Journal of Safety Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Safety Research

journal homepage: www.elsevier.com/locate/jsr

Finding statistically significant high accident counts in exploration of occupational accident data

Tuula Räsänen [a,*], Arto Reiman [b], Kai Puolamäki [c,d], Rafael Savvides [c], Emilia Oikarinen [c], Eero Lantto [a]

[a] Finnish Institute of Occupational Health, Finland
[b] Industrial Engineering and Management, University of Oulu, Finland
[c] Department of Computer Science, University of Helsinki, Finland
[d] Institute for Atmospheric and Earth System Research, University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 6 November 2020
Received in revised form 24 May 2021
Accepted 18 April 2022
Available online xxxx

Keywords:
Occupational accident
Silent signals
Workplace
Prevention

ABSTRACT

Introduction: Finnish companies are legally required to insure their employees against occupational accidents. Insurance companies are then required to submit information about occupational accidents to the Finnish Workers' Compensation Center (TVK), which then publishes occupational accident statistics in Finland together with Statistics Finland. Our objective is to detect *silent signals,* by which we mean patterns in the data such as increased occupational accident frequencies for which there is initially only weak evidence, making their detection challenging. Detecting such patterns as early as possible is important, since there is often a cost associated with both reacting and not reacting: not reacting when an increased accident frequency is noted may lead to further accidents that could have been prevented.
Method: In this work we use methods that allow us to detect silent signals in data sets and apply these methods in the analysis of real-world data sets related to important societal questions such as occupational accidents (using the national occupational accidents database).
Results: The traditional approach to determining whether an effect is random is statistical significance testing. Here we formulate the described exploration workflow of contingency tables into a principled statistical testing framework that allows the user to query the significance of high accident frequencies.
Conclusions: Our results show that we can use our iterative workflow to explore contingency tables and provide statistical guarantees for the observed frequencies.
Practical Applications: Our method is useful in finding useful information from contingency tables constructed from accident databases, with statistical guarantees, even when we do not have a clear a priori hypothesis to test.

1. Introduction

Before undertaking preventive or corrective occupational safety actions, risks of accidents must be identified through rigorous management of information (Ross et al., 2005). Accident statistics information has been analyzed as defining different characteristics of occupational accidents by, for example, Pietilä et al. (2018), Ciarapica and Giacchetta (2009), Hovden et al. (2010), Papazoglou et al. (2015), Cruz Rios et al. (2017), and Jacinto and Guedes Soares (2008). In this paper, we present a method to find unusually high accident counts, which allows iterative exploration of data and gives a statistical guarantee for the observed counts. We call these patterns *silent signals*, "silent" because they are easy to miss with more traditional approaches, and "signals" because they may be informative about emerging patterns or changes in the data.

The digital era in which we now live provides even more possibilities for complex data gathering and analysis (Badri et al., 2018). Technological developments have made it possible to collect and analyze different kinds of data from various sources using highly developed tools and methods. However, this development trend has not eliminated the role of humans as those who determine whether the data are actually useful for accident prevention purposes (Badri et al. 2018). The concept of 'big data' is used to describe this entity of handling and processing massive data sets. For instance, Wu and Li (2019) highlight the complexity of accident database analyses and suggest applying entropy theory to be able to more deeply understand the dynamic nature of occupational safety.

* Corresponding author.
E-mail address: tuula.rasanen@ttl.fi (T. Räsänen).

https://doi.org/10.1016/j.jsr.2022.04.003
0022-4375/© 2022 The Author(s). Published by National Safety Council Elsevier Ltd.

Finnish workplaces are legally required to insure their employees against occupational accidents. Insurance companies are then required to submit information about occupational accidents to the Finnish Workers' Compensation Center (TVK), which then publishes occupational accident statistics in Finland together with Statistics Finland.

Therefore, Finland's statistics and data have good coverage of blue-collar and white-collar employees' occupational accidents. In this paper we use a data set of occupational accidents in Finland from 2010–2015, available from TVK. Each accident is described by a date and 15 categorical variables (see Supplementary Material 1). Each variable consists of numerical codes that correspond to, for example, different industries, job types, accident causes, and injured body parts. We study the problem of finding accident counts that are larger than could be expected by random chance alone.

Accident frequencies can be displayed through contingency tables (or cross tabulation, pivot tables). We present a permutation testing-based statistical framework for exploring data through these tables. Contingency tables (such as Table 1) may be relatively large and contain multiple cells corresponding to accident frequencies, and multiple tables may be viewed through an iterative process. It is often the case that the analyst observes an unusually high accident frequency and does not know whether this is a significant finding or just a random effect. If the analyst can determine that it is a random effect, they can then focus on more promising hypotheses and avoid wasting resources on spurious findings or taking actions that are not based on the evidence at hand. For example, in Table 1 the two highest frequencies (405 and 149) seem to be a significant finding, while it is difficult to determine this for accidents with a lower absolute frequency (e.g., 37 or 4).

A traditional approach for determining whether an effect is random is statistical significance testing. Using, for example, a common statistical test, such as the chi-square test of independence (Agresti, 2019; Cacha, 1997), they can answer questions such as: "How unlikely is it to observe the counts in Table 1, if the variables are independent?" yielding a p-value of $\leq 10^{-16}$. The low p-value indicates that the cell values that were observed in Table 1 are extremely unlikely if the variables were independent, and thus there is evidence against independence.

However, these common statistical tests for contingency tables suffer from several shortcomings. First, they test a specific hypothesis and provide a single determination, or p-value, for the whole table. If the analyst is interested in a single accident frequency in the table, they are unable to obtain more focused answers. For example, after obtaining a low p-value using a chi-square test on Table 1, they know there is a significant finding, but cannot investigate which cells influenced this determination. If they attempt to naively test every cell in the table, they risk false discoveries due to the multiple comparisons problem (Dudoit et al., 2003). Second, most statistical tests have a specified null hypothesis that is formed before viewing the data. These are problems in practice. Answering questions such as 'what else is there in the data?' is not possible, because it would require formulating a new hypothesis that somehow takes into account what has already been observed, and then testing it on unseen data.

In practical data analyses, hypotheses are often formed after viewing the data during an iterative process, which is not in line with the assumptions made in traditional statistical testing, in which the hypothesis about the data should be formed before even observing the data at all. Therefore, there is a need for a statistical methodology that allows for testing hypotheses *during* the iterative workflow of viewing contingency tables. In this paper, we present such a methodology (initially introduced in Savvides et al., 2019)

Table 1
Contingency table of accident frequencies for Specific physical activity and Cause of accident variables. A p-value is included in parentheses (a dot signifies a p-value equal to one). *p*-values with $p \leq \alpha$=0.1 are statistically significant. The *p*-values have been corrected for multiple testing using the minP method (see text).

