **2019**, *4*, 2. [CrossRef]
30. Hitcho, E.B.; Krauss, M.J.; Birge, S.; Claiborne Dunagan, W.; Fischer, I.; Johnson, S.; Nast, P.A.; Costantinou, E.; Fraser, V.J. Characteristics and Circumstances of Falls in a Hospital Setting. *J. Gen. Intern. Med.* **2004**, *19*, 732–739. [CrossRef]

31. López-Soto, P.J.; Smolensky, M.H.; Sackett-Lundeen, L.L.; De Giorgi, A.; Rodríguez-Borrego, M.A.; Manfredini, R.; Pelati, C.; Fabbian, F. Temporal Patterns of In-Hospital Falls of Elderly Patients. *Nurs. Res.* **2016**, *65*, 435–445. [[CrossRef](#)]
32. Aranda-Gallardo, M.; Morales-Asencio, J.M.; Canca-Sanchez, J.C.; Toribio-Montero, J.C. Circumstances and Causes of Falls by Patients at a Spanish Acute Care Hospital. *J. Eval. Clin. Pract.* **2014**, *20*, 631–637. [[CrossRef](#)]
33. Robertson, M.C.; Gillespie, L.D. Fall Prevention in Community-Dwelling Older Adults. *JAMA* **2013**, *309*, 1406–1407. [[CrossRef](#)]
34. De Gomes, G.A.O.; Cintra, F.A.; Batista, F.S.; Neri, A.L.; Guariento, M.E.; de Sousa, M.d.L.R.; D’Elboux, M.J. Elderly Outpatient Profile and Predictors of Falls. *Sao Paulo Med. J.* **2013**, *131*, 13–18. [[CrossRef](#)] [[PubMed](#)]
35. Kobayashi, K.; Ando, K.; Suzuki, Y.; Inagaki, Y.; Nagao, Y.; Ishiguro, N.; Imagama, S. Characteristics of Outpatient Falls That Occurred in Hospital. *Nagoya J. Med. Sci.* **2018**, *80*, 417.