Cause of accident Specific physical activity	1100 ground level buildings/surfaces/structures	2699 other portable/mobile machines	2703 machines/chemical processes	2706 machines, other processes	2799 other fixed machines	2802 elevators/lifts/hoists/jacks etc.	2803 cranes/hoisting machines with suspended load	2811 non-lifting load transporting devices	2816 forklift trucks	2819 other handling mobile devices	2899 transport/storage systems not listed	4200 chemical/radioactive/biological substance
00 No information	4 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	11 (<0.01)
10 Operating machine	5 (.)	6 (0.032)	4 (0.51)	5 (<0.01)	7 (<0.01)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	1 (.)	12 (.)
20 Working with hand-held tools	8 (.)	1 (.)	4 (0.17)	1 (.)	2 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)	12 (0.96)
30 Driving/being on board a means of transport or handling equipment	11 (.)	3 (0.77)	1 (.)	1 (.)	0 (.)	0 (.)	1 (0.68)	0 (.)	7 (<0.01)	1 (1)	1 (.)	0 (.)
40 Handling of objects	37 (.)	7 (.)	12 (0.097)	2 (.)	8 (.)	1 (.)	0 (.)	3 (.)	1 (.)	1 (.)	0 (.)	149 (<0.01)
50 Carrying by hand	36 (.)	6 (0.14)	0 (.)	1 (.)	1 (.)	1 (.)	0 (.)	1 (.)	0 (.)	1 (.)	4 (0.045)	7 (.)
60 Movement	405 (<0.01)	2 (.)	1 (.)	3 (.)	5 (.)	3 (.)	0 (.)	2 (.)	5 (.)	2 (.)	3 (.)	16 (.)
70 Presence	6 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	2 (1)	0 (.)	0 (.)	23 (<0.01)
99 Other specific physical activities	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	14 (0.047)

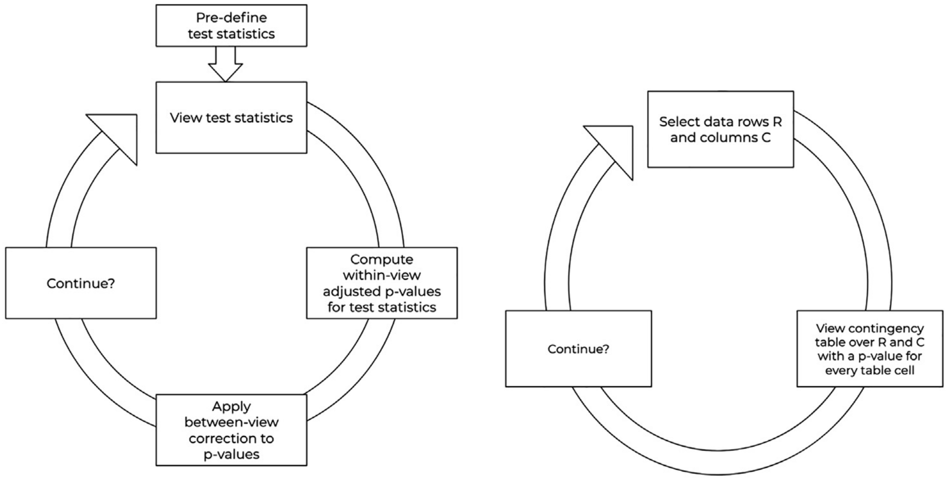

Fig. 1. Flowchart of the statistical testing procedure (left) and the practical workflow from the user's perspective (right).

and examples of finding novel features from a data set of occupational accidents in Finland.

The number of compensated accidents at work in Finland has been quite steady over the last 10 years. Registers show that over 126,000 occupational accidents occurred among wage earners in 2018 (Workers' Compensation Center, 2019). The vast majority (82%) of these accidents occurred at workplaces. Calculations by the Ministry of Social Affairs and Health show the annual costs of occupational accidents and injuries to be around EUR 2–2.5 billion in Finland (Rissanen & Kaseva, 2014). In addition, the human suffering of the injured person and their families and co-workers causes indirect costs that are difficult to estimate (Manuele, 2011).

Despite several measures to strengthen accident prevention and occupational safety, the statistics show a disparity between practical working life and the ambitious goal of zero accidents. Clearly, new approaches to improving occupational safety and accident prevention should be introduced. To contribute to this discussion, we introduce a new approach to large-scale occupational accident statistics categorization to achieve a more in-depth understanding of accidents for occupational accident prevention purposes.

Our objective is to detect *silent signals,* by which we mean patterns in the data such as increased occupational accident frequencies for which there may initially be only weak evidence, making their detection challenging. Detecting such patterns as early as possible is important, since there is often a cost associated with both reacting and not reacting: not reacting when an increased accident frequency is noted may lead to further accidents that could have been prevented. In this work we use methods that allow us to detect silent signals in data sets and apply these methods in the analysis of real-world data sets related to important societal questions such as occupational accidents (using the national occupational accidents database).

1.1. Motivating example

We next present an example motivating our approach. The example demonstrates how our approach works compared to a standard method. Using our approach, an analyst may ask more specific questions than standard methods.

Suppose that an analyst explores accident data using contingency tables. Table 1 displays one such contingency table (or cross tabulation, pivot table), in which each cell corresponds to an accident count in the chemical industries in Finland. If a cell appears to have a high frequency, the analyst may wish to know whether the high value is statistically significant. A standard approach, such as a chi-square test of independence, provides a *single* p-value for the whole table ($p \leq 10^{-16}$). The low p-value indicates that the table is statistically significant, and the analyst has made a "discovery." However, it is unclear which cell of the table is significant, which is especially problematic when the table is large.

In our approach, we determine whether a cell value is significantly high by computing a p-value for *every* cell in the table. The *p*-values are computed using a *permutation test*, in which the test statistic is the cell value and a p-value is computed by simulating the distribution of the test statistic under the null hypothesis. Our approach works as follows. We use a null hypothesis of independence between the variables of the table, and we simulate the null distribution by permuting each column in the data independently. This permutation scheme preserves the value distribution within each column and breaks any dependencies between columns. We permute the data multiple times and compute a contingency table on each permuted data set (Fig. 1b). This process provides a distribution of values for each cell in the table, which corresponds to the null distribution of each test statistic. A p-value can then be computed for each cell by comparing the simulated null distribution of the test statistic with its value in the original table. Finally, as we perform multiple tests, the *p*-values need to be adjusted for the multiple hypotheses problem. We adjust the *p*-values using a resampling-based adjustment procedure, called minP, which is discussed in the Methods section.

By computing a p-value for every cell, we can answer more specific questions. For example, the *p*-values in Table 1 communicate how likely it is to observe a count as high as that in the table when the 'Specific physical activity' and 'Cause of accident' variables are independent. In contrast, a standard approach, such as a chi-square test of independence, that provides one p-value for the whole table corresponds to the question: how likely is it to observe Table 1, when the 'Specific physical activity' and 'Cause of accident' variables are independent.

Another disadvantage of common statistical tests (besides not being able to test single cells), is that the analyst is unable to test more interesting hypotheses of independence. For example, how unlikely is it to observe Table 2, when the variables are *independent over most of the data, excluding a subset in which they are dependent*? In order to answer this question, we construct a permutation test, using a modified permutation scheme. Instead of permuting each column independently (as in Table 1), we now independently permute *tiles*. A tile is simply a subset of rows and columns (Fig. 1). It can act as a constraint on the permutation process, in that the rows in every tile are permuted independently to other tiles. In the pre-

Table 2
An alternative hypothesis is tested in Table 1. The accident frequencies are the same as in Table 1, while the *p*-values are computed differently. In addition to independence, we use *tile constraints* to fix the relationship of the variables in the subset of the data with Specific physical activity=(10. Operating machine). As a result, all cells with Specific physical activity = 10 are insignificant, whereas the other cells that were insignificant in Table 1 are now significant, e.g., (50. Carrying by hand, 2699. other portable/mobile machines) and (20. Working with hand-held tools, 2703. machines/chemical processes).

Cause of accident Specific physical activity	1100 ground level buildings/surfaces/ structures	2699 other portable/mobile machines	2703 machines/chemical processes	2706 machines, other processes	2799 other fixed machines	2802 elevators/ lifts/ hoists/jacks etc.	2803 cranes/ hoisting machines with suspended load	2811 non-lifting load transporting devices	2816 forklift trucks	2819 other handling mobile devices	2899 transport/ storage systems not listed	4200 chemical/ radioactive/ biological substance
00 No information	4 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	11 (0.015)
10 Operating machine	5 (.)	6 (.)	4 (.)	5 (.)	7 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	1 (.)	12 (.)
20 Working with hand-held tools	8 (.)	1 (.)	4 (0.098)	1 (.)	2 (0.98)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)	12 (0.92)
30 Driving/being on board a means of transport or handling equipment	11 (.)	3 (0.55)	1 (.)	1 (.)	0 (.)	0 (.)	1 (0.67)	0 (.)	7 (<0.01)	1 (1)	1 (.)	0 (.)
40 Handling of objects	37 (.)	7 (.)	12 (0.022)	2 (.)	8 (0.7)	1 (.)	0 (.)	3 (1)	1 (.)	1 (.)	0 (.)	149 (<0.01)
50 Carrying by hand	36 (.)	6 (0.042)	0 (.)	1 (.)	1 (.)	1 (.)	0 (.)	1 (.)	0 (.)	1 (.)	4 (0.049)	7 (.)
60 Movement	405 (<0.01)	2 (.)	1 (.)	3 (.)	5 (.)	3 (.)	0 (.)	2 (.)	5 (.)	2 (.)	3 (.)	16 (.)
70 Presence	6 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	2 (0.99)	0 (.)	0 (.)	23 (<0.01)
99 Other specific physical activities	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	14 (0.023)

vious example of Table 1, we permuted each column independently, which is equivalent to having a tile constraint over each column and then permuting each tile independently. In this second example of Table 2, we add a tile constraint over each column, as before, and we also apply another tile constraint over a subset of rows and all columns in which 'Specific physical activity'=10. The tile constraint causes the dependencies between the columns to be preserved in this subset of the data. As a result, every permuted data set has fixed the relationship of these variables in this subset of the data, which modifies the null distribution and the computed *p*-values. In practice, this means that the contingency table for every permuted data set is identical to the original data for 'Specific physical activity'=10, and the *p*-values are insignificant. Therefore, by using modified permutation schemes we can answer questions based on what was observed before, such as "what else is there in the data that is not explained by the observed accident counts?" This is not possible with standard methods.

In this paper, we present a permutation testing-based statistical framework for exploring data through contingency tables, based on the article published by Savvides et al. (2019). The framework includes two contributions: (1) application of a powerful statistical test that computes a p-value (adjusted for multiple testing) for every cell in a contingency table, and (2) a sequential exploratory procedure that is adjusted for multiple testing.

2. Material and methods

2.1. Data

The basic reporting of occupational accidents is done by companies. They investigate each accident and fill out an accident report form for the insurance company. When compiling statistics that present the conditions under which occupational accidents occur, the TVK uses ESAW variables (European Statistics on Accidents at Work). As described in the introduction, in this study we use a data set of occupational accidents in Finland from 2010–2016. In TVK's data, each accident is described by a date and 15 categorical variables (see Supplementary Material 1). In addition to the 7 ESAW variables, TVK use their own more specific variables. Each variable consists of numerical codes that correspond to different industries, job types, accident causes, and injured body parts.

2.2. Methods

2.2.1. Overview

We study the accident data by calculating contingency tables. For example, an expert chooses the 'industry' and 'body part' variables and views a table that contains accident frequencies for every combination of industry and body part.

In this exploratory workflow, the user may observe unusually high or low accident counts, which may be true phenomena in the data or merely random artefacts. One traditional method for discarding findings that cannot be distinguished from random noise, is statistical significance testing. Hence, here we formulate the described exploration workflow of contingency tables into a principled statistical testing framework that allows the user to query the significance of high accident frequencies.

We follow an approach presented in our previous paper Savvides et al. (2019), which uses a permutation test. In this article, the authors provide a novel realization of the method for contingency tables and a new iterative correction method based on alpha investing. The test requires a test statistic and its null distribution. If the test statistic computed on the observed data is extreme compared to its null distribution, then it is significant. In our example, the test statistic corresponds to an accident frequency (a value of the cell in the contingency table) and the null distribution is defined as a model of the user's knowledge of the data as defined in Puolamäki et al. (2021). The user's knowledge is parameterized as a probability distribution over all possible data sets, and samples are drawn from this distribution to form the empirical null distribution of the test statistic.

Our null hypothesis assumes that the marginal distributions of the variables are fixed and that all possible data sets can be obtained by permuting the columns of the data sets. A sample from the null hypothesis can be obtained by computing the contingency table for such a permuted data set. Without any constraints, the columns are permuted independently and at random, which results in a null hypothesis corresponding to situations in which the data attributes are independent of each other and any relation between them is broken. During the exploration, the user's knowledge is updated, using observed contingency tables as constraints: when a pattern is observed, the permutations are constrained so that the attributes shown for the user in the permutation table are permuted together, after which all samples produce the observed contingency table. We can informally say that a test statistic is significant if it is exceptionally high compared to the user's expectations (i.e., if the test statistic has a low *p*-value).

Two aspects in the exploration process require attention. Firstly, multiple test statistics are often viewed and hence tested simultaneously. A multiple testing correction is required in order to avoid false discoveries. Secondly, we assume that the user views the data more than once (i.e., the exploration is an iterative process). If the user looks at the data enough times, they will eventually discover something significant by chance alone. This adds another level of multiple testing, which also requires a correction. We next formally describe our procedure that incorporates these two levels of multiple testing corrections to control the family-wise error rate (FWER) at a chosen level.

2.2.2. Details

In this section, we formally describe the testing procedure, as initially described in our previous paper Savvides et al. (2019). The novel contributions in this paper are the application to the domain of accident data using contingency tables, two theorems, and the use of alpha investing as an iterative correction.

Let Ω denote the *sample space* (i.e., the set of all possible data sets), and $\omega_0 \in \Omega$ the observed data set, which has been sampled from an unknown probability distribution Pr_D over Ω. As discussed above, we implicitly assume the user's knowledge is parametrized as a probability distribution of Pr_U over Ω.

Our goal is to formulate a statistical testing procedure in which Pr_U is the null distribution and the test statistic corresponds to a pattern observed in the data. Intuitively, we call the pattern significant if the test statistics (counts in contingency table) are extreme compared to Pr_U.

Test statistics. We define test statistics as functions $T_i : \Omega \to \mathcal{R}, i \in [n_T]$, which measure the 'strength' of an observed pattern and where we have used the notation $[n_T] = \{1, \ldots, n_T\}$. In this paper, the observed patterns are the counts in a contingency table.

Iterative exploration. We assume that the user is shown a finite *sequence of n_V views* of the data. Each view V_t with $t \in [n_V]$ contains a subset of counts T_i shown in one contingency table and is defined as an index set, i.e., $V_t \subseteq [n_T]$. The idea is that in view V_t, the user observes the values of the test statistics on the observed data $T_j(\omega_0)$ for all $j \in V_t$.

Null distribution. The user's knowledge Pr_U can be updated with the use of *constraints* $C_i : \Omega \to P(\Omega)$ (where $P(\Omega)$ denotes the power set of Ω), which restrict the possible data sets to those that have a test statistic equal to the observed data set, i.e., $C_i(\Omega) = \{\omega \in \Omega : T_i(\omega) = T_i(\omega_0)\}$. We identify a set of constraints using an index set $I \subseteq [n_T]$ and we denote the set of possible data sets that satisfy a set of constraints $I \subseteq [n_T]$ as $\Omega_I = \cap_{i \in I} C_i(\Omega) = \{\omega \in \Omega : T_j(\omega) = T_j(\omega_0) \forall j \in I\}$.

In each view V_t, the null distribution is the user's *current* knowledge Pr_U, which has been updated on the basis of the test statistics I_t observed so far. We define *constrained p-values* as:

$$p_{i|I} = \frac{Pr_U(\{\omega \in \Omega_I : T_i(\omega_0) \leq T_i(\omega)\})}{Pr_U(\Omega_I)}$$

The null hypothesis that corresponds to a constrained p-value $p_{i|I}$ is that the distribution Pr_D (from which the observed data ω_0 is sampled) satisfies the following condition for any $\omega \in \Omega_I$:

$$\frac{Pr_D(\omega)}{Pr_D(\Omega_I)} = \frac{Pr_U(\omega)}{Pr_U(\Omega_I)} \tag{1}$$

The intuitive interpretation for the null hypothesis is that if true, then the conditional distribution of the data, given the constraints, is equal to the corresponding distribution assumed by the user.

Within-views correction. A view contains multiple test statistics, which are used simultaneously for testing the null hypothesis (user's knowledge). Since multiple tests are performed, a multiple testing correction is warranted. We use the *step-down minP procedure* (Westfall-Young, 1993) to compute FWER-adjusted *p*-values for the test statistics in a single view.

The minP algorithm is summarized as follows: given a vector of observed test statistics $X_0 = (x_1, \cdots, x_n)$ and a matrix of m samples of test statistic vectors from the null distribution $Y = (X_1, \cdots, X_m)$, the minP algorithm computes a vector of FWER-adjusted *p*-values $P = (p_1, \cdots, p_n)$.

An implementation of the minP algorithm in the R programming language (R Core Team, 2020) is provided in the Supplementary Material 2.

Between-views correction. The user is shown multiple views in a sequential manner and in each view, a hypothesis is tested. In addition to the within-views correction, an additional multiple testing correction is warranted for the sequence of views. If the number of views is known in advance, we can apply any multiple testing correction, such as a Bonferroni correction. However, if the number of views is not known in advance, we instead apply an online multiple testing correction, such as alpha-investing (Foster & Stine, 2008). In alpha investing, the user has an alpha wealth of total acceptable error that they may "invest" in hypotheses. If the hypothesis provides a significant result, then the alpha investment is returned and can be reused in future hypotheses.

A simple online multiple testing procedure is a generalization of the Bonferroni, called a *weighted Bonferroni correction* (Holland & Copenhaver, 1988). The weighted Bonferroni correction is summarized as follows: given a sequence of *p*-values p_t $t \in [n_V]$, multiply each p-value with a factor w_t such that $\sum_{t=1}^{\infty} 1/w_t = 1$. Then the p-values $\underline{p_t} = \min(1,\ w_t\ p_t)$ are adjusted for FWER.

2.3. Testing procedure

The elements described above (test statistic, null distribution, iterative exploration, within-view correction and between-view correction) are combined into a statistical testing procedure. The testing procedure consists of the following steps, for a given data set $\omega_0 \in \Omega$ sampled from Pr_D, number of views n_V and weights w_t for each view $V_t, t \in [n_V]$:

1. Set $t \leftarrow 1, V_0 \leftarrow \{\}, I_0 \leftarrow \{\}$
2. Set $I_t \leftarrow I_{t-1} \cup V_{t-1}$

3. View values of test statistics $T_j(\omega_0)$ where $j \in V_t \subseteq [n_T]\{I_t$
4. Compute within-view adjusted p-values p_{tj} using minP algorithm and apply between-view correction to obtain final adjusted p-values $\underline{p_{tj}} = \min\left(1,\ w_t\ p_{tj}\right)$
5. Set $t \leftarrow t + 1$
6. If $t \leq n_V$ continue from Step 2, else terminate

The flowcharts below (Fig. 2)) visualize the statistical procedure (left) and how it translates to a workflow for the user (right). The statistical procedure is generally applicable to data exploration with any visualization, while the workflow presented here is a special case where the visualization is a contingency table. The workflow is presented in the section Case Example.

The procedure is an iterative process, as denoted by the "Continue?" block in the flowchart. The process terminates either when the user wishes to end the exploration, or when the exploration has no practical reason to continue, for example if the data are fully constrained (i.e., everything has been observed already), or the specified alpha budget is depleted when using an alpha investing method. In practice, the lack of a termination criterion means that the user is free to explore as long as there are discoveries to be made in the data and there is enough available alpha budget.

The following theorems show that the above procedure controls the family-wise error rate (FWER) at a chosen level α, both within each view and overall for the whole procedure. The theorems are a novel contribution that extends our previous work (Savvides et al., 2019).

Theorem 1 (within-view)

Let V_t be a contingency table containing test statistics T_j where $j \in V_t$, and p_j are the corresponding p-values as computed with the minP algorithm using Pr_U as a null distribution.

Then for any given constant $\alpha \in [0, 1]$ and for every $j \in V_t$ we have that $Pr(p_j \leq \alpha) \leq \alpha$, i.e., the probability of at least one false discovery is at most α.

Proof. Assume we have m data samples from Pr_U, denoted by ω_i. Let $X_0 = \left[T_j(\omega_0)\right], j \in V_t$ be a vector of test statistics for the observed data set, $X_i = \left[T_j(\omega_i)\right], j \in V_t$ be a vector of test statistics for data sample ω_i, and $Y = [X_1, \cdots, X_m]$ be a matrix of m test statistic vectors.

Then the p-values $p_i = MINP(X_0, Y)$ are FWER-adjusted, since the minP algorithm controls FWER.

Theorem 2 (between-views)

Let $S = (V_1, \cdots, V_{n_V})$ be a sequence of views and p_{tj} the p-values in each view, as computed with the minP algorithm using Pr_U as a null distribution and corrected with the weighted Bonferroni correction.

Then for any given constant $\alpha \in [0, 1]$, for every $t \in [n_V]$ and every $j \in V_t$ we have that $Pr(p_{tj} \leq \alpha) \leq \alpha$.

Proof. We use $W_t \subseteq V_t$ to denote the views whose p-values obey the null hypothesis, according to the definition of Eq. (1), and by $S = (W_1, \cdots, W_{n_V})$ the respective sequence of views. We denote by $P_t = \min_{j \in W_t} p_{jt}$, with $P_t = 1$ if $W_t = n$ull, the minimal p-value in view $t \in [n_V]$ in which the null hypothesis is true. Since the p-values in each view have been corrected for *FWER*, we know that $P(P_t \leq \alpha') \leq \alpha' \forall \alpha'$.

Consider an iteration $t \in [n_V]$. The user has the option of choosing any subset of test statistics $j \in [n_T]$ to V_t. It can be shown that for all test statistics, including P_t, for which the null hypothesis is true, it holds that they are stochastically no larger than the uniform distribution. Then, P_t is multiplied by w_t, which means that the probability of a false positive at iteration $t \in [n_V]$ is therefore at most ${w_t}^{-1}\alpha$, resulting in a total false positive probability of at most α when summed over all iterations in $[n_V]$, since $\sum_{t=1}^{n_V} w_t^{-1} \leqslant 1$.

3. Case example

In this section, we demonstrate the statistical testing procedure using case studies and discuss their results. As a first case study, we focus on occupational accident data from the chemical product industry in Finland in 2010–2015. The idea is to explore the accident data using the testing procedure in order to obtain insights into unusually high accident frequencies in the chemical product industry. Focusing the analysis on one industry enables the selection of variable categories that are relevant to that industry, for example, standard variables in accident reports. Selecting only a subset of categories reduces the number of multiple hypotheses and hence improves statistical power.

Note that alternative approaches to find similar results are limited or are not typically used, to our knowledge. For example, using a standard test, such as a chi-square test of independence, we can compute a p-value for each table (Cacha, 1997). However, the test provides a *single* p-value for the whole table (as opposed to one for each table cell) and the p-value does not account for previously observed significant patterns (whereas here the user's knowledge is updated, which affects future p-values). Therefore, traditional methods are not directly comparable to our presented framework, as discussed in the Motivating example. Two general methods that can act as baselines are an approach where no corrections are performed, and an approach where the corrections are overly strict. The first approach may lead to spurious findings and no control of the error rate, which our method controls. The second approach may lead to no findings due to lack of statistical power, while our

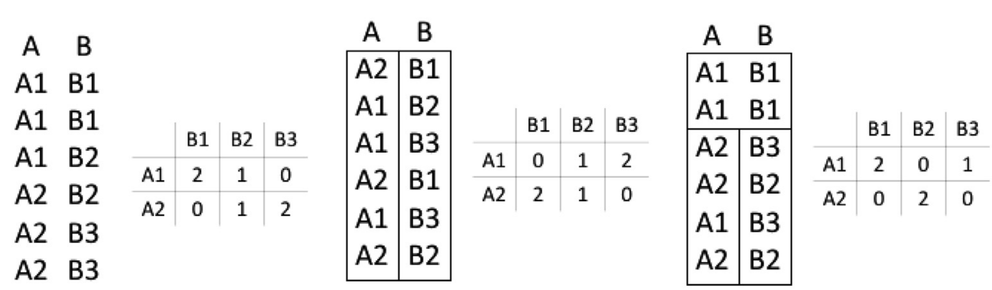

Fig. 2. Illustration of permutation with *tiles* and its effect on contingency tables. *Left*: Data set D with variables A and B. A contingency table is computed from D by counting all combinations of values in variables A and B. *Centre*: Variables A and B are permuted independently. The permutation is realized through tile constraints. A tile is placed over each column and each tile is permuted independently. A contingency table is computed in the same manner. *Right*: Variables A and B are permuted independently, except for a subset of rows where B = B1. Tile constraints are placed in each column, as before, and an additional tile is placed on a subset of rows where B = B1. The result is that variables A and B retain their relationship within the subset where B = B1. This is illustrated in the contingency table in which the counts containing B1 are the same as those in the original data.

approach retains statistical power through the powerful minP correction. In addition, the correctness of the results of the case study cannot be demonstrated experimentally, since there is no "ground truth" to compare to. The validity of the approach is provided by the mathematical proofs in the methods section.

We now describe a case example of using the testing procedure to explore a data set of occupational accidents. The exploration consists of three iterations (i.e., three contingency tables are viewed sequentially). In each iteration, the contingency table is determined by selecting two variables (columns) and a subset of data points (rows). For each cell in the table, a FWER-adjusted p-value is computed using as a within-view correction the minP algorithm and as a between-view correction a standard Bonferroni procedure for a predetermined number of three iterations. After viewing a table, the observed accident frequencies are used to update the user's background distribution and are therefore not significant in future tables.

Iteration 1. We start by viewing a contingency table of the Specific physical activity and Cause of accident variables for the whole data. For the Cause of accident variable we view 12 out of 73 categories that are relevant to the chemical product industry (e.g., chemicals, logistics and machinery). For the Specific physical activity variable we view all nine categories (e.g., using machinery or handling objects).

We discover eight statistically significant accident frequencies in Table 3. These frequencies are unusually high compared to the current knowledge of the user, as parameterized by the null distribution.

After viewing Table 3, the user's knowledge is updated so that if Table 3 is viewed in future iterations, it does not contain significant findings. The user's knowledge is updated by modifying the null distribution through a tile constraint {R = all rows, C = (Specific physical activity, Cause of accident)} that fixes the relationship of the variables in Table 3.

Iteration 2. The table for the next iteration is determined by the user, by selecting a subset of rows and the two variables of the contingency table. The next table can be completely independent from the current one or (as in this example) it can be based on the findings of previous tables. We now focus on a subset of the data that was significant in Table 3, denoted by R1 = {Specific physical activity = 60 Movement, and Cause of accident = 1100 ground level buildings/surfaces/structures}. In subset R1, we view a contingency table of the Industry (4 digit) (using 18 out of 587 categories which, based on our knowledge, are relevant for the chemical product industry) and Working process (using all 32 categories) variables. The contingency table is presented in Table 4 (see Supplementary Material 3 for full table) and we discover one statistically significant result.

The effect of Iteration 1 on Iteration 2 has two parts. First, subset R1 was selected on the basis of the findings from Iteration 1. Second, the constraints from Iteration 1 on the null distribution may affect the *p*-values of Iteration 2. In this case, the constraints have no overlap with the data of Table 4, and as such have no effect on the *p*-values.

After viewing Table 4, the user's knowledge is updated, similarly to Iteration 1. The null distribution is updated by adding a tile constraint {R = R1, C = (Industry (4 digit), Working process)} that fixes the relationship of the variables in Table 4 for the viewed subset R1 (i.e., not for all the data). After fixing this result for subset R1, we can now test whether the result is significant for the rest of the data. We do this by using the whole data set (instead of only subset R1) to view the same variables as in Table 4.

Iteration 3. A significant result is discovered in Iteration 2, for subset R1 of the data. We now repeat the steps of Iteration 2 using all the accident reports in the chemical product industry, to investigate whether accident frequency is also significantly high for the

Table 3
Contingency table of Cause of accident and Specific physical activity variables. Each cell contains the accident count for a category of each variable. A FWER-adjusted *p*-value is computed for each cell and is contained inside parentheses. Insignificant p-values ($p = 1$) are denoted by a dot. *p*-values with $p \le \alpha = 0.1$ are statistically significant. The cell in bold is the subset R1 used in Table 4.

Cause of accident / Specific physical activity	1100 ground level buildings/surfaces/structures	2699 other portable/mobile machines	2703 machines/chemical processes	2706 machines, other processes	2799 other fixed machines	2802 elevators/lifts/hoists/jacks etc.	2803 cranes/hoisting machines with suspended load	2811 non-lifting load transporting devices	2816 forklift trucks	2819 other handling mobile devices	2899 transport/storage systems not listed	4200 chemical/radioactive/biological substance
00 No information	4 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	11 (0.022)
10 Operating machine	5 (.)	6 (0.097)	4 (.)	5 (0.022)	7 (0.022)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	1 (.)	12 (.)
20 Working with hand-held tools	8 (.)	1 (.)	4 (0.5)	1 (.)	2 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)	12 (.)
30 Driving/being on board a means of transport or handling equipment	11 (.)	3 (.)	1 (.)	1 (.)	0 (.)	0 (.)	1 (.)	0 (.)	7 (0.022)	1 (.)	1 (.)	0 (.)
40 Handling of objects	37 (.)	7 (.)	12 (0.29)	2 (.)	8 (.)	1 (.)	0 (.)	3 (.)	1 (.)	1 (.)	0 (.)	149 (0.022)
50 Carrying by hand	36 (.)	6 (0.41)	0 (.)	1 (.)	1 (.)	1 (.)	0 (.)	1 (.)	0 (.)	1 (.)	4 (0.13)	7 (.)
60 Movement	**405 (0.022)**	2 (.)	1 (.)	3 (.)	5 (.)	3 (.)	0 (.)	2 (.)	5 (.)	2 (.)	3 (.)	16 (.)
70 Presence	6 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	2 (.)	0 (.)	0 (.)	23 (0.022)
99 Other specific physical activities	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	14 (0.14)

Table 4

Contingency table of Working process and Industry (4 digit) variables for the subset of the data defined by R1 (significant result from Table 3). Only a part of the table is shown here for clarity; refer to the Supplementary Material 3 for the whole table. Each cell contains the accident count for a category of each variable. A FWER-adjusted p-value is computed for each cell and is contained inside parentheses. Insignificant p-values ($p = 1$) are denoted by a dot. p-values with $p \leq \alpha = 0.1$ are statistically significant.

Working process Industry 4 digit	00 no information	11 production, manufacturing, processing	12 storing	19 other manufacturing and storing	21 excavation	22 new construction, building	23 new construction, roads, bridges, dams, ports	24 remodelling, repairing, building maintenance	25 demolition	29 other construction
2011 Manufacture of industrial gasses	2 (.)	4 (.)	2 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2012 Manufacture of colours and pigments	0 (.)	2 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2013 Manufacture of other non-organic basic chemicals	6 (.)	27 (.)	1 (.)	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2014 Manufacture of other organic basic chemicals	1 (.)	4 (.)	0 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2015 Manufacture of fertilizers and nitrogen compounds	0 (.)	3 (.)	2 (.)	0 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)
2016 Manufacture of plastic materials	1 (.)	29 (0.05)	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2017 Manufacture of synthetic rubber raw material	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2020 Manufacture of pesticides and agriculture chemicals	1 (.)	6 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2030 Manufacture of paints, printing inks and enamels	0 (.)	17 (.)	9 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)

Table 5

Contingency table of Working process and Industry (4 digit) variables for the whole data set. Only a part of the table is shown here for clarity; refer to the Supplementary Material 3 for the whole table, which contains more statistically significant cells. Each cell contains the accident count for a category of each variable. A FWER-adjusted p-value is computed for each cell and is contained inside parentheses. Insignificant p-values ($p = 1$) are denoted by a dot. p-values with $p \leq \alpha = 0.1$ are statistically significant.

Working process Industry 4 digit	00 no information	11 production, manufacturing, processing	12 storage	19 other manufacturing and storage	21 excavation	22 new construction, building	23 new construction, roads, bridges, dams, ports	24 remodelling, repairing, building maintenance	25 demolition	29 other construction
2011 Manufacture of industrial gasses	12 (.)	47 (.)	21 (.)	3 (.)	1 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)
2012 Manufacture of colours and pigments	1 (.)	27 (.)	2 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2013 Manufacture of other non-organic basic chemicals	17 (.)	168 (.)	20 (.)	15 (.)	1 (.)	1 (.)	0 (.)	0 (.)	1 (.)	0 (.)
2014 Manufacture of other organic basic chemicals	5 (.)	53 (.)	8 (.)	10 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2015 Manufacture of fertilizers and nitrogen compounds	4 (.)	32 (.)	7 (.)	3 (.)	0 (.)	0 (.)	0 (.)	1 (.)	0 (.)	0 (.)
2016 Manufacture of plastic materials	6 (.)	123 (.)	13 (.)	14 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2017 Manufacture of synthetic rubber raw material	2 (.)	21 (.)	2 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2020 Manufacture of pesticides and agriculture chemicals	2 (.)	25 (.)	1 (.)	1 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
2030 Manufacture of paints, printing inks and enamels	10 (.)	209 (0.083)	50 (.)	10 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)

whole data. In Table 5, we discover seven significant results (see Supplementary Material 3). However, these do not include the significant result from Iteration 2. In other words, we observe a significantly high accident frequency in subset R1 of the data and, after taking it into account in the null distribution, the same accident frequency is not significantly high in the whole data set, even though more data are used. This suggests that the significantly high accident frequency is somehow related to the constraints added during the exploration: [{R = all, C=(Specific physical activity, Cause of accident)} + {R = R1, C=(Industry (4 digit), Working process)}]. To further illustrate this relationship, we contrast the above with a scenario in which there are no tile constraints from previous iterations (i.e., when the user has not viewed the previous two tables). In this scenario, the significantly high count in a subset (from Iteration 2) is also significant for the whole data set.

The findings obtained from the above three iteration steps are products of an exploratory data analysis. The benefit over existing approaches is that the analyst is allowed to look at the data and still be able to obtain a statistical guarantee that the observed accident counts are not due to random chance alone.

For an occupational safety analyst, the results of these three iterations in this case study propose that in the manufacture of plastic materials there may have been additional haste in production, leading to relatively many slip, trip and fall-related injuries. Now having this statistical guarantee, the analyst could start looking at other data from the industry (such as production volumes) that could explain the result according to their hypotheses. Finding larger than expected accident counts is an ubiquitous problem across safety research, for which our approach provides a practical solution.

4. Discussion

Responsive methods for accident statistics analyses have traditionally been used in safety management (Goel et al., 2017) and a selection of predictive methods have been introduced to supplement these. Predictive models developed in recent years are able to predict, for example, the number and severity of accidents at work, but silent signals that can anticipate safety situations are still poorly recognized by the commonly used analysis methods. We see that more attention should be paid to identifying silent signals and modern analytics tools in order to succeed in accident prevention. By identifying information sources that anticipate critical safety incidents and utilizing data mining, data collection and analysis can focus on relevant issues and be more cost-effective.

In practical working life, occupational accidents are often approached through uni- and bi-variate distribution analyses that show the distribution of incident characteristics in absolute numbers or percentages. In more sophisticated use, incident concentration analyses try to identify clusters of incidents with common characteristics utilizing variables similar to ours to prioritize safety measures (Kjellén and Albrechtson, 2017). Our analysis approach utilizes similar data, with the purpose of identifying silent signals from the data set of occupational accidents in Finland.

The 'traditional' way to conduct a scientific study on accident statistics data has been to form a hypothesis and then use statistical testing methods to see if the hypothesis is true (e.g., some frequencies are high). The methods presented in this article enable us to draw more fine-tuned conclusions and also perform the analysis iteratively, as the approach we present allows creating hypotheses during the analysis based on viewing contingency tables created from the data. This method would be useful for detecting 'silent signals' for informed decision making, for example, even if they concern only small portions of the data (e.g., one branch, city, company).

Previous accident analysis models suffer from the fact that it is not always obvious if the found patterns are valid in a statistical sense. The methodology presented in this paper provides a straightforward, understandable, yet powerful framework to find hidden signals and weed out random artefacts. In the examples of this paper we used raw data sets provided by TVK and only had the human expert's knowledge and intuition at hand. In principle it would have been possible to use other variables in this context (such as a person's income level, health status, etc.). However, this would have required combining different databases. As an example, Pietilä et al. (2018) similarly combined two different databases; an accident statistics database of one accident insurance company and an employee health database of an occupational health care provider.

New approaches to data analysis are needed when human capacity is not sufficient to analyze available data efficiently and reliably. Occupational safety management is facing such a challenge when it comes to utilizing fragmented information as well as large materials; this creates its own challenges for information management. In information management, information can be divided into explicit and indirect information. The collection and use of this indirect or tacit information can be of significant benefit in the prevention of accidents at work (Podgorski, 2010). Data-driven safety management, which takes advantage of more than just accident data, enables continuous improvement (Wang et al., 2018). This is what many employers strive for, as reducing accident rates with traditional analytics and data is limited.

In principle, it would be possible to use an AI method, for example, to suggest views of the data and our method to independently assess the statistical validity of the results, or augment the data set by attributes (e.g., risk indices) estimated by supervised learning models. In this article, the focus was on the 15 variables used in the TVK data. However, the TVK data also contained small verbal descriptions of every accident. This part of the data we excluded, as our focus was on statistical testing. Combining these two parts of data would be an inspiring new approach for a future study on this topic. As we have learned from the studies by, for instance, Jocelyn et al. (2016), Nanda et al. (2016), Valmuur et al. (2016) and Marucci-Wellman et al. (2017), machine learning has been successfully tested in analyzing accident descriptions. We believe that such an approach, going into the verbal data in depth, should be studied further.

5. Conclusions

The presented method is generic and can, in principle, be used to explore any data set from which one can compute contingency tables and for which contingency tables are an informative 'visualization.' Even though the examples in this paper are from quantitative measurements, there is no reason why the same approach could not be applied to qualitative data, from questionnaires, for instance. Furthermore, machine learning algorithms such as classifiers are often used to find relations in the data and to estimate unobserved variables. For example, in the case of this work accident data set we could try to estimate some of the properties using a classifier. The prediction given by a classifier could be added as a new variable to the data.

Large data sets contain a great deal of potentially useful and valuable information. Often, there is no one great question that is clear in advance; finding the useful parts requires first exploring the data. After we see something, it is then important to have some confidence in the fact that the observed patterns – in our case accident frequencies – are 'real' and not just random artefacts.

In this paper, we have proposed a method to do this on a publicly available occupational accident database. Our approach is based on iterative exploration of confidence tables. Although the underlying mathematics and algorithms require some understanding, the outcomes are easily understandable, namely contingency tables and knowledge, if any of the contingency table elements are larger than they would be expected to be by chance.

Future studies could focus on studying combined material in larger samples as they introduce an interesting possibility to gain more in-depth information. Analysis of large-scale data sets with richer information about the employer, workplace and organizational practices could provide more insight into their effects on occupational accidents.

Acknowledgements

We thank Janne Sysi-Aho and TVK for their help and insightful discussions.

This study was supported by the Academy of Finland (decisions 326280 and 326339).

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsr.2022.04.003.

References

Agresti, A. (2019). An introduction to categorical data analysis. *Wiley series in probability and statistics*. Hoboken, NJ: John Wiley & Sons.

Badri, A., Boudreau-Trudel, B., & Souissi, A. S. (2018). Occupational health and safety in the industry 4.0 era: A cause for major concern? *Safety Science, 109*, 403–411.

Cacha, C. A. (1997). *Research design and statistics for the safety and health professional*. New York, NY: Van Nostrand Reinhold.

Ciarapica, F. E., & Giacchetta, G. (2009). Classification and prediction of occupational injury risk using soft computing techniques: An Italian study. *Safety Science, 47*(1), 36–49.

Cruz Rios, F., Chong, W. K., & Grau, D. (2017). The need for detailed gender-specific occupational safety analysis. *Journal of Safety Research, 62*, 53–62.

Dudoit, S., Shaffer, J. P., & Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science, 18*(1), 71–103.

Foster, D. P., & Stine, R. A. (2008). α-investing: A procedure for sequential control of expected false discoveries. *J Royal Statistical Soc B., 70*(2), 429–444.

Goel, P., Datta, A., & Mannan, S. (2017). Application of big data analytics in process safety and risk management. *2017 IEEE international conference on big data*.

Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. *Psychological Bulletin, 104*(1), 145–149.

Hovden, J., Albrechtsen, E., & Herrera, I. A. (2010). Is there a need for new theories, models and approaches to occupational accident prevention? *Safety Science, 48*(8), 950–956.

Jacinto, C., & Guedes Soares, C. (2008). The added value of the new ESAW/Eurostat variables in accident analysis in the mining and quarrying industry. *Journal of Safety Research, 39*, 631–644.

Jocelyn, S., Chinniah, Y., & Ouali, M.-S. (2016). Contribution of dynamic experience feedback to the quantitative estimation of risks for preventing accidents: A proposed methodology for machinery safety. *Safety Science, 88*, 64–75.

Kjellén, U., & Albrechtson, E. (2017). *Prevention of accidents and unwanted occurrences: Theory, methods, and tools in safety management* (Second Edition). Boca Raton, FL: CRC Press.

Manuele, F.A. (2011). Accident costs. Rethinking ratios of indirect to direct costs. Professional Safety, 39-47.

Marucci-Wellman, H. R., Corns, H. L., & Lehto, M. R. (2017). Classifying injury narratives of large administrative databases for surveillance - A practical approach combining machine learning ensembles and human review. *Accident Analysis and Prevention, 98*, 359–371.

Nanda, G., Grattan, K. M., Chu, M. T., Davis, L. K., & Lehto, M. R. (2016). Bayesian decision support for coding occupational injury data. *Journal of Safety Research, 57*, 71–82.

Papazoglou, I. A., Aneziris, O., Bellamy, L., Ale, B. J., & Oh, J. I. (2015). Uncertainty assessment in the quantification of risk rates of occupational accidents. *Risk Analysis: an official publication in the Society of Risk Analysis, 35*(8), 1536–1561.

Pietilä, J., Räsänen, T., Reiman, A., Ratilainen, H., & Helander, E. (2018). Characteristics and determinants of recurrent occupational accidents. *Safety Science, 108*, 269–277.

Podgorski, D. (2010). The use of tacit knowledge in occupational safety and health management systems. International Journal of Occupational Safety and Ergonomics (JOSE) 2010, 16(3), 283–310.

Puolamäki, K., Oikarinen, E., & Henelius, A. (2021). Guided visual exploration of relations in data sets. *Journal of Machine Learning Research, 22*, 1–32.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rissanen, M., & Kaseva, E. (2014). *Cost of lost labour input [e-publication]*. Finland: Ministry of Social Affairs and Health, Department for Occupational Safety and Health. Access method: http://stm.fi/documents/1271139/1332445/Cost+of+lost+labour+input_en.pdf/d5790088-8e3e-4d13-a5cd-56c23b67de0c.

Ross, A. J., Davies, J. B., & Plunkett, M. (2005). Reliable qualitative data for safety and risk management. *Process Safety and Environmental Protection, 83*(2), 117–121.

Savvides, R., Henelius, A., Oikarinen, E., & Puolamäki, K. (2019). Significance of patterns in data visualisations. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1509–1517).

Valmuur, K., Marucci-Wellman, H. R., Taylor, J. A., Lehto, M. R., Corns, H. L., & Smith, G. S. (2016). Harnessing information from injury narratives in the 'big data' era: Understanding and applying machine learning for injury surveillance. *Injury Prevention, 22*(Suppl 1), i34–i42.

Wang, B., Wu, C., Huang, L., & Kang, L. (2018). Using data-driven safety decision-making to realize smart safety management in the era of big data: A theoretical perspective on basic questions and their answers. *Journal of Cleaner Production, 210*(2019), 1595–1604.

Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-Value adjustment*. John Wiley & Sons, Mathematics. 360 pages.

Workers' Compensation Center [Tapaturmavakuutuskeskus] (2019). Työtapaturmat 2009-2018. Tilastojulkaisu [Occupational accident statistics 2009-2018]. Workers' Compensation Center. Access method: https://indd.adobe.com/view/baa94c89-d2b1-4fa3-b10c-f421c41208a4.

Wu, D., & Li, Z. (2019). Work safety success theory based on dynamic safety entropy model. *Safety Science, 113*, 438–444.

Tuula Räsänen is a researcher at the Finnish Institute of Occupational Health. She has over 30 years' experience of occupational safety research. She completed her PhD in 2007 at Tampere University of Technology about the management of occupational safety and health information in Finnish production companies.

Arto Reiman is a research team leader at the University of Oulu in Finland. His research interests include ergonomics & human factors and occupational safety, and how they can be included in the design and development processes to improve well-being and productivity at work.

Kai Puolamäki is Associate Professor of computer science and atmospheric sciences in the Department of Computer Science and Institute for Atmospheric and Earth System Research (INAR) at the University of Helsinki. He completed his PhD in 2001 in theoretical physics. His primary interests lie in the areas of exploratory data analysis, machine learning, and related algorithms. He has a website at http://www.iki.fi/kaip/.

Rafael Savvides is a doctoral student at the Exploratory Data Analysis group at the University of Helsinki, Finland. His research interests include visual and interactive data exploration.

Emilia Oikarinen is university lecturer at Department of Computer Science at University of Helsinki, Finland. Her research interests broadly span artificial intelligence research, ranging from knowledge representation, reasoning, and optimization to explorative data analysis with applications in a wide variety of domains.

Eero Lantto M.Soc.Sci, is a researcher at the Finnish Institute of Occupational Health, working in the Occupational Safety unit. Before FIOH Eero gained experience at Eurofound on work and employment related research topics. During the two and a half years Eero has spent at FIOH he has participated in various research and development projects in different industries